Technologies and Systems for
Assembly Quality, Productivity and Customization

Proceedings of the 4th CIRP Conference on
Assembly Technologies and Systems

Editor: Professor S. Jack Hu

May 20-22, 2012, Ann Arbor, Michigan, USA
Preface

Welcome to the 4th CIRP Conference on Assembly Technologies and Systems, held in Ann Arbor, Michigan from May 20 – 22, 2012. Following the traditions of the past three conferences held in Trondheim (2010), Toronto (2008), and Stuttgart (2006), this conference brings together active researchers in assembly technologies and systems from around the world for two and half days to exchange ideas and explore future research directions in assembly, an area critical to successful product realization and manufacturing competitiveness.

Many of the consumer products we encounter everyday are assembled products. The product realization processes for such products usually begin with the design of the products as assembly and end with the assembly processes in the assembly factories. Assembly decomposition, supply chain, tolerance and variation management, product variety, robotics and automation, reconfiguration, joining technologies, etc. all play a critical role in the quality of assembled products and the efficiency of the production systems. As more companies embrace the concept and practice of sustainable manufacturing, disassembly and remanufacturing (re-assembly) are becoming importance new research areas in assembly technologies and systems. All these and many more areas are important topics for continued research and innovation.

We would like to express our appreciation to the keynote speaker, Dr. Daniel Whitney of Massachusetts Institute of Technology, for his outstanding overview and enlightening vision on assembly technologies. Dr. Whitney has been a leader for the past several decades in this area. Additional thanks to Professor Yoram Koren of the University of Michigan for his direction and guidance on the conference. Members of the scientific committee spent their precious time reviewing the manuscript under a tight time constraint. Their help is greatly acknowledged.

Finally, I would like to thank the members of the Organizing Committee, Dr. Tae Hyung Kim and Ben Palumbo, for their efforts in putting together this conference.

Sincerely,

S. Jack Hu
J. Reid and Polly Anderson Professor of Manufacturing Technology
University of Michigan, Ann Arbor
Organizing Committee

Conference Chairs
S. Jack Hu, University of Michigan, USA
Yoram Koren, University of Michigan, USA

International Scientific Committee

Tamio Arai, The University of Tokyo, Japan
Olga Battaia, Ecole des Mines de Saint-Etienne, France
Hulmut Bley, Saarland University, Germany
April Bryan, University of the West Indies, Trinidad and Tobago
Wayne Cai, General Motors Co., USA
Jaime Camilio, Virginia Tech, USA
Umesh Gandhi, Toyota Research Institute North America, USA
Dini Gino, Universita di Pisa, Italy
Jurgen Fleischer, Karlsruhe Institute of Technology, Germany
Erica Klampfl, Ford Motor Company, USA
Jeonghan Ko, Ajou University, Korea
Joerg Krueger, Technical University of Berlin, Germany
Terje Kristoffer Lien, Norwegian University of Science and Technology
Soh Khim Ong, National University of Singapore, Singapore
Matthias Putz, Chemnitz University of Technology, Germany
Gunther Reinhart, Technical University of Munich, Germany
Moshe Shpitalni, Technion – Israel Institute of Technology, Israel
Hui Wang, University of Michigan, USA
Wencai Wang, University of Michigan, USA
John Wang, General Motors Co., USA
Lianhong Zhang, Tianjin University, China

Local Organizing Committee

Tae Hyung Kim, University of Michigan, USA
Ben Palumbo, University of Michigan, USA
Table of Contents

Invited Keynote
What is Assembly? ......................................................................................................................................... 1  
  D. E. Whitney

Assembly Processes and Technologies

PR1: Single-sided piercing riveting for adhesive bonding in vehicle body assembly ............................ 3  
  Y. Liu, L. Zhang, W. Liu, P. C. Wang

PR2: Ultrasonic-assisted adhesive handling of limp and air-permeable textile semi-finished products in  
  composites manufacturing ................................................................................................................... 7  
  J. Fleischer, A. Ochs, S.F. Koch

PR3: Process technology and device for joining of metal with non-metal components for composite-  
  metal structures .................................................................................................................................. 11  
  R. Neugebauer, M. Putz, M. Pfeifer, E. Jäger, R. Marx

PR4: Spatial alignment of joining partners without fixtures, based on component-inherent markings..... 17  
  J. Fleischer, J. Elser

PR5: Gripper design for tolerance compensating assembly systems .................................................... 21  
  F. Dietrich, J. Maaß, K. Kaiser, A. Raatz

PR6: A cladistics approach to classification of joining and fastening methods....................................... 25  
  A. Ziout, A. Azab

PR7: Cell stacking process of high-energy lithium-ion cells .................................................................. 33  
  J. Kurfer, M. Westermeier, G. Reinhart

PR8: Interpretation of multiaxial gripper force sensors ........................................................................ 39  
  K. Tracht, S. Hogreve, S. Bosse

PR9: Calibration of reconfigurable assembly cell using force feedback and vision............................ 43  
  S. Dransfeld

PR10: Picking precision in assembly using robots and flexible feeders .............................................. 47  
  S. Dransfeld, L. E. Wetterwald

PR11: Influence of welding sequences on compliant assembly geometric variations in closure  
  assemblies ........................................................................................................................................ 51  
  A.F. Nagy-Sochacki, M.C. Doolan, B.F. Rolfe, M.J. Cardew-Hall

PR12: Estimation of the Weldability of Wingle-Sided Resistance Spot Welding................................. 55  
  D. Kim, Y. Cho, Y. H. Jang
Reconfigurable Assembly Systems

R1: MS design methodology for automotive faming systems BIW .......................................................... 59
    A. Al-Zaher, Z. J. Pasek, W. ElMaraghy

R2: Assessing the structural complexity of manufacturing systems layout ............................................. 65
    V. Espinoza, H. ElMaraghy, T. AlGeddawy, S.N. Samy

R3: Optimising the process chain and achieving flexibility in rear axle alignment – an integrated view
    of chassis assembly ............................................................................................................................ 71

IDEAS

IDEAS1: Configuration model for evolvable assembly systems............................................................... 75
    P. Ferreira, N. Lohse

IDEAS2: Evolvable Assembly Systems: entering the second generation.............................................. 81
    M. Onori, J. Barata, F. Durand, J. Hoos

IDEAS3: Operational characterization of evolvable production systems............................................. 85
    H. Akillioglu, A. Maffei, P. Neves, J. Ferreira

IDEAS4: IADE – IDEAS agent development environment: lessons learned and research directions ...... 91
    L. Ribeiro, R. Rosa, A. Cavalcante, J. Barata

IDEAS5: Distributed bayesian diagnosis for modular assembly systems – a case study ......................... 95
    M. S. Sayed, N. Lohse

Assembly Quality

Q1: Geometry assurance versus assembly ergonomics - comparative interview studies in five
    manufacturing companies .................................................................................................................. 101
    M. Rosenqvist, A. Falck, R. Söderberg

Q2: Compensation of shape deviations for the automated assembly of space frame structures............ 105
    J. Fleischer, M. Otter

Q3: The impact of clamping and welding sequence on the dimensional outcome of a single-station
    compliant assembly: a production case study .................................................................................. 109
    T.I. Matuszyk, M.J. Cardew-Hall, P. Compston, B.F. Rolfe

Q4: Non-nominal path planning for increased robustness of robotized assembling............................ 113
    D. Spensieri, J. S. Carlson, R. Söderberg, R. Bohlin, L. Lindkvist

Q5: Statistical shape modeling in virtual assembly using PCA-technique ............................................. 119
    B. Lindau, L. Lindkvist, A. Andersson, R. Söderberg
Q6: A bio-inspired approach for self-correcting compliant assembly systems ................................. 125
L. J. Wells, J. A. Camelio

Man-Machine Collaboration

M1: Human operator and robot resource modeling for planning purposes in assembly systems .......... 131
J. Provost, Å. Fasth, J. Stahre, B. Lennartsson, M. Fabian

M2: The influence of assembly design methods on work exposures – an analytical examination ........ 137
J. Egbers, G. Reinhart, A. Hees, W. P. Neumann

M3: Training by augmented reality in industrial environments: a case study .................................... 141
S. Iliano, V. Chimienti, G. Dini

M4: Interaction between complexity, quality and cognitive automation ............................................. 145
T. Fässberg, Å. Fasth, F. Hellman, A. Davidsson, J. Stahre

M5: A hybrid human-robot assistance system for welding operations - methods to ensure process quality and forecast ergonomic conditions ................................................................. 151
F. Busch, C. Thomas, J. Deuse, B. Kuhlenkoetter

M6: Design for collaboration: a development of human-robot collaboration in assembly ......................... 155
J. T. C. Tan, T. Inamura, T. Arai

Assembly System Planning

PL1: Adaption of processing times to individual work capacities in synchronized assembly lines .......... 161
G. Reinhart, M. Glonegger, M. Festner, J. Egbers, J. Schilp

PL2: Automatic assembly path planning for wiring harness installations ........................................... 165
T. Hermansson, R. Bohlin, J. S. Carlson, R. Söderberg

PL3: Conceptual DFA method for electric vehicle battery systems .................................................. 169
A. Tornow, A. Raatz

PL4: Beyond human tetris: simulation-based optimization of personnel assignment planning in sequenced commercial vehicle assembly ................................................................. 175
L. März, W. Mayrhofer, W. Sihn

PL5: Assembly path planning by distance field based shrinking ...................................................... 179
S. Björkenstam, J. Segeborn, J. S. Carlson, R. Bohlin

PL6: A classification of carrier and content of information ............................................................... 183
T. Fässberg, Å. Fasth, J. Stahre

PL7: Cost impact assessment of production program changes: a value stream oriented approach ....... 187
J. Gottmann, W. Mayrhofer, W. Sihn
PL8: Discovering design structure matrix for family of products ............................................................ 191
  M. Kashkoush, T. AlGeddawy, H. ElMaraghy

PL9: Enhanced mixed integer linear programming for flexible job shop scheduling ....................... 195
  V. Roshanaei, H. ElMaraghy, A. Azab

PL10: Allocation of maintenance resources in mixed model assembly systems ............................... 199
  W. Guo, J. Jin, S. J. Hu

PL11: Product architecting for personalization ...................................................................................... 203
  C. Berry, H. Wang, S. J. Hu

PL12: Automatic creation of virtual manikin motions maximizing comfort in manual assembly
  processes ............................................................................................................................................... 209
  R. Bohlin, N. Delfs, L. Hanson, D. Högberg, J.S. Carlson

PL13: An assembly decomposition model for subassembly planning considering imperfect inspection
  to reduce assembly defect rates ............................................................................................................. 213
  J. Ko, E. Nazarian
What is Assembly?

D. E. Whitney

Keynote presentation to the 4th CIRP Conference on Assembly Technologies and Systems

May 21, 2012

Summary:

“Assembly” is a process that puts parts together to create a product or system.

An “assembly” is such a product or system.

The fact that the words are the same obscures the differences between them in spite of the links between them that are established during the process of designing an assembly and designing its assembly process.

Assemblies are complex systems that are intended to meet specific customer needs. These needs must be converted into specific geometry, materials, dimensions, and tolerances. In particular, the performance of the assembly depends on achievement of particular dimensional and geometric relationships between different elements (typically called parts) of the assembly. The assembly process consists of joining the parts and ensuring that the required geometric relationships are achieved. Depending on materials and constraints, the process may require the presence and use of temporary parts called fixtures, which mimic or supplement some of the geometric relationships when the parts are unable to do so. The process of designing an assembly and its assembly process can be aligned with the System Engineering V, a view of product-process design that is driven by the original customer needs and is intended to achieve them.

There is abundant research on some aspects of assembly design as defined above. Some other aspects are under-researched. These second aspects will be emphasized in this presentation.
Single-sided piercing riveting for adhesive bonding in vehicle body assembly

Y. Liu\(^{a}\), L. Zhang\(^{a}\), W. Liu\(^{a}\), P. C. Wang\(^{b}\)

\(^{a}\) Tianjin Key Laboratory of Equipment Design and Manufacturing Technology, School of Mechanical Engineering, Tianjin University, Tianjin 300072, China
\(^{b}\) General Motors R&D Center, Warren, MI 48090, USA

Abstract: With increasing use of advanced high strength, lightweight materials and alternate vehicle architectures, materials joining issues have become increasingly important in automotive vehicle body assembly. Among the various joining methods, adhesive bonding is one important technique for joining dissimilar materials in vehicle body assembly. During curing of the adhesive, it is necessary to fix and control the gap between the sheet panels to ensure the bonding strength and geometric quality. In this paper, a novel process—the single-sided piercing riveting (SSPR), is proposed for fixing and controlling the gap between the sheet panels during adhesive bonding of vehicle body assembly. The SSPR is a cold forming process for joining two or more sheet parts by driving a specifically designed U-shaped rivet using an impact force. The rivet pierces through and flares in the sheets to form a mechanical interlock between the sheet parts. The process is easy and convenient to implement without the need of a direct back support, and can meet the limited space requirement of vehicle body assembly. Experiments and joining performance tests were carried out to validate the effectiveness of the SSPR process.

Keywords: Single-sided piercing riveting (SSPR), Adhesive bonding, Vehicle body assembly

1. Introduction

Lightweight is essential to reduce fuel consumption and gas emission of vehicles [1, 2]. Great effort has been made to reduce vehicle mass, especially through high strength/lightweight material development and application [3-5]. Most of the high strength/lightweight materials, such as advanced aluminium alloys, magnesium alloys, and non-metallic composites, have poor weld compatibility due to their low electrical resistivity and higher thermal conductivity [6]. This lightweight effort thus been faced with challenge in joining for vehicle assembly [7]. Adhesive bonding [8-10] has emerged as a sound alternative to the joining challenge since its compatibility to materials from metallic to non-metallic, improved crash toughened formulation of full temperature range capability, and cost competitiveness [11, 12].

For adhesive bonding, there exist two processing prerequisites: 1) properly controlling the thickness of the adhesive or the gap between the sheet parts; 2) fixing the parts in location until full completion of adhesive curing. To satisfy these prerequisites, auxiliary fixtures are needed. Rivet joining or other similar methods, seems to be an alternative to the auxiliary fixtures. Traditional solid rivet riveting needs predrilled holes, which makes the riveting process complex and time consuming [13]. Conventional clinching (also named press joining) [14], dieless clinching [15], and SPR [16-18], though without the need of a predrilled hole, all necessitate a specific die or anvil to provide back support to the riveting punch load. These joining processes are not competitive to be auxiliary fixtures for adhesive bonding of vehicle body assembly on considerations of easy and convenient handling, operating space limitation, and cost effectiveness.

To develop a solution to the auxiliary fixtures for the adhesive bonding, a novel process—the single-sided piercing riveting (SSPR), is proposed in this paper. The SSPR is able to join two or more sheet parts by driving a specifically designed and made rivet that pierces through and fastens the sheet parts using an impact force. It is easy and convenient to use without the need of a direct back support, and therefore can meet the limited space requirement of vehicle body assembly.

This paper presents a preliminary experimental study on SSPR and the joining performance of the SSPR related joints to validate the effectiveness of the SSPR process. The paper is organized as follows: Section 2 introduces the principle and try-out experiments of the SSPR, Section 3 presents the Performance of the SSPR and the SSPR-bonding joints, and Conclusions are presented in Section 4.

2. Principle and try-out of the SSPR

The idea of SSPR is driven by the practical needs of industry—to provide a convenient and easy to handle joint to control the proper thickness of the adhesive or gap between sheet parts, and comes from the observation of the phenomenon of a rapid moving sharp nail penetrating target objects. As the great ratio of mass and inertia of the target object to the nail, the target object will keep almost static while being impacted or even penetrated through by the nail without the need of a counteractive support.

The prerequisites of SSPR are: 1) the rivet must be stiffer than the sheet parts; 2) the rivet must reach a critical speed; 3) the sheet parts are of plasticity.

Therefore, the SSPR can be conducted by shooting and making the rivet pierce into the sheet parts to be joined. To explore the possibility of SSPR, three commercial shoot devices, three commercial nails or staples, and three newly designed and made rivets were prepared and tested to pierce or rivet eight groups of sheet combination, as shown in Table 1.

The try-out results of the SSPR are presented in Table 1. The commercial powder nail gun with its consistent commercial nail was too powerful for piercing the double steel sheets. The commercial F30 and 442J air guns together with its consistent commercial nails were only able to successfully pierce the Al single sheet due to the dull tip of the nails. While with the tailor-made I-shape sharp tip nail and U-shape rivets of chamfered sharp tip legs, both the Al and steel double sheets were...
mechanical locks were achieved, which is able to fasten or fix. Worked with the tailor-made U-shape rivets, permanent available at the lab. Maximum pressure air (0.7 bar compressed were driven via Note: the air guns. Table 1: Try-out of the SSPR. Shooting devices Rivets Sheets Effects Notes Com’t Powder nail gun Com’t nail diam: 2.0 mm, MAT.: AISI-1566 steel. 1.0 mm. Double sheets. MAT.: AISI-1035 steel. The steel sheets were pierced but distorted via the great energy of the powder. Com’t F30 air gun Com’t nail diam: 1.2 mm, MAT.: soft steel. 1.5 mm. Single sheet. MAT.: AISI-6063Al. The Al single sheet was successfully pierced. Com’t staple diam: 1.2 mm, MAT.: soft steel. 1.5 mm. Double sheets. MAT.: AISI-6063Al. The Al and steel double sheets were successfully pierced. Note: the air guns were driven via 0.7 bar compressed air (0.7 bar was the maximum pressure available at the lab). Com’t 422J air gun Legs outer chamfered. 1.5 mm. Single sheet. MAT: AISI-6063Al. The Al single sheet was successfully pierced. But the legs interfered. The single and double Al sheets were successfully pierced. Fasten lock was formed between the double sheets. Legs inside chamfered. 1.5 mm. Single sheet. MAT: AISI-6063Al. The Al single sheet was successfully pierced. But the legs interfered. The single and double Al sheets were successfully pierced. Fasten lock was formed between the double sheets. U-shape rivet diam: 1.2 mm. MAT.: AISI-1566 steel. 1.5 mm. Single sheet. MAT: AISI-6063Al. The Al single sheet was successfully pierced. But the legs interfered. The single and double Al sheets were successfully pierced. Fasten lock was formed between the double sheets. Effects: Note: The Al and steel sheets were dry-laid. It can be found from Table 1 that the chamfer on the legs of the tailor-made U-shape rivets is positive to the joining effect. This is because the chamfer can lead the two rivet legs flare inside or outside during piercing into the sheets to form a mechanical locks among/between the sheets. In comparison the lab U-shape rivet with inner chamfered legs is superior to that with the outer chamfered legs, since the outer chamfers may cause interference of the two legs and weaken the joining effect. The try-out of SSPR demonstrates that it presents expectant joining effect with the combination of the commercial 422J air gun and the newly lab designed and made U-shape rivet with inner chamfered legs. This combination of shooting device and rivet is further verified with improved design in Section 3. 3. Performance of the SSPR and the SSPR-bonding joints To validate the effectiveness of the SSPR and to verify the performance of the SSPR and the SSPR-bonding joints, the U-shape rivet was redesigned as shown in Figure 1, and prepared with a specific setup via 1.2 mm diameter AISI-1566 steel wires of hardness HRC50. Sheet specimens were fabricated with AISI-5052 Al sheets of 2 mm in gage. To ensure the consistency of the SSPR and SSPR-bonding joints, the speed of the piston of the air gun was measured with a high-speed camera and calibrated to 40 m/s. The bonding adhesive for the joints was Henkel Terokal 5087-02. A schematic illustration of the SSRP-bonding process is shown in Figure 2. The sheet surfaces were firstly cleaned to remove the oil and dust. Then the adhesive was uniformly pasted on one of the cleaned surfaces, and some glass balls of 0.2 mm in diameter were dispersed into the adhesive for controlling the thickness of the adhesive layer. Then the other sheet was put on above the adhesive pasted sheet to form a sheet-adhesive-sheet semi-assembly. Afterwards the semi-assembly was clamped to ensure the thickness of the adhesive layer to be 0.2 mm for optimal bonding (Figure 2 (a)). After clamping, the SSPR process was started and the rivet was forced by the piston rod of the air gun to pierce through the upper sheet and into the bottom sheet, and finally the SSPR joint was formed (Figure 2 (b), (c) and (d)). Following the SSPR, the SSPR joined sheet-adhesive-sheet semi-assembly was baked at 180°C for 30 minutes to make the adhesive fully cured and form the SSPR-bonding joint. Figure 5 shows the typical shear performance of the joints. The lap-shear and coach-peel tests were adopted to measure the static mechanical performance of the joints. Lap-shear coupons of four joint groups: SSPR, SSPR-bonding, bonding, and self-piercing riveting (SPR) bonding, together with coach-peel coupons of two joint groups: SSPR-bonding and bonding, were prepared for joining performance test and comparison, as shown in Figures 3 and Figure 4. The lap-shear and the coach-peel tests were performed on a CSS-44100 universal test machine via a cross-head speed of 2 mm/min. Figure 5 shows the typical shear performance of the joints. It can be found that the adhesive is dominant to the shear performance of the x-bonding joints. Both the SSPR and the SPR have poor ultimate shear load and fracture extension. The SPR is stronger than the SSPR due to its fastener riveting. This fastener riveting of the SPR lessens the adhesive bonding greatly because its strong clamping load forces the adhesive out of its large rivet.
affecting area from between the sheets (Figures 6 and 7(c)). As shown in Figure 6(a), the adhesive layer of the SSPR-bonding joints is uniform except in the vicinity of the rivet legs. Moreover, the damaged area of the adhesive layer is two small circle areas with the diameter of 1.2 mm. While for the SPR-bonding joints, the bond-line thickness of adhesive layer is gradually thinner from the edge to the center of the joint, as shown in Figure 6(b), and there is a large circular area with no adhesive around the rivet. Diameter of the area without adhesive is about three times of the SPR rivet leg diameter of 5.3 mm. This phenomenon can be validated through the failure interfaces of the SSPR-bonding joints and SPR-bonding joints, which is shown in Figures 7(b) and 7(c). According to the measurements for the failure interfaces of the SSPR-bonded and SPR-bonding joints, the whole damaged area of the adhesive layer in SSPR-bonding joint is about 7 mm², while it is about 18 mm² in SPR-bonding joint. For the above reasons, the SSPR-bonding joint presents almost the same shear performance as that of the adhesive bonding joint before the ultimate shear load is reached. Furthermore, the SSPR-bonding joint shows a residue strength tail after the ultimate shear load when the adhesive bonding fails and loses its full strength (Figure 5). This residue strength tail improves the toughness of the SSPR-bonding joint as compared with the bonding joint. Thereby, using the SSPR-bonding joint is desirable for the safety of vehicle structures.
the rivet legs of the SSPR-bonding joint; and consequently the secondary dull peak load is generated the failure of the joint is delayed. This means that the SSPR can serve as a peel-stopper when peeling reaches the SSPR rivet legs. The peel-stopper absorbs peeling energy and delays the fracture, which may benefit the crash safety of vehicles.

The lap-shear and the coach-peel test show that the SSPR can provide effective clamping and fixing to adhesive bonding of sheet assembly, while presenting minimum influence to the shear performance and improvement to the peel performance of the SSPR-bonding joint.

4. Conclusions

The single-sided piercing riveting (SSPR) is proposed in this paper for fixing and controlling the sheet panel positions during adhesive bonding for vehicle body assembly. The SSPR process was explored via try-out experiments and validated via joining performance tests. The following conclusions can be drawn from the research:

(1) The SSPR is easy and convenient to handle without the need of a direct counteractive support, which can meet the limited space request of vehicle body assembling.

(2) The SSPR can provide effective clamping prior to and during curing of the adhesive to ensure the bonding strength.

(3) The SSPR has less influence on the adhesive distribution and thus ensure the mechanical performance of the SSPR-bonding joints comparable to that of the adhesive bonding joints.

(4) The rivet of SSPR can serve to improve the shear toughness performance and act as a peel stopper to improve the peel performance of the SSPR-bonding joints.

Acknowledgements

The authors would like to express their sincere thanks to GM R&D Center for providing support for the research.

References


Abstract: The handling of limp and air-permeable textile semi-finished products is a huge challenge for the automated production of continuous-fiber reinforced polymers. This is why semi-finished textile products are mainly provided and fed manually. It is against this backdrop that this article begins by defining the requirements for an effective gripper system for semi-finished textile products based on the individual process steps. The state-of-the-art of the gripper systems is then evaluated according to these criteria. Based on the findings, a novel, vibration-based gripper system was set up and simulated using the Finite Element Method at the Institute of Production Science.

Keywords: Composite, Handling, Automation, Ultrasonic, Flexibility

1. Introduction
Continuous-fiber reinforced polymers (FRP) are superior to metal materials in many aspects, e.g. weight-related stiffness and strength [1]. FRPs offer other benefits such as design leeway [2] and adjustable thermal expansion coefficients. The use of FRP in motor sports and in the aerospace industry has become state-of-the-art. The possibilities of transferring findings to series production of vehicles are limited for reasons of profitability and of technical feasibility. The tried and tested methods for the production of these light-weight products, i.e. mainly carbon fiber or glass fiber reinforced products, such as Resin Transfer Molding (RTM), mainly include manual operations because there is still a lack of automation strategies in this field. The focus on manual labor represents an obstacle to the economic production of FRP components. The handling of limp and air-permeable semi-finished textile products or preforms, which are used for almost all manufacturing processes, is a huge challenge for the automated production of FRPs [3,4,5].

2. Gripper systems for textile semi-finished products
The manufacturing process for advanced composites always consists of multiple stages. The big challenges arising from the automation of such process chains [3] are highlighted by the example of the RTM technology. RTM is one of the technologies that are suitable for the large-volume production of continuous-fiber reinforced structural components with complex shapes [6] used in the automotive industry.

2.1 Process chain and resulting requirements for gripper systems
For conventional RTM processes, the basic materials of composite fiber components, i.e. textile fibrous semi-finished products, are generally provided in the form of rolls. Textile semi-finished products are provided as random fiber mats, fiber layers and woven fabric. In the first process step, the textiles are cut according to the intended component contour. These sheet cuts are then sorted and are further processed into one or several preforms. By definition, a textile preform is a dry fiber material with a contour that is close to the intended end contour and with a fiber structure that is sufficiently resistant to the intended load application. Once the reinforcement fibers have been positioned in the mold, the two-piece mold is closed. Then, a low-viscosity thermo-setting blend of resin and hardener is injected into the mold. The resin begins to cure depending on the resin formula and on the temperature. During curing, the resin system starts to cross-link until a dimensionally stable composite fiber component is formed.

Handling is one of the main operations within the automated production of composite fiber components. In that respect, handling does not mainly refer to transportation processes, but to all operations where semi-finished fiber products must be picked up and gripped carefully. Damage affecting material characteristics caused by handling, such as the “trelis” effect, fiber elongation, crease formation, fiber stretching and fiber displacement, causes local deficiencies and optical damage in the components. Therefore, this damage must be avoided to the maximum possible extent. Fig. 1 shows the main individual process steps required for the RTM production of components and the handling operations required for them.

The handling operations mainly serve to link the different processes. From the roll to the finished part, the textile semi-finished product undergoes several process steps. During the process the contour, the dimensional stability, the air permeability, the sensitivity (see damages affecting the material characteristics), the surface adhesiveness and the basis weight (Fig. 1) can change. From the above it can be seen that the requirements that must be met by gripper elements must be derived from the characteristics of the semi-finished products, of the end components and from the individual process steps.

The handling technology and the gripper systems in particular are most challenged during the sorting of the sheet cuts. This is when the mostly dry textile semi-finished product is highly sensitive, highly air-permeable and very susceptible to the
afore-mentioned damage that affects material characteristics due to the fact that the dimensional stability is very low.

Different binder systems are used to transform the semi-finished product into a preform. The binder systems are mainly applied to hold the individual textile layers in place. However, this setup in combination with the thermoplastic and thermosetting binder material application and the layer construction leads to a decrease of the air permeability and the sensitivity to and an increased dimensional stability. When a thermosetting binder system, i.e. a spray adhesive most of the times, is used, the surface adhesiveness of the preform will also increase. The sensitivity of the textiles against damage affecting the material characteristics is lower downstream of the preform step, but the 3D structure and the adhesiveness of the preform surface lead to additional requirements for the gripper system.

After infiltration and curing of the components, the requirements that must be met by the gripper systems are lower. From that time on, the components are dimensionally stable and air-impermeable, which is why they can be gripped with established technologies, such as vacuum systems, for example.

2.2 State-of-the-art and evaluation of gripper systems

Different gripper systems [5,7,8,9,10,11,12] are used to handle limp materials. Fig. 2 shows a selection of these gripper systems.

<table>
<thead>
<tr>
<th>Physical principles for grippers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form fit</td>
</tr>
<tr>
<td>Force fit</td>
</tr>
<tr>
<td>Material closure</td>
</tr>
<tr>
<td>Surface hooking</td>
</tr>
<tr>
<td>Low pressure</td>
</tr>
<tr>
<td>Electrostatic forces</td>
</tr>
<tr>
<td>Molecular forces</td>
</tr>
</tbody>
</table>

- Needle gripper
- Coanda gripper
- Bernoulli gripper
- Electrostatic gripper
- Freezing gripper

The holding force of mechanical gripper systems required to handle the components is applied through either force or form closure. Needle grippers are the most frequently used system for textiles. The counter-directional needles pierce the fabric, thus creating a form closure connection between gripper element and component. The holding forces of pneumatic gripper systems are generated by creating a vacuum. Among the established systems you will find Bernoulli grippers, low-pressure surface suction [5] grippers and coanda grippers [7]. The vacuum technology has become the state-of-the-art for this kind of application. These systems are based on the low-pressure principle. Adhesive grippers create a connection with the textile material based on force closure or closure by adhesive force. Current research activities are looking into the use of electrostatic grippers and freezing grippers [11,12], which are tried and tested systems for classic textile applications.

Fig. 3 gives an evaluation of the state-of-the-art based on component and material specific requirements and on the requirements derived from the process. Needle grippers are characterized by the secure retention of the textile semi-finished products. However, it is difficult to pick up single textile layers, because the needles can penetrate and thus, grip several layers at once. If this kind of gripper system is used, the needles must be precisely adjusted, thus determining the penetration depth accurately. In addition, the punctures caused by the needles are considered problematic. Depending on the application, the punctures can affect the optical appearance of the product or can be detrimental to its quality.

The benefits of the pneumatic systems are the simple design and the gentle handling of the textiles. It has become possible to single the textile layers reliably by arranging the individual grippers accordingly. In addition to Bernoulli grippers, more and more grippers based on the Coanda principle [7] are used. Compared to the other systems, the Coanda gripper has the advantage of allowing the transfer of normal and lateral forces and of being flexible in terms of the materials that it can handle. As with all grippers that are based on the low pressure principle, the following disadvantages apply: only comparatively low lateral forces can be transferred, dirt particles are drawn in and large suction surfaces for the suction elements are required. Because of their design and function, Bernoulli grippers can only absorb lateral forces via additionally fitted devices such as rubber elements. These elements have the potential to damage the textiles. Furthermore, tests have shown that the airflow required for the Bernoulli gripper can cause the material to fray. This applies to the fringes of limp textiles. Compared to the above-mentioned classic systems, adhesive systems have benefits regarding gripping reliability and transferable holding forces. The high flexibility of adhesive systems as regards the range of materials and components for which they can be used is another huge benefit [8]. The suitability of freeze grippers for RTM components is questionable because many material systems are incompatible with even small amounts of water.

The above leads to the conclusion that adhesive systems are beneficial over conventional system with regard to the handling of textile semi-finished components, in particular. If combined with the use of system-compatible media, these systems have a huge potential for use in composite fiber applications.

3. Ultrasonic-assisted gripping

Research carried out at the wbk is looking into the development of novel gripper systems that are able to provide
optimal support for each of the previously defined process stages. The system that has been developed is based on piezo-electrically generated vibrations. The vibrations are applied to release textiles that are held in place by means of adhesive force. If the vibrations are applied selectively, the adhesive forces can be overcome when the semi-finished products are placed. The effect of mass inertia that objects adhered to an element have is used in order to overcome the effective adhesive forces. The adhesives used are either system-compatible adhesives or binder systems commonly used for preforming processes. An ultrasonic resonance system is used to achieve the acceleration required to detach the materials. The basic structure, i.e., converter, booster and sonotrode, is provided in the form of an ultrasonic weld system. Previous tests have proved the handling principle to work [13]. This system shows comparable advantages like the freezing gripper technique, but a wide range of system-compatible adhesives can be used. However, there are cost disadvantages which are primarily caused by the used piezo technology. A particular challenge with this process is the minimally detachable mass and therefore the calculation of the producible forces. For this reason, the focus of the work presented in this contribution concentrates on the illustration of the operating principle and on the basis of calculation of the producible detachment forces.

The acceleration response \( s(t) = \frac{d^2s(t)}{dt^2} \) is as follows:

\[
s(t) = 3 \cdot 4 \cdot \pi^2 \cdot f^2 \cdot \sin(2 \cdot \pi \cdot f \cdot t) \tag{2}
\]

The maximum acceleration \( s_{\text{max}} \) is effective whenever the \( \sin \) term is equal to one (3). For the test set-up, amplitudes of approximately 32 \( \mu \)m are achieved at the tip of the sonotrode for a frequency of 35 kHz. This corresponds to a maximum acceleration of more than 1,500,000 \( \text{m/s}^2 \) at the tip of the sonotrode. The adhering part is released (3) when the generated mass inertia forces calculated from the maximum acceleration at the tip of the sonotrode \( s_{\text{max}} \) and the effective component mass \( m_{\text{dyn}} \) divided by the cross-sectional area \( A_{\text{adh}} \) exceed the breaking stress of the adhesive \( \sigma_B \).

\[
\sigma_B > \frac{s_{\text{max}}}{A_{\text{adh}}} \tag{3}
\]

If the geometrical specifications of the gripper tip, the breaking stress of the adhesive \( \sigma_B \) and the maximum achievable acceleration (3) at the tip of the sonotrode are used as a basis, the minimum mass \( m_{\text{min}} \) (4) required to dissolve the adhesive joint can be calculated. This correlation works both ways. If the range of components that must be handled with the system is known, it is possible to determine on that basis what strength the adhesive can have as a maximum \( \sigma_{B, \text{max}} \) (4) to still allow for the reliable release of the material.

\[
m_{\text{min}} = \frac{\sigma_B \cdot A_{\text{adh}}}{3 \cdot 4 \cdot \pi^2 \cdot f^2} \tag{4}
\]

The geometrical specifications of the sonotrode tip and the maximum achievable accelerations are determined by the design and by the process. However, the required strength of the adhesive is determined by the medium that is used and can vary accordingly. The mass that is relevant for the calculation of the effective forces is of particular importance.

### 3.2 Dynamically effective volume

In addition to the geometrical specifications of the sonotrode tip, there are three factors that must be taken into consideration in particular: the maximum achievable acceleration at the tip of the sonotrode, the breaking stress of the respective adhesive and the mass that is relevant for the generation of force are required to calculate the detachment force. The design of the vibration-assisted system is developed using the FEM, which is a common approach for all classic ultrasonic weld systems. For this purpose, an eigenfrequency analysis for the validation of the resonance frequency is carried out at first and subsequently a quasi-static, dynamic mode analysis (fixed support at the zero point of the booster). With the mode analysis stimulating forces are calculated by using the mass of the converter and its vibration behavior and applied as pressure load. A modal damping approach serves for the adaptation of the oscillation amplitude at the sonotrode tip to the manufacturer’s instructions of the real ultrasonic unit. The values for force amplitude and damping are made further use of for simulation with a bonded textile. The breaking stress of the adhesive can either be seen from the data sheets or can be determined through tests. The process of determining the mass that is relevant for the generation of forces is a lot more complex. This is why the therm and concept of the ‘dynamically effective volume’ will be introduced. It describes the vibration characteristics of the adhesive in combination with the adhering
part under high-frequency excitation. It can be observed that only part of the adhering mass can be used for the calculation because of the low level of stiffness of the textile material. Only a small part of the material has an impact on the generation of detachment forces caused by the effective mass inertia. Tests and simulation have shown that the dynamically effective volume depends on the excitation frequency and the material parameters of the respective adhesive and the textile.

3.3 Determination of the dynamically effective volume and the achievable forces through the performance of an FEM

A FEM simulation was performed in order to simulate the complete handling system consisting in ultrasonic resonance system, adhesive and textile. The results are illustrated by Fig 5. The textile was modeled of 2D elements. Here, the mesh size corresponds to the width of a roving so that each of the meshing elements represents rovings. For the adhearing 3D elements were used. Their mesh is modeled in a significantly finer manner than in comparison to the textile or the ultrasonic system. An improvement of the outcome quality is not achieved by means of finer meshes. Using the simulation, the dynamically effective volume of the textile and the effective stress within the adhesive joint can be determined, for example. For the evaluation of the dynamically effective volume, the amplitude distribution of the textile is considered in relation to the overall surface area. There are three relevant ranges that can be identified (Fig. 5):

- Range 1: quasi-constant acceleration amplitude
- Range 2: linearly decreasing acceleration amplitude
- Range 3: inactive range; Actions of moments offset each other mutually, whereas actions of forces offset each other approximately.

![Simulated amplitude distribution of the textile](image1)

![Simulated stress (Mises) curve of the adhesive](image2)

**Figure 5.** Dynamically effective volume and achievable stresses (FEM)

The example shows that the chosen approach is the only viable option. For the simulation, a carbon fiber fabric with binder material with a square base area of 50 mm x 50 mm and an overall mass of 1.79 g was chosen. The material behavior of the bound textile has been assumed to be orthotropic and estimated in a highly simplified way with values according to the Voigt-Reuss models. Both the diameter at the sonotrode tip and the diameter of the adhering are presumed to be the same size of 14 mm. For these parameters, the dynamically effective volume would be 17.35 % of the overall volume of the textile. The dynamically effective mass, which acts on the adhesive joint in the form of mass inertia forces, is only 0.31 g. The same simulation also allows determining the distribution of stress within the adhesive layer (epoxy adhesive). The stress in the fringe area amounts to approximately 2.65 MPa. The calculation of the detachment force is based on the assumption that the adhesive has cured completely. The analytical stress calculation from (5) can be adapted with the findings from the FEM simulation as well. The modified calculation is then applied on the regions of quasi-constant acceleration (Fig. 5, Range 1) and the linear decreasing amplitude (Fig. 5, Range 2).

\[
\sigma_{\text{b,max}} < \frac{m_{\text{dyn,1}} \cdot \dot{s}_{\text{max}} + \int_{r_1}^{r_2} m_{\text{dyn,2}}(r) dr \cdot \frac{1}{r_2} \cdot \frac{1}{r_1} \cdot \frac{1}{r_2} \cdot \frac{1}{r_2}}{A_{\text{adh}}}
\]

Taking into account the dynamic effective regions from the FEM simulation, the previously calculated results for acceleration and the mass of the adhesive, the analytical calculation from (6) results in a stress of 2.46 MPa. Under consideration of the dynamically effective volume, the congruence between FEM and analytic calculation can be rated as very good.

4. Conclusion

The vibration-assisted gripper system described in this article combines many advantages of conventional systems. The main benefit of the technology is that it allows the secure and gentle gripping of the preforms. Based on these research activities, a new gripper element will be developed. It will combine the advantages of the vibration-assisted system and those of a low-pressure surface suction gripper. If both systems are combined, an adjusted gripper system can be developed that can be used for the entire process chain from the sorting of the cuts to the placing of the preform in the mold.

References


Process technology and device for joining of metal with non-metal components for composite-metal structures

R. Neugebauer, M. Putz, M. Pfeifer, E. Jäger, R. Marx
Fraunhofer Institute for Machine Tools and Forming Technology, Chemnitz, Germany

Abstract: Modern composite materials are increasingly used in the sheet metal fabricating industry, especially in the automotive sector. Future joining and assembling technologies are absolutely necessary to handle the multi-material combinations that occur. The paper focuses on the development of a new riveting technology. An innovative robot guided riveting tool has also been designed. Several further applications seem to be possible, e.g., clinching, blind riveting, screwing and friction-stir spot welding. Design matters are discussed as well as real test results. An immediate application of this flexible joining system is planned especially for small and medium sized businesses.

Keywords: Composite; joining; robot tool

1. Introduction
Joining technology has developed in recent years into an increasingly sophisticated production technology, since the trend in the field of materials technology is towards increasing use of modern high-strength and lightweight materials.

To increase energy efficiency significantly and minimize the use of materials for sustainable and resource-efficient production, high-strength steels (St), aluminium (Al), as well as fibre-reinforced plastic composites (FRPs) are finding more and more applications in the construction of vehicles, especially in the bodywork, and railway vehicles. Fibre composites have already shown their worth in the aerospace industry due to their excellent material properties, such as very high strength at low specific material weights. Since the manufacturing processes for FRPs are gradually becoming cheaper, these new materials are finding increasingly widespread use in mass production at automobile manufacturers. There is already a multitude of fibre-reinforced lightweight components and subassemblies integral to production, and joints to metal components have to be made. The preferred use of FRPs, mainly based on carbon fibres (CFRPs) represents a new challenge for joining technology [1 - 8].

At the Fraunhofer Institute for Machine Tools and Forming Technology (IWU), efforts are being made to develop an adequate bonding technology tailored to the problem of composite construction and material mixtures. For this purpose, together with the Voith Engineering Services Ltd, the framework for a research project was set.

The focus of the research is the development of a new point-like mechanical joining technology for multi-material design (MMD). Firstly, this includes the development of a new type of rivet and the associated setting process and, secondly, the design and construction of a flexible robotic joining system.

2. Market situation and research approach
2.1. Problems for joining technology in composite construction
The nature of the FRP varies greatly depending on the application; the reinforcing fibres used and their arrangement can also be very different. For high-strength components, mostly continuous fibres such as glass (GRP), carbon or aramid are used in the form of mats and weaves. Together with a thermoplastic or thermosetting plastic matrix, the "shaping" of the composites is realized in a variety of manufacturing processes. Even short-reinforced materials can be found in small and attached parts again, typical joining elements are often just body bolts in such cases. However, long-reinforced high-strength components made straight from CFRP fabrics (typically carbon-looking) require more sophisticated joining processes for the potential of the fibre to be exploited to its best. Material-fit connection technologies, such as bonding and mechanical spot-joining techniques, such as collar or blind rivets, are currently used for hybrid combinations.

Problems for mechanical joining methods are mainly caused by the necessary pilot hole operation with the subsequent positioning of the connecting element and the requirement for surface flatness in the joint area. Sheet plane connectivity technologies such as conventional self-pierce riveting have so far been unable to sufficiently exploit the strength of high-strength fibres. The insertion of the connecting element, depending on the fibre volume content, splits the majority of the fibres and thus the force distribution is disturbed or interrupted [see Figure 1].

The damage and fracture of fibres in the contact area means that the flow of forces is often transmitted only through the polymer [1]. In this case, creep processes and the small form fit at the rivet head often mean that no long-term stability can be achieved. There are additional problems in the contact zone of the carbon fibres, since galvanic corrosion occurs due to the different electrochemical potentials and the joint connection is additionally and permanently damaged [2].

This means there is a great demand for innovative joining methods or the adaptation of conventional technologies for the material mix of fibre composite and metallic base material.
2.2. Solution approach and technology concept

The research goal is to develop a novel high-strength mixed connection, flush on one side, for FRP components/semi-finished products in combination with metallic materials. The solution approach to be elicited is based on the basic principle of solid punch riveting. A significantly increased form-fitting of the rivet head should improve the load coupling to the upper fibre composite. To meet the desired flatness requirement without performing any additional preparatory operations, the space required by the connecting element in the fibre composite must be created during the joining process. The scientific challenge lies in the development of the setting process involving superposition and synchronization with the rotary movement of the riveting punch. The punching out of the typical slug and the stamping of the annular groove on the metallic underside of the connection remain part of the process chain for solid punch riveting [see Figure 2].

Increased friction at the underside of the head during insertion/forming causes the base polymer to heat up and facilitates the displacement of the material matrix. Depending on the sinking behaviour, the rivet head geometry will then be optimized in terms of material displacement. In addition, there is the assumption that the plastification in the contact zone inhibits the tendency of the fibre to corrode. Studies of corrosion resistance, including with special surface coatings, are also being undertaken to characterize the connection properties.

The research project should therefore include, in addition to the basic investigation and the development of riveting technology, experimental analysis and the determination of characteristics. The focus is on material combinations of steel, aluminium and magnesium with thermoplastic GRP and CFRP. However, the required joining equipment will not only be designed for a special riveting technology.

2.3 System development and test station

Standard commercially available devices for rivet setting may not fully reflect the exacting movement sequences of the “new technology”. In addition to the freely synchronizable rotary motion for the actual translation of the riveting punch, a coaxially arranged clamping device is required. The joint partners have to be pre-tensioned with appropriate force against each other over the underlying die before the setting operation. Thus, three independent degrees of freedom are necessary (see Figure 3), while normal joining systems are usually equipped with only one or one and a half.

Since composite construction uses a mixture of various conventional joining and machining technology, it also makes sense to develop a flexible, versatile and above all robotic joining system [3]. Often, operators are forced to select a joining method based on commercially available and specially adapted riveting technology. A multi-functional tool would cover a greater range and diversity of technology and also reduce investment costs and thus be especially suitable for small-and medium-sized businesses as well as for prototyping.

Due to the great process forces in the manufacture of mechanical joints, the system must be equipped with a C-frame designed for the load. As a class, similar systems are also designated as clinch machines. The difficulty and therefore the challenge for the development of the system lie in a highly compact design and an advantageous kinematic chain. The safe working loads of eligible robots must not be exceeded. Equally, good accessibility to the joints is to be ensured. Another focus is the control technology, which must permit the synchronization of movements and ease of operation of the clinch machine.

3. Development of the riveting technology

3.1 Basic investigations

After extensive research of the prior state-of-the-art in the field of point joints in multi-material design (MMD), the boundary conditions for the riveting technology to be developed were identified [4][7]:

- single-cut overlap joining
- Arrangement with FRP on the punch side, sheet metal on the die side
- accessibility on both sides for the joint tools (punch and die)
- one-sided flatness of the joint
- no pilot-hole operation

Experimental studies on conventional solid punch riveted joints initially looked into different material combinations (GRP–Al, GRP-St, CFRP–Al, CFRP-St) to determine their joining properties. This was to serve as a reference for the newly developed riveting technology. In the transverse section of the joints (see Fig. 1) it can be seen that the punching of the rivet through the FRP results in a severe disturbance in the immediately surrounding fibre directions. The delaminated area here is greater than the rivet head diameter.

In the static shear and head-pulling test the joints fail, as would be expected, by the rivet head being pulled out from the FRP. In comparison to the base material strength, due to the
fibre layer in the FRP and therefore dependent on direction relative to impact, only weak joint strengths can be achieved.

From the comparison of experimental and theoretical maximum tensile strength of the solid punched rivet under examination (see Figure 4), it can be seen, firstly, that the preparation of such joints is possible in principle. However, these are unsuitable for use in multi-material design, since the transferable strengths are lower here than in pure St–St or Al–Al joints.

![Figure 4: Maximum tensile strengths (determined experimentally and theoretically using a hole-broadening cross section) of solid punch riveted joints in different material combinations](image)

3.2 Optimization of riveting technology

Due to the pronounced relaxation behaviour of the FRP at point joints in mixed materials, no friction lock can be generated by clamping forces, only a form fit between the rivet element and the base material. The hole-broadening resistance of joints rises with increasing rivet head diameter. The essential approach for riveting technology was derived from this relationship and the investigations of the solid punch riveted joints. The novel joining element is therefore characterised by a much enlarged countersunk rivet head diameter in comparison to a conventional solid punch rivet. The introduction of the rivet head is achieved by a superimposed feeding and rotary motion. The rotary motion is meant to cause the plasticization of thermoplastic matrix of the FRP by frictional heating, and thus facilitate the moulding of the rivet head. The formation of a plastic film between the rivet element and the reinforcing fibres may cause electrical insulation, thus improving the corrosion properties.

![Figure 5: Schematic solution approach to the development of the new riveting technology](image)

The rivet geometry was designed based on the conventional solid punch rivet with respect to rivet diameter and rivet lengths. The countersunk head geometry was developed using finite element method and optimized to improve the joint strength. The distribution of equivalent stress in the longitudinal specimen axis under shear-stress testing, presented in Figure 6, shows that for the optimized geometry, the areas of greatest tension no longer occur in the FRP, but in the area of the rivet shank and the die-side sheet metal.

To transfer the rotary motion from the punch to the rivet, a hexagonal geometry was initially selected on the basis of a decision matrix.

![Figure 6: FEM simulation of the stress distribution [MPa] in the joint under tensile stress](image)

3.3 Experimental investigation

In addition to the numerical method, the rivet geometry was optimized by experimental investigations. These consisted primarily of producing and evaluating cross-sectional images. The formation of the undercut by the shaping of the metallic base material into the shank of the rivet and the structure of the FRP can be characterized solely by metallographic sections.

To determine the achievable joint strengths, static shear and head-pulling tests were performed on samples made according to standard DVS 3480 [3]. For GRP–Al joints, the tolerable head-pulling strength can be increased with the new technology by about 20% compared to the conventional solid punch riveting. Shear tests revealed 60-90% higher tolerable strengths (see Fig. 7 and Figure 8).

![Figure 7: Cross-section of the new rivet joint](image)

![Figure 8: Experimentally determined maximum head-pulling and shear tests of conventional and novel rivet joints](image)
Fraunhofer IWU is currently studying the fatigue strength of the new joining technology. Determination of the corrosion behaviour by the VDA alternating test is planned, whereby in addition to stainless steel rivets, rivets with metallic coatings (ZnNi, ALMAC) will be used.

Since multi-material design often uses point mechanical joining technologies in conjunction with adhesives (the joining points only transfer stresses until the adhesive has cured) to pursue investigations into the displacement behaviour of the adhesive during insertion of the rivet.

4. Development of the riveting device
4.1 Starting conditions and requirements

For the development of the riveter and its kinematics, the constraints and requirements have been first set out in a specification, considering general knowledge and analogies from existing systems. Since this is to be a truly flexible experimental tool, the drive parameters must be generously proportioned. A final parameter optimization will be applicable only after experimental investigations of the technological requirements.

Conventional riveting punch forces are in a range of up to 50 kN, and this value is therefore taken as a guide. In order to cope with a variety of different technologies, the feeding of the punch must be without backlash in either direction. The coaxial and freely movable clamping device should meet the same requirements in terms of generating forces. The travel of the two concentric linear axes should be 100 mm each. The thrust must be available over the entire operating range, since sheet thickness and joint position accessibility can vary greatly. In order to achieve the shortest possible cycle times for future industrial use, the travel speeds must be 20 mm/s under load and 100 mm/s without load. The drive parameters for turning the rivet cannot be predicted exactly. The maximum torque and the optimal rotary speed can only be estimated for the time being due to the wide range of materials. In order, nevertheless, to cover as large an application field as possible torques up to 100 Nm and engine speeds up to 5000 rpm are planned.

The energy supply for the drives should be exclusively electric. The riveting frame, called the C-frame of the riveter, must meet the requirement for small deformation at the lowest possible mass, because the load capacity of the handling robot is limited. In order to meet standard safety requirements, a force measuring system should be built in for both linear axes. In addition, the readings can be used for evaluating the technology via a suitable interface for data acquisition. The operation of the system, particularly with it being a test station, should be possible via a central control panel with two-hand operation and visualization.

4.2 Development of a load-appropriate C-frame

After defining the system requirements, the development of the C-clamp is performed simultaneously with that of the riveter kinematics.

The total static load of the riveting punch and clamping device is 100 kN, the riveter width and projection are based on conventional riveters at 200 x 250 mm. The mechanical robot interface is provided with a standard quick release coupling to facilitate riveter changes.

After creating the CAD model the first topology optimization for the anticipated maximum load case can be carried out. The design variants thus obtained are then compared with variants that are intuitively developed and designed to be appropriate for the forces involved. As a comparison criterion, the ratio of deformation to mass (ß) is assessed, as there are no known technical requirements for deformation, such as a maximum upward bending of the riveter, for example (see Figure 9). The base material for this aspect of the frame structure is steel, since its behaviour is characterized by isotropic material properties and a high modulus of elasticity, and thus small deformations.

![Variant 1](image1.jpg)  
ß = 19.7 µm/kg

![Variant 2](image2.jpg)  
ß = 18.3 µm/kg

![Variant 3](image3.jpg)  
ß = 28.2 µm/kg

![Variant 4](image4.jpg)  
ß = 28.2 µm/kg

Figure 9: Comparison of C-frame variants

Variant 2 was selected as the preferred alternative and design reference, and the adjustment to the riveting kinematics is carried out in the design implementation.

4.3 Development of the riveting kinematics

The principle of the spindle gear is in general the best suited to creating powerful and uniform linear motion. Through the use of special threaded roller screw drives [see Figure 10], the pitch can be kept very small [8]. This enables a very compact design with relatively high power transmission, because the number of load-bearing thread flanks is great. In addition, split spindle nuts at defined pre-tensioning ensure the required backlash-free reverse operation [9].

![Figure 10: Principle and design of threaded roller screw drive](image5.jpg)

The riveting kinematic chain is developed for different arrangement variants and subjectively evaluated. Using a cost-
benefit analysis in accordance with VDI [11] at the technical, geometric, and economic level, the variant shown below emerged [Figure 11].

Figure 11: Chosen riveting kinematics arrangement

Figure 12: Implementation of the design

The rivet feed and clamping device each use a threaded roller screw drive on the moving spindle principle. The spindle nuts are rotatably mounted in the kinematic housing and in each case is driven by an eccentrically arranged servo motor, offset by 90°, via a backlash-free synchronous belt (toothed belt). The radial support and anti-rotation protection of the linearly movable spindles is provided by corresponding sliding surfaces in the housing. So that the clamping device and riveting punch can be aligned coaxially, one spindle nut system is set above the other. The lower clamping device spindle is hollow for the riveting die to pass through. The threaded spindle of the riveting punch is also hollow, and this is where the backlash-free, adjustable mounting of the rotatable stamping shaft fits. At the upper end of the rotatable shaft, the torque for turning the rivet is delivered over the entire stamping motion using a spline. The associated, also eccentrically located rotary drive covers the entire working range, but not without additional translation. The solution approach provides a removable/replaceable belt drive for this, so that the preferred operating point can be individually set. High speeds or high torques are then possible with a corresponding ratio.

Based on the design approaches, a virtual prototype was designed from the rough draft [see Figure 12].

4.4 Measurement systems and safety

Before the virtual prototype can be converted into a functional model, it is important to develop an appropriate measurement and safety concept. Appropriate measurement and safety equipment are required for the imaging of programmed movements, and also for ensuring the safety of the operator and system.

All three servo motors are equipped with a multi-turn absolute encoder for detecting the position of the linear axes and the rivet rotation. The software-based collision avoidance is also assured with end position sensors. For monitoring forces in the riveting punch and clamping device, each have 4 strain gauge strips on the housing above and below the interface between the kinematics and the C-clip in the load path [see Figure 13].

During monitoring, the induced strain is converted into the related force via an experimentally determined calibration value. The signal from the strain gauges is simultaneously recorded for the implemented data collection, so the process forces can be used for studying the joining technology and determining characteristic values.

A protection circuit, just like the two-hand operation, is firmly integrated into the control panel of the test station [see Figure 14]. After verification of the technology in this separated system, the prototype riveter will be applied on the handling robot in a second development step and tested [see Figure 15].

Figure 13: Application of strain gages

Figure 14: Test station with control panel

Figure 15: Robot application
5. Summary

The paper shows the process of development and several practical approaches of a new joining technology system, aimed for application for fiber-reinforced – metal composites. One of the development targets has been flexibility, especially to use the system for different types of spot joining processes.

An outstanding criterion was the interaction of process parameters, shape of the joining element and design decisions of the device. The resulting compact solution enables the handling and application of a robot systems and further automation.

In the future, a complex simulation model should be developed to make it possible to map and thereby perform calculations on the use of riveting, including the related process parameters for complex application areas.

Acknowledgement

The work in the project is funded by the European Regional Development Fund (ERDF) and the Free State of Saxony.

References


Spatial alignment of joining partners without fixtures, based on component-inherent markings

J. Fleischer, J. Elser
wbk Institute of Production Science, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Abstract: Light-weight aluminium space frame structures are frequently used for small-volume products, such as sports cars. The assembly of these products has so far been mainly manual and requires the use of complex and expensive fixtures. To increase the profitability the research conducted at wbk Institute of Production Science is aiming to achieve an automated, fixtureless assembly of such structures by the use of industrial robots. To achieve the required accuracies regarding the alignment of the joining partners, a new approach based on component-inherent markings has been developed. This article describes the theoretical foundations of the required measurement approach.

Keywords: Assembly, Automation, Flexibility, Alignment

1. Introduction

The energy efficiency of machines, vehicles and systems can be increased if the dynamically moving masses are reduced. One possibility to reduce the weight would be to use light-weight space frame structures made from aluminium extrusion profiles. These structures have already found their way into the automotive and aerospace industries and are used frequently [1, 2] (Figure 1). At the same time, there is a clear tendency of the automotive industry to cater to individual customer requirements. As a result, the number of variants is increasing, more and more modular designs are used, and hence the overall number of parts is split between different products [3, 4].

![Figure 1: Examples for space frame structures](image)

The above-mentioned developments lead to new challenges for production and assembly processes. Assembly processes must be highly flexible to ensure that different structures can be joined to the same production facility. A large proportion of assembly processes is still manual because the production facilities lack flexibility [5]. Automated assembly without fixtures through the use of industrial robots and flexible gripper systems is one way to increase profitability, system flexibility and repeatability [5].

The joining partners must be aligned with a high level of accuracy in order to ensure that the joining process can be implemented and that the required tolerances are met. For laser welding, which is a method that is frequently used for the joining of frame structures, the maximum permitted gap width and thus, the required accuracy, is 0.2mm [6]. However, the accuracy of flexible manufacturing processes is limited compared to inflexible manufacturing processes in large-volume production. This leads to increased deviations of the actual component geometry from the target geometry. Also, there are inaccuracies that result from the fact that the absolute accuracy of the flexible handling systems (e.g. industrial robot) and the flexible gripper systems is limited. According to research conducted by [7], the absolute accuracy of non-calibrated standard six-arm robots coming straight from the factory varies between 0.4 - 1.2 mm depending on the respective pose. Calibration increases the accuracy to 0.2 - 0.4 mm. However all types of errors combined add up to deviations of a lot more than 1 mm at the joining spot. Therefore achievable alignment accuracy is not sufficient for the required joining processes.

For the components to be aligned correctly and assembled with industrial robots and without fixtures in spite of that, an additional system with a higher accuracy as mentioned above is required. With such a system it will be possible to determine at first the exact position and orientation of the components and their deviation from the nominal position and orientation. After that, the alignment of the components can be performed by moving the industrial robot in small steps relative to the current position of the robot instead of moving the industrial robot to a new absolute position and orientation. The smallest resolvable step width of an industrial robot hereby is about 0.01 mm.

2. Requirements and state-of-the-art

Systems for the three-dimensional measurement and fixtureless alignment of joining partners must meet the following requirements:

- For the assembly of space frame structures the position and orientation of several components must be measured and aligned in 6 (x, y, z, A, B, C) degrees of freedom.
- Both hardware and software must be useable for different components without requiring machine changeover or adjustments of the algorithms for the evaluation of the measurement data in order provide maximum flexibility.
- Measurement and alignment of 6 degrees of freedom (x, y, z, A, B, C) also for rotationally symmetrical components, such as tubes or balls.
- The alignment process must be error-proof and must allow joining partner alignment without force application if possible.
Approaches that would meet these or similar requirements for fixtureless assembly processes with industrial robots have already been defined. However, these approaches are generally limited to specific applications and offer only limited flexibility. 

[8, 9, 10, 11] represent some of the most important approaches. They use optical measuring systems to measure the spatial location and orientation of the component. In most cases the detection process for the components is based on the identification of characteristic features such as edges and bores. When the shape of the component is known, it is possible to define the spatial location and orientation of the component based on that. Rotationally symmetrical parts such as spheres and all profiles with a circular cross section do not have any characteristic features that could be measured accurately for a determination of the component orientation. Therefore, these methods are unsuitable for these kinds of components. Furthermore, the flexibility of the method is limited because the algorithms used for measurement data evaluation and sometimes also the hardware must be adjusted whenever a new and different component is handled.

[1, 12, 13] describe another approach for single component positioning by the example of aluminium extrusion profile machining. First, the geometric specifications of different three-dimensionally curved extrusion profiles are determined using a component-inherent scale. The scale consists in several, regularly spaced markings applied onto the component surface by a marking laser. After the components have been measured, their three-dimensional geometry and the exact position of the scale markings along the profile axis are known. Based on that, the location of the profile can be determined accurately in two degrees of freedom (position in longitudinal profile direction and rotation about this axis) in the following process steps by measuring the individual marking using a digital image processing system. The remaining degrees of freedom must then be defined using a fixture. The benefit of this method is the flexibility in detecting component location and orientation, because the algorithm for the evaluation of the markings is independent of the shape of the profile.

3. System description

In preparation for assembly and alignment of the joining partners, the joining partners - aluminium extrusion profiles in our example - are pre-positioned in relation to each other using one industrial robot each or similar flexible handling tools (Figure 2, a). First, a safety distance between the profiles is set in order to prevent them from colliding. The process for the accurate spatial alignment of the joining partners is the continued development of [1, 12, 13] and uses the markings applied to the component surface. One of the component-inherent markings of each of the joining partners close to the respective joining spot is scanned with an optical measuring system (Figure 2, b). This measuring system can be fixed locally right where the measurement takes place or attached to another industrial robot.

Based on the results, the location and orientation of the joining partners can be calculated by means of a mathematical description of the correlation between the position of the marking and the profile geometry (Figure 2, c). In the next step, the relative distance between the joining partners at the joining spot can be determined based on the measured and calculated data (Figure 2, d). If the deviations are not within the allowed tolerance, the position is adjusted by the industrial robot that holds the respective profile (Figure 2, e). Once the setup has been adjusted, the position is checked again and if required, the position of the joining partners relative to each other is further corrected until it is within the specified tolerance. This ensures a closed loop for the positioning of the joining partners. After the profiles have been arranged correctly, the components can be joined (Figure 2, f).

![Diagram](image)

**Figure 2:** Closed-loop system for three dimensional alignment of joining partners (6 DOF)

The elements for a successful implementation of the approach are the accurate mathematical description of the joining partners, the application, detection and evaluation of the component-inherent markings and the control algorithm for the displacement of the component with industrial robots.

4. Mathematical description of joining partners

The components are described by means of 3D coordinate systems in order to identify their location and orientation based on the location and orientation of the markings. These coordinate systems represent the decisive elements for joining partner alignment. These elements comprise the joining spot, the component-inherent markings and the gripper position of the respective handling units (Figure 3). If the location and orientation of these coordinate systems relative to each other are known, it is possible to describe the respective component accurately. The location and orientation of the coordinate systems relative to each other are described mathematically using displacement vectors and rotation matrices \( M \).

The way in which \( v \) and \( M \) are determined depends on the manufacturing process of the components. If the absolute accuracy and the repeat accuracy of component production and marking application are sufficient, \( v \) and \( M \) can be calculated based on the reference data. However, if these accuracies are insufficient, the elements must be measured first before \( v \) and \( M \) can be determined.

This means that the position of the marking on the component must be known accurately and that the marking must be designed in such a way that accurately describes a 3D coordinate system.
The joining partners are aligned by moving the component until the coordinate systems overlap on the joining spot.

The joining partners are aligned by moving the component until the coordinate systems overlap on the joining spot.

5. Design and detection of markings

The component-inherent markings are applied for the accurate definition of a 3D coordinate system. This requires the shape of the markings to theoretically have three clearly identifiable characteristic features which are not located on the same axis and whose spatial positions \((x, y, z)\) must be known. For the measurement of the spatial positions, a stereo camera system performing 3D measurements was selected by a systematic selection process.

Figure 4, a) shows a possible marking by the example of a flat surface. The \(x\) axis is determined by the mathematical connection of the centres of two circles. The \(y\) axis is determined by placing a vertical connecting line between the third point and the line connecting the first two points. Then, the \(z\) axis can be calculated based on the scalar product. The circle diameters, which differ in size, are used to determine the direction of the coordinate axes.

This marking only complies with the requirements described in 2. to a limited extent. Therefore, a marking consisting of five sections was defined on the basis of the basic marking consisting of three sections (Figure 4, a). This marking accurately describes a 3D coordinate system and is illustrated in Figure 4, b) and c) by the example of a linear tubular profile. The shape of the marking is designed in such a way that it can be used for different, multi-dimensionally curved tubular profiles.

For the interpretation of the marking and thus, for the calculation of the 3D coodination system described by the markings, five steps must be carried out (Figure 4, b):

1. Determination of the spatial location of the centre \((x, y, z)\) for all circles using the stereo camera system (C1-CS5).
2. Definition of the centre of circle of C1 as the point of origin of the 3D coordinate system.
3. Definition of the centre points C2 and C3. Definition of this direction as the \(x\) direction.
4. The \(y\) direction is theoretically formed by the straight line crossing centre points C4 and C5. Because of variance in the application and detection of the markings it cannot be ensured that the straight line will be perpendicular to the \(x\) direction. Figure 5 shows the resulting angle between the \(x\) direction and the C4-CS direction as a function of the probability distribution of the measurement values.

The markings will be positioned as close to the joining spots as possible with the distance that is required to ensure a reliable detection of the marking by the measuring system. This is necessary because angle errors, which can be caused by errors in the application of the marking or in the detection, for example, lead to translational errors that increase along with an increase in the distance between joining spot and marking. Figure 6 illustrates this correlation by the example of a linear profile.

For the interpretation of the marking and thus, for the calculation of the 3D coodination system described by the markings, five steps must be carried out (Figure 4, b):

1. Determination of the spatial location of the centre \((x, y, z)\) for all circles using the stereo camera system (C1-CS5).
2. Definition of the centre of circle of C1 as the point of origin of the 3D coordinate system.
marking coordinate system is independent of the distance between marking and joining spot.

6. Closed-loop control for alignment

The compensation movement adjusting the position of the joining partners is calculated using a control algorithm implemented in Matlab. The input data for this algorithm are the profile geometries, the positions of the profile elements that are relevant for the mathematical description of the joining partners along the longitudinal profile axes (joining spots, component-inherent markings and gripper positions) and the location and orientation - determined with the stereo camera system - of the component-inherent markings used for alignment within a superordinate world coordinate system.

First, displacement vectors \( \mathbf{v} \) and rotational matrices \( \mathbf{M} \), required for the description of the profiles using the 3D coordinate system, are calculated based on these data (cf. 4.). Based on the location and orientation of the component-inherent markings, the profiles/joining partners are virtually arranged in a 3D setting in the second step. Then, the deviation between the joining spots of the joining pair, which must be aligned relative to each other, is determined (Figure 1 c, d).

The joining partners must not be displaced too much in order to ensure that no forces are applied and that there are no collisions during the correction process. In addition to the safety distance maintained during the preliminary positioning process (cf. 3.), the compensation movement carried out for the more accurate positioning of the joining partners is broken down into several small steps. These steps are decreasing along with the decreasing distance to the end position. The step width is determined with a sine function (Figure 7, a).

Figure 7, b) illustrates the simulation of such a movement by the example of the movement of a 3D coordinate system which describes a joining spot. The starting position is the position that results from the defined safety distance set during the preliminary positioning process. In this example, the end position of the target coordinate system corresponds to the joining spot of the second joining partner, which will maintain its position. After every step of the compensation movement, the actual location and the actual orientation of the components are checked by measuring the component-inherent markings in order to identify any variance and any errors of the movements and compensate them. This ensures that the joining partners can be joined without the use of force.

6. Experiment

In order to validate the approach presented in this document there has been set up a first test stand that will examine the achievable accuracy when detecting the markings with a 3D image processing system. Initial tests reveal that an accuracy of approximately ±0.03 mm in the direction of \( x \), \( y \), and \( z \) as well as an accuracy of ±0.01° for the orientations \( A \), \( B \), \( C \) can be reached.

7. Conclusion

A novel approach for the spatial alignment of joining partners has been developed at wbk Institute of Production Science for the assembly of frame structures. The approach allows the alignment of joining partners without the use of fixtures. It is based on component-inherent markings. The approach entails the mathematical description of extrusion profiles on the basis of 3D coordinate systems, the design, detection and evaluation of component-inherent markings and an algorithm for the alignment of the component position.

The fact that component-inherent markings applied to the component surface are used represents an advantage because the location and orientation of rotationally symmetrical profiles, which are often used for frame structures, can be identified accurately.

Acknowledgement

This paper is based on investigations of Transregional Collaborative Research Center SFB/TR10, which is kindly supported by the German Research Foundation (DFG).

References

Gripper design for tolerance compensating assembly systems

F. Dietrich\textsuperscript{a}, J. Maaß\textsuperscript{b}, K. Kaiser\textsuperscript{a}, A. Raatz\textsuperscript{a}

\textsuperscript{a}TU Braunschweig, Institute of Machine Tools and Production Technology (IWF), Braunschweig, Germany
\textsuperscript{b}Skysails GmbH, Hamburg, Germany

Abstract: This article presents a new design concept for grippers used in assembly, which enlarges the range of parts while maintaining the reliability. Such grippers are of interest as soon as robots can accomplish assembly tasks in the presence of geometric uncertainties. The design task is approached by morphological analysis, where candidate combinations are elaborated at first. From the set of viable combinations, one concept is selected, designed in detail and realized as a lab demonstrator. The presented gripper features three jaws, where two of them are actuated by servo-motors via belt drives. The third jaw consists of an inflatable / evacuable rubber pillow filled with granulate, a feature which lets the jaw adapt to arbitrary shapes. Practical experiments showed that the pillow is a key feature, which gives the ability to grip many items of different shape and size reliably.

Keywords: Assembly Automation, Shape-Adaptive Gripper Design, Industrial Robot

1. Introduction
Advances in the fields of specification and execution of robotic tasks \cite{1, 2} advocate the usage of sensor-based compensation of geometric uncertainties in industrial assembly situations. For example, recent variants of the task frame formalism \cite{3, 4} in combination with force feedback master peg-in-hole problems, also when manufacturing tolerances are present \cite{5, 6}. This exemplary task, although a challenge for control, can be specified on the basis of a very small set of robot commands, provided that the task frame formalism is available \cite{7, 8}.

Once robots can accomplish tasks that involve geometric uncertainty, the need for exact gripping loses importance. More specifically, it is less important to have exact knowledge about the geometrical relation between gripper frame and the frame of the item to be gripped. This shift in paradigm has considerable impact on the design of industrial grippers. Previously, such grippers restricted the range of items to be gripped by shape and size heavily, in order to fix the item repeatably. Now, since exact fixture is of less importance in the context described above, grippers can be redesigned to cover a wider range of items.

What the present article investigates is the question of gripper designs that enlarge the range of parts, while conserving the reliability of gripping and having only few degrees of freedom to be actuated. Along this topic, the remainder of the article is organized as follows. The next section gives an introduction to related work in terms of grippers and design elements. The subsequent section discusses benchmarks for such grippers and proposes two assemblies as reference scenarios. Thereafter, morphological analysis is applied, and subsequently, one of these concepts is designed in detail and implemented in lab scale. Finally, this demonstrator is validated in experiments and its potential for the reference scenario is shown. The article closes with a conclusion and an outlook on upcoming work.

2. Mechanical Grippers with Form-Adaptive Elements
In industrial applications where mechanical grippers are used, two- or three-fingered grippers are presumably the most common types. This fact is reflected by suppliers’ catalogues which provide numerous solutions off-the-shelf, which are usually designed to maximize repeatability of the gripping position. For situations where repeatability is of minor interest, additional designs have appeared. Some of them contain flexible design elements; for example, \cite{9} offers such a form-adaptive gripper concept featuring two jaws with inflatable / evacuable chambers, filled with granulate. From the underlying idea to maximize the contact surface, another gripper evolved \cite{10}, using a matrix of rigid sliders that adapt to the shape of the item to be gripped. Other approaches increase the degrees of freedom to obtain the adaptability, accepting the drawback that additional actuators are required \cite{11}. Such concepts are beyond the scope of the present article, which restricts its focus on the class of grippers that have form-adaptive elements. The remainder of the article will show that enlargement of the range of items to be gripped is not necessarily achieved by adding degrees of freedom.

3. Proposal of Benchmark Scenarios
Many assembly tasks can be abstracted to generic peg-in-hole situations with additional constraints on the assembly trajectory. Presumably the most common constraints are those that limit orientations (e.g. gearwheel pairing, keyed joint on a shaft) or narrow the path of the assembly trajectory (e.g. screw). A second characteristic of importance to gripper design is the shape and size of the items to be gripped. Since the present article is interested especially in the range of parts that can be handled, the items to be gripped shall vary in size and shape.

This article proposes two benchmark scenarios for assembly systems. The first proposal is a planetary gearing, displayed in Fig. 1. There are two properties of interest for the comparison of grippers. A gear to be mounted has to be aligned to its particular shaft, and its teeth have to be aligned to the teeth of the other gears that will have contact in the assembled configuration. Moreover, the individual gears have very different sizes.

The second proposal, a pegging game as displayed in Fig. 2, features very different shapes of parts, ranging from round to rectangular, from flat disks to pin-shaped geometries, and from small to large. If it is desired to increase the level of difficulty furthermore, the parts can be fed to the assembly process in a magazine that involves clearance in its fixtures. This second scenario models industrial assembly in several aspects: First, the assembly sequence is unique; second, there are multiple joining
operations which only allow one single orientation, and third,
the key-shaped items require, in similarity to bayonet joints, the
two-step assembly that consists of the steps “insert” and “turn”.

Figure 1. Benchmark 1: planetary gearing.

Figure 2. Benchmark 2: a pegging game.

Figure 4. Scores assigned to pairs of features during morphological analysis.

4. Proposal for Gripper Design
Morphological Analysis

Morphological analysis was applied to the present design problem in order to identify and evaluate principles that suit the requirements stated above. During the analysis, 11 categories of design elements were defined. Iterating through each of these categories, multiple features were proposed. The following list states the categories and, in brackets, the number of features proposed per category:

1. Principle of gripping (3)
2. Principle of object enclosure (8)
3. Direction of gripping (2)
4. Number of gripping jaws (4)
5. Adaption to different sizes of object (4)
6. Adaption to different shape of object (2)
7. Reproducibility of object position in gripper (4)
8. Principle used to provide elasticity (4)
9. Location of elasticity (3)
10. Variability of stiffness (4)
11. Variability of damping (1)

In total, there were 39 features proposed, where each of them falls in one of the 11 categories stated above.

If the number of combinations would be small enough, manual evaluation of all of them could take place, but the fact that the list above yields theoretically 294,912 candidate combinations asks for another reduction of the complexity. Instead of evaluating combinations involving all categories, this can be done effectively by considering pairs of features from two categories at first. For each pair, a score is assigned, saying whether the pair is not feasible (0), feasible, but not recommendable (1), feasible (3), or independent (-). These scores can be inserted into a matrix, as displayed in fig 4. An algorithm may then use this matrix of preferable pairs to check candidate combinations for feasibility. In this way, whole branches of the tree of combinations can be removed due to one non-feasible pair of features, and the remaining candidates may be ordered according to their sum of scores. A piece of software has been developed that implements such an algorithm and that compiles the set of most interesting solutions. Since each individual solution has been annotated by a sketch prior to the execution, the user may browse the sets of sketches that correspond to the most viable combinations afterwards. These sets of sketches formed the basis for further, manual, selection and elaboration.

What challenges the designer in particular is the wide range of sizes and shapes that shall be gripped. For this purpose, the jaws have to travel long distances, while sufficient gripping forces must be maintained. For example, the concept of the 2-jaw belt gripper displayed in Fig. 5 offers such displacements while only requiring one single linear actuator. The limitation of such mechanisms lies in the transmission: the more the gripper closes, the nearer the transmission mechanism approaches its singularity, increasing the forces inside the mechanism and the item to be gripped. Vice versa, the more it opens, the lower the forces imposed on the item become. It is very likely that such a gripper is a rather poor compromise between constancy of gripping force, maximum span, actuation speed, linearity of movements, and tactility. One may be tempted to add more actuators to close such a gripper, but then other major drawbacks must be accepted. In summary, the design studies elaborated from the morphological analysis conclude in the following statements:

1. It is disadvantageous if stiffness of the gripper decreases, when jaws are wide open.
2. Jaws should be small in order to minimize the probability of collision.
3. Individual actuation of the jaws increases adaptability, but requires more design space.
4. The combination of frictional contact and form closure increases robustness of gripping.

**Figure 5.** Gripper concept using a belt or chain elements for enclosure.

**Design Proposal: the Hybrid Belt Gripper**

Besides other concepts, the morphological analysis judged grippers with three jaws very promising. Commonly, such designs feature rotational symmetry, where all three jaws are (at least approximately) symmetrical. Some variants additionally feature common actuation, where all jaws are attached to one single actuator. In contrast to such concepts, the present work proposes, due to the absence of the requirement for symmetry and centering, to use different designs of jaws. Figures 5, 6 and 9 show such a design. Two jaws are actuated individually by belt-drives to travel along linear guideways, and the third jaw is fixed. The mobile jaws are more pin-shaped whereas the fixed jaw exposes a plane to the object to be gripped.

When the mechanism is commanded to close, the actuated jaws move towards the fixed jaw. Fully closed, all jaws touch each other, being able to grip very small items. The angle in which the two jaws travel towards the fixed jaw is 60°, which results desirably in an equilateral triangle of contact points, if the item is gripped symmetrically. Figure 7 draws an abstraction of this issue. Notice the three green rectangle on top, representing the fixed jaw, and the green trajectories of the moving jaws, approaching from lower left / lower right (green lines). The symmetrical configuration is drawn in blue.

The instance in which the gripper picks up an item is crucial to the reliability [12, 13], and hence studied to obtain a design rule to choose the width of the fixed jaw. Commonly, industrial assembly tasks require that the gripper picks the items from a feeder or a magazine, where some clearance is present. Of particular interest is whether the gripper can pick the item reliably, also when the item is displaced from the ideal position inside the feeder. Such feeders are modeled by the reference scenarios in such a way that the items have to be picked up from moulds. In these moulds, there is clearance to the item, which is modeled in Fig. 7 as a translational offset from the centered position. More restrictions are applied to this model, so that only one degree of freedom and no rotations are accepted. Moreover, suppose for the following gripping analysis that the items are of circular or squared shape, as shown in Figs. 7 and 9. This figure draws the centered configuration in blue and the offset configuration in red. Now that the preliminaries have been set up, the criterion for the analysis is formulated: It is considered that the gripper can pick the item at a particular offset position, if both trajectories of the moving jaws touch the item on their way towards the closed configuration. The application to Fig 7 means that the green trajectories of the moving jaws intersect with the area of the item under investigation (circle / square). This analysis, although a strongly simplified representation, provides a valuable rule of thumb to choose the width of the fixed jaw without exaggerating the computational effort. In the figure, the offset $s$ is denoted, denoting the minimum width of one half of the jaw.

![Mounting Flange](image)

**Figure 6.** Three-jaw gripper featuring hybrid friction / form closure. Perspective from front side.

![Centered Configuration](image) ![Displaced Configuration](image)

**Figure 7.** Gripping analysis: robustness against tolerances in feed position. Blue: Ideally centred configuration, red: displaced configuration (Error model), see explanation in text.

According to the summary of the morphological analysis, the combination of frictional contact and form closure is advantageous for the gripping process. In the present design, frictional forces are maximized by dedicated coatings of the jaws, i.e. rubber. Maximizing form enclosure means maximization of the contact area between the item to be gripped and the jaws. This requirement is suited by a particular design feature in the fixed jaw. This jaw consists of an inflatable / evacuable rubber pillow filled with granulate, i.e. sand. An item to be gripped is pushed into the non-evacuated pillow by the actuated jaws. This action forces the pillow to deform and to adapt to the shape of the item, and hence maximizes the surface which is in contact with the item. When vacuum is
applied to the pillow, the granulate inside is consolidated and the geometrical shape of the pillow is frozen.

Figure 8. Fixed jaw with inflatable / evacuable rubber pillow.

Experiments: gripping of exemplary pieces from scenario 2.

Figure 9.

5. Experiments and Lessons Learned

Basic experiments with the gripper were carried out. A test bed was developed, carrying the gripper itself, the control electronics, and power supply. The experiments involved the items of the two benchmark scenarios. All of these parts were held by the gripper fixed enough to perform the assembly sequence. Figure 9 exemplary gives some impressions of these experiments. Practical experience showed that the evacuable pillow adds a lot of robustness to the gripping process. Especially, when the pillow is evacuated to harden it, removal of a gripped item required much more force than before. In summary, the pillow works as desired, and is definitely a key feature in gripper design, when adaptability to multiple shapes is required.

6. Conclusion

This article covers design aspects of grippers which are used for assembly in the presence of geometric uncertainty. Benchmarks for such grippers are proposed and discussed. Also, this article elaborates categories for morphological analysis of grippers that can adapt to multiple shapes and sizes of object. On this basis, a design for such a gripper is detailed and its lab implementation is presented. Results from basic experiments are reported, confirming the potential of shape-adaptive design elements in grippers.

The scientific contribution of this article is in the field of gripper design for production automation. The advances proposed in this article are based on the fact that, once robots can compensate tolerances, the design of grippers receives new degrees of freedom. Exemplarily, the article proposes to use flexible elements to enhance the range of items to be gripped.

Ongoing work enhances the gripper’s functionality by a tactile sensing algorithm, similar to [14], so that objects with impedance and fragile objects can be handled safely. Also, plans are made to integrate the gripper into the HEXA robot, a parallel kinematic demonstrator developed and realized during the Collaborative Research Center 562 (DFG-SFB 562).

References
A cladistics approach to classification of joining and fastening methods

A. Ziout, A. Azab
Intelligent Manufacturing systems centre, Windsor, Ontario.

Abstract: Joining and fastening methods have evolved over many centuries. This evolution is driven by many factors like product functionality, increased parts count, and emergence of new materials. This paper aims at classifying past and recent joining/fastening methods in a manner to allow for extraction and study of the development and evolution of assembly methods and mating features. Cladistics was used to plot the relationship and co-evolution between assembly features on one hand and corresponding joining/fastening methods on another. This, in turn, is used in this study to predict the future development of joining/fastening methods.

Keyword: Disassembly, Joining/fastening methods, Classification.

1. Introduction

Industry is facing many challenges. This includes common ones; such as quality, competitiveness, customer satisfaction, and economical viability, and also non-traditional challenges, whether it is environmental, societal, or rapid technological changes. Inflation and emergence of new geographical economies placed industry in more challenging situation.

Manufacturing has responded to these challenges in many different ways; Green manufacturing, changeable and reconfigurable manufacturing systems, and sustainable product design are few alternate means. Yet more ways need to be discovered and followed.

Assembly is a major process in product manufacturing and realization; improving assembly process will improve manufacturing sector’s ability to adapt to previously mentioned challenges. Improvement in assembly process can lead up to 42% in labour cost, 54% reduction in part count, 60% reduction in assembly time, and 45% reduction in product development cycle time. [1]

Within the assembly process, selection and use of proper joining and fastening method is a crucial decision that involves the overall assembly process improvements. Hence, joining and fastening methods have been recently receiving more attention in academia and industry alike. Aerospace and marine industries have been among the pioneers in developing new joining and fastening technologies. Vast research is being done in design and development of fasteners and joining methods-for comprehensive review see Askinder’s [2]. Design for assembly, disassembly, service, and recycling are examples of different design methods that largely revolve around fastening and joining encountered in product design. For example, snap-fit joints have been developed to facilitate assembly process, whereas fasteners made of shape memory alloys have been more advantageous when it comes to disassembly.

Joining methods cover a wide spectrum, and they include different working principles; many researches attempted to classify joining methods. These attempts are driven by different purposes. Some classifications are intended to help designers and users to choose the right fasteners or joining method along with the right parameters; these can be found mostly in design manuals or handbooks. Other classifications focused on the working principles applied in joining methods; apparently, the purposes of these classifications are better understanding of the joining methods, and to keep the door open for development of new methods and technologies based on existing and potentially new working principles. This paper reviewed both types of classifications- see section 2.

Review of joining methods classifications is useful. It helps examine the gradual development of joining methods over time. This leads in turn to the fact that structure-based joining methods are well developed, while material-based joining methods are under development and not very robust yet.

This work presents a new classification approach for fastening and joining methods, which borrows from Biology and utilizes their well-established taxonomy method “Cladistics”. Use of cladistics for classifying joining methods is a novel approach that has not been explored yet, to the authors’ knowledge. Also cladistics are used to study the mutual effect “co-evolution” between joined parts and the corresponding used joining methods, and to suggest directions for future development.

2. Classification of joining methods in literature

Literature surveyed during this research showed that there are two major categories of classification of joining methods. First category of classification is based on the final shape of joint, or process used to create the joint. The main purpose behind this type of classification is to help designers select the optimal joint for their purposes. Handbooks and design manuals follow this type of classification. The second category is classification based on working principles applied in joining. This category is more used in academic research and books where focus is on innovation and development.

2.1. Classification based on the process, material, or application involved.

Due to the wide range of joining and fastening methods, it is not practical to have one classification criterion covering the whole range. Messler used both joint shape and manufacturing process of joint as criteria for classification. Based on this, three families of joining methods are identified [3]:

1. Mechanical joints: it includes: (a) threaded and unthreaded mechanical fasteners such as screws, bolts, studs, dowels, and pins; (b) integral attachments, such as rigid mechanical interlocks, elastic snap-fit, and plastic deformation in part or fasteners to achieve joining (rivets); (c) others, such as
stapling, stitching, sewing, ropes, knots, wraps, coupling, clutches, and magnetic connections.

2. Adhesive bonding and cementing: It is a joining process where two parts (adherents) are joined using bonding material (adhesive). Adhesive classification is very complicated due to the fact that no single criterion can satisfy all purposes of classification. Adhesives are classified based on the following criteria:
   b. Chemical composition.
   c. Physical form.
   d. Curing or setting mechanism.
   e. Specific adherent or application.

3. Welding: is a joining process where two materials are joined at atomic level through excessive use of heat. Heat source could be viewed as the main source to distinguish among the different types of welding. Brazing and soldering are two subsets of welding, except that they occur at a temperature lower than that in welding.

   This classification signalled the development of joining methods over time. Mechanical joining is the oldest method of joining, adhesives were known later; natural adhesives and mix of clay and straw were known in very old ages. Lately welding methods were known due to availability of many technologies. MESSLER proposed another approach of classification which is based on joint’s material. Five categories were identified [3]:
   1. Joining of metals, alloys, and inter-metallics
   2. Joining of ceramics and glasses
   3. Joining of polymers
   4. Joining of composite materials and structures
   5. Joining of living tissues.

   The main purpose of this classification is to facilitate the selection process of right joining method for specific material or application. This classification shows the importance of material properties of joined parts; physical and chemical properties are main factors for selection of proper joining method. It is found that the same joining method is used in different materials, this lead to the fact that different materials with similar key physical properties can use same joining method.

   Parmley classified joining methods based on three criteria: type of fastening component, type of joining process, and type of application [4]. Under fastening, the following joining methods are identified: threaded fasteners, pins, retaining rings, locking component, industrial rivets, wire rope and cables, and shafts and couplings. Under joining process, three joining methods are identified; welding, adhesive bonding, and injected metal assembly. Finally, under application, the following are identified: Concrete fastening, lumber connection, structural steel connections, electrical connection, aerospace fastening, sheet metal connections, and seals and packaging.

   Though, this classification is comprehensive enough, it doesn’t help in tracing joining methods development. It uses many classification criteria and hence, it was hard to have clear cut between different joining methods; i.e., same joining method could be found in many categories. Main benefit of this classification was ease of selection of suitable joining method.

   A common factor between mentioned classification methods is all of them served one goal, that is “ease of selection”. These classifications are developed to facilitate designer’s job. For example, designer can choose a set of candidate joining methods by knowing the application or material of joined parts. These classification methods does not clearly contribute to proper understanding of development of joining methods, they does not show interrelationships between joining methods, such as methods’ capabilities and characteristics. They also do not refer to working principle behind each joining method. Next section surveyed classification methods which classified joining methods based on working principle behind each method.

2.2. Classification based on working principle

   A proper way of studying development or “evolution” of joining methods is through studying the working principle of these methods. Pahl and BIEZT classified joining methods based on three working principles [5]:

   1. Material connections: the working principle is to use molecular and adhesive forces to join two or more parts of same or different materials. Welding, brazing, soldering, and adhesives are examples of methods that use such working principle.

   2. Form connections: the working principle is based on normal forces between mating surfaces due to applied pressure. Joining methods that apply this principle can also deliver functions other than joining; it can also seal, insulate, or transmit forces or motion. Wedges, bolts, pins, rivets, shaft-hub, locating elements, snapping, and clamping are examples of joining methods that use form connection as working principle.

   3. Force connections: three types of forces form the working principle of these joining methods; friction forces, elastic forces, and field forces. Interference joints and bolted joints are examples of joints that use friction forces to realized connection between parts. Field forces such as electrical or magnetic forces can be used to join parts together.

   This classification gives clear directions for exploring new avenues for further development and innovations in joining methods. Exploring concepts that use atomic or molecular bonding, innovative shapes “form”, and potential field forces could lead to development of new joining methods.

   In attempt to classify and design joining methods considering assembly and disassembly requirements Sonnenberg defined the following working principles for studying and potentially classifying joining methods [6]: (1) adhesion, (2) cohesion, (3) negative pressure, (4) magnetism, and (5) friction. The working principle adhesion was divided into more three sub categories, Chemical adhesion, as soldering, micro structural mechanical adhesion, as gluing, and macro structural mechanical adhesion, as snap fit.

   A primitive classification of joining methods was proposed by Kaplan and Brandon. All joining methods were classified under three major categories [8]:

   1. Mechanical methods: two working principles were identified; point attachment, and line attachment. Pins, screws, or bolts are examples of point attachment. Folding a sheet metal along its edge is an example of line attachment joining method.
2. Chemical joining methods: the working principle depends on joining two materials using bonds result from chemical reaction. Adhesives use reaction between resins and hardeners to achieve the required bonding forces. Brazing and soldering use fluxing agents to achieve proper joining.

3. Physical bonding: working principle of physical joining depends on phase transition from liquid to solid state, or diffusion process where atoms move within solid solution. Solvent-based adhesives, welding, brazing, soldering are example of joining methods that use physical bonding.

Quite clearly, this classification is flawed in logic, since there are many joining methods that use both mechanical and physical bonding at the same time, chemical and physical bonding also exist simultaneously in adhesive joining method. This affected the usefulness of this classification.

Review of related literature showed that there were many classifications for joining methods, this was because each classification served different purpose. Next section showed how different purposes could lead to different classification of same items.

2.3. Purpose of classification

In general, humans practice classification in their daily life to learn things, and to know relationships within their surroundings. Classification helped humans to differentiate between birds and animals as groups, and between different birds at the same group. Process of classification evolves to allow deeper and deeper differentiation between things based on differences and/ or similarities.

Three major purposes of a classification were identified by Messler [7]. First purpose was to facilitate learning and recall of information and knowledge, second to see familial relationships between things, and lastly to build a taxonomy or arrangements of related things.

In particular, classification of joining methods is expected to serve one or more of the following purposes:

1. To have overall taxonomy of joining methods.
2. To position certain joining method within the overall joining method taxonomy.
3. To identify, recognize, and use the key attributes or characteristics of a joining method.
4. To help understand principle of operation, performance, and potential applications.

The contribution of this paper is to provide a new classification of joining method to serve a new purpose, which is to study the development “evolution” of joining methods, and to identify evolving features and characteristics of joining methods. Review of old and current joining methods shows that there are mutual interaction between joining methods and joined parts; development of any of them affects the other. This fact is used here to study the co-development “co-evolution” between joining methods and joined parts. To achieve this, a classification methodology called cladistics is used. Cladistics is a common classification methodology in biology. Its use in engineering is relatively new and limited. Cladistics was used by ElMaraghy et.al. to study the evolution of manufactured products and their manufacturing systems [9]. Section 3 gives more explanation on cladistics and its implementation in this research.

3. Cladistics

3.1. Definition

Cladistics is a method of classification that groups taxa hierarchically into discrete set and subsets [10]. It can be used to categorize any type of data, but its application is well known in the field of biology as “phylogenetic systematic”.

The output of cladistics is tree-like diagram called cladogram, which shows the hierarchical relationships between classified items or taxa. Taxa are the set of items or entities need to be classified into sets or subsets. Cladistics classifies taxa based on their characteristics. Character is the attribute or feature inherited in taxon that makes it similar or different from other taxon.

There are different kinds of characters:

1. Qualitative and quantitative characters: characters of qualitative values are preferred in cladistics over ones of quantitative values
2. Continuous and discrete characters: cladistics deal with discrete values, so for continuous characters discretization is needed.
3. Overlapping and non overlapping characters: non overlapping is preferred over overlapping character, cut-off criteria are needed to exclude overlapping data. Cladistics’ preferences for characters are followed in this research, meaning qualitative, discrete, and non overlapping values are used, necessary modification is done to cope with these preferences.

3.2. Cladogram construction

Cladogram construction starts with selecting taxa (variants or items) to be classified; then characters which distinguish relationships between taxa and encoding follows. Characters coding means assigning values to each character. Kitching et.al. mentioned five possible coding methods [10]; the method which consider characters as independent and multi state variable is followed in this research. It is worth mentioning that this method is used, since it is also used by the software selected to construct the cladograms. There are many software available online to construct cladograms. These software use different algorithms. Irrespective of what algorithm is used, all software is looking for the most parsimonious cladogram, a cladogram with the minimum steps to represent relationships between studied taxa.

4. Methodology

4.1. Scope of the study

The scope of this study is the co-evolution between joined parts and their corresponding joining methods. Co-evolution is borrowed from biology, where it signifies the gradual change in characters of two dependent biological systems over time bringing upon a new system. In this study co-evolution is synonymous to co-development which is gradual mutual/related development and changes occurred in characters of both the joining methods and joined components parts. Different joined parts (joints) were randomly selected, studied and analysed. Key features were determined and were coded, i.e., were given values. Each joint was linked to joining method used in realizing
the joint. Key capabilities for joining methods were identified and coded.

This study is limited to joints made of similar materials, meaning, the mating parts are of the same material. This constraint is made to keep analysis simple and to reduce problem complexity. Consequently, results of this study are limited to this type of joints.

4.2. Software used in the study

Since the objective of this study is not how to establish a cladogram, online software is chosen to do this task. The software is called TNT “tree analyses using new technology”. This program was designed by Globoff et.al.[11], and was made freely available with subsidy from Willi Hennig Society [11]. See acknowledgement.

4.3. Joining methods’ characters and characters states

To construct a cladogram that shows the familial relationship between joining methods, their characters need to be identified. Character is a basic component in constructing cladograms. Joining methods’ characters or capabilities are selected carefully. Many factors are considered in selecting these capabilities; force, environmental, and economical requirements are examples of such requirements. Five capabilities are identified, see table 1, they are:

1. Load carrying capacity: type and magnitude of load that can be taken by a joining method is considered as key character. Some joining methods are good for compression loading but not for tension. To overcome the continuous nature of this character, three discrete values are assigned; high, medium, or low.
2. Motion: some joining methods produce fixed joints, others provide motion in one or multi directions.
3. Assembly: technical and economical factors affect assembly capability of joining methods, this capability is coded into three levels; high, medium, or low.
4. Disassembly: this capability is important to serve purposes like service, maintenance and recycling. This capability is coded similar to assembly.
5. Weight adding: many applications require the joining method not to add weight to the structure. Joining methods use filler materials or fasteners add considerable weight to the structure, others do not. Table1 has all key capabilities of joining methods with possible values of each one. These values are coded in numerical format.

4.4. Joined parts (joint) characters and characters states

Constructing a cladogram for joined parts requires the same as for joining methods. Eight characters (features) are identified and coded with the appropriate values. Table 2 lists these features and its values.

Table 2: Joined parts’ features and its states

<table>
<thead>
<tr>
<th>Capability (character)</th>
<th>Value (State)</th>
<th>Description*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Strength (0) = low (1) = medium (2) = high</td>
<td>Below 100 Mpa</td>
<td>between 100-500 Mpa Higher than 500 Mpa</td>
</tr>
<tr>
<td>2. fusion point (0) = low (1) = medium (2) = high</td>
<td>Below 300 °C</td>
<td>Between 300 - 900 °C Higher than 900 °C</td>
</tr>
<tr>
<td>3. plasticity (0) = low (1) = medium (2) = high</td>
<td>Yield strength below 100 Mpa</td>
<td>Yield strength between 100-500 Mpa Higher than 500 Mpa</td>
</tr>
<tr>
<td>4. effective thickness (0) = thin (1) = thick</td>
<td>Sheet thickness</td>
<td>Plate thickness</td>
</tr>
<tr>
<td>5. effective area (0) = small (1) = large</td>
<td>Relative to structure surface area</td>
<td>Relative to structure surface area</td>
</tr>
<tr>
<td>6. type of loading (0) = compr. (1) = tension (2) = shear</td>
<td>Type of load applied at the joint</td>
<td>Type of load applied at the joint Type of load applied at the joint</td>
</tr>
<tr>
<td>7. working environment (0) = neutral (1) = harsh</td>
<td>No harmful working conditions</td>
<td>Harmful working conditions exist</td>
</tr>
<tr>
<td>8. sealing/insulation (0) = yes (1) = no</td>
<td>Sealing/insulation is wanted</td>
<td>Sealing/insulation is not necessary</td>
</tr>
</tbody>
</table>

*: based on information found in [12]

4.5. Cladograms construction and mapping

The logic and sequence showed in figure 1 is the logic followed in this paper to construct required cladograms, and to extract required knowledge.

5. Case study

To demonstrate the implementation and usefulness of proposed approach, a case study is analysed using this approach. Nine randomly selected products are chosen for this case study. Products are normal products that can be found in daily life uses. It ranges from wooden coffee table to air manifold found in care engine- see Appendix. From each product one joint is selected for analysis. The only constraint in selecting joints is that the joint has to be made of similar material, i.e., joined parts had to be made of same material. Table 3 lists the selected nine joints with values “states” of each feature “character” according to value assignments found in table 2.
According to the software results, 2207 rearrangements were fed to the software to get the most parsimonious cladograms. Files according to the format accepted by TNT software. Then it minimum number of change in characters state is followed. Based on data presented in table 4, where the concept of \{2110010\}, \{2111110\}. Cladogram in figure 3 is constructed two changes took place - the thickness and the effective area.

For all joints found in table 3, its corresponding joining method is identified and analysed as shown in table 4.

Table 4: Analysis of selected joining methods

<table>
<thead>
<tr>
<th>Joining method</th>
<th>Load carrying</th>
<th>Motion</th>
<th>Assemblability</th>
<th>Disassemblability</th>
<th>Weight adding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threaded fasteners</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Synthetic rubber</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Rigid Mechanical interlocking</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Elastic Mechanical interlocking</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Elastic Mechanical interlocking</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Vibration welding</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Soldering</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Adhesive Mortar</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Arc welding</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Characters in table 3 are used in construction of the cladogram (figure 2). The process starts with generating a large number of possible cladograms. The cladogram with minimum number of changes in characters’ states is chosen. For example the last branch in figure 2 contains Copper_Pipe and Stainless_S_Tank. This is due to the fact that this is the minimal change in characters states between the two joined parts. Only two changes took place - the thickness and the effective area. (2110010), (2111110). Cladogram in figure 3 is constructed based on data presented in table 4, where the concept of minimum number of change in characters state is followed.

Data in table 3 and 4 is transferred into two separate input files according to the format accepted by TNT software. Then it is fed to the software to get the most parsimonious cladograms. According to the software results; 2207 rearrangements were examined to find most parsimonious cladogram for joined parts, and 3670 rearrangements for joining methods’ cladogram.

6. Results and discussion

6.1. Inferences from joined parts’ cladogram

The cladogram in figure 2 shows the assumed evolutionary trend of the nine studied joints, which was the result of the analysis of most evolved features on the joints’ part. The cladogram classified wood_coffee_table, concrete_brick, and glass_mirror as the least evolved parts. This can be justified due to their material with its primitive features.

Polypropylene_cup, Polyseter_sink, and Nylon_manifold are classified to be very close together, and that is also due to their material properties which significantly affect their ability to have evolvable features. Metals; Stainless_s_cup, and Copper_pipe were pushed to the end of the tree, metals are versatile materials which have excellent properties that make metals viable material with good evolvability.

Another inference can be extracted from joined parts cladogram is that the cladogram placed old known materials at the beginning (less evolved) - wood, concrete, glass, rattan, and pushed newly known materials (more evolved) to the ends of the tree. See figure 2.

Another inference can be extracted from joined parts cladogram is that the cladogram placed old known materials at the beginning (less evolved) - wood, concrete, glass, rattan, and pushed newly known materials (more evolved) to the ends of the tree. See figure 2.

6.2. Inferences from joining methods’ cladogram

Figure 3 represents the classification tree of the studied joining methods. To value the resulted classification, it is important to point out that this classification is not based on one or two criteria as the case in traditional classification (see section 2). This cladogram is based on the five capabilities and their state values as in table 4, the classification results from examining the change in characters state of joining methods, so the classification results from this approach is expected to classify and group methods based on their ability to evolve.

At node 9 in figure 3, fasteners appeared in a separate branch isolating themselves from all other joining methods; this means that fasteners are well-established and have less chance to evolve in the future. At node 12 in same figure, joining methods are divided into two major branches; first branch is mechanical methods which depend on the form of used joint to achieve joining, second branch is joining methods that depend on material to achieve joining. So other than fasteners, joining methods can be either form-based methods, or material-based methods. According to their location in the cladogram, material-based methods have higher ability to evolve and develop in the future. They are at the far end of the evolution tree, see fig. 3.
6.3. Inferences from mapping two cladograms

Considering the direction of evolution in both cladograms, it was found there is a significant match between them, which proves the authors’ assumption/intuition claimed earlier in this paper about the co-development between joined parts and corresponding joining methods. Figure 4 demonstrates this co-development, joined parts in red group matched with joining method with the same color, the same applies for green and blue groups.

Figure 3: Most parsimonious cladogram of joining methods

Results from both cladograms showed the future trends and development on the joining methods side would be applying material-based joining, supported with developments in metallic materials and new polymers.

Figure 4: co-development between joined parts and joining methods.

6.4. Impact on assembly/disassembly technologies and systems

One of the findings of this study is the use of the mimicry from the sciences of Biology to help plot a road map for new concepts in joining technology. It is believed it conveys clearly the direction of potential areas where innovative joining methods and materials could be used. The study proves that material based joining methods have potential for development higher than form based methods. It is also indicative that metals and plastics are sought to be materials with more potential for improvement of the joining/assembly features.

6. Conclusions

A novel approach for classification of joining methods was developed. Cladistics was used for the first time to classify existing joining methods and to assess directions of future development. Unlike current classification methods that consider few criteria for classifying joining methods, cladistics consider a group of characters along with change in their states to establish a classification for entities possess these characters. The objective of cladistics is to find familial relationships between classified entities, and to show their path of evolution.

With the help of case study, new classification for joining methods was developed, also directions of evolution were identified. Co-development between joining methods and joined parts were studied, matching between joining methods cladogram and joined parts cladograms proved the existence of co-development between them.

This study was limited to joints of similar materials. For future work, joints of dissimilar materials can be studied. Also the significance of this work can be improved by considering larger number of joints, and consequently larger number of joining methods.

7. Acknowledgment

The authors wish to express their gratitude to Willi Hennig society who made TNT Software freely available, which saved the authors a lot of effort. Special thanks to Prof ElMaraghy W. for his invaluable support and supervision. Also, we would like to thank Dr ElGeddawy T. for sharing with us his expertise in manufacturing system co-evolution.

References


# Appendix A: List of products used in the case study

<table>
<thead>
<tr>
<th>Product</th>
<th>Description and source</th>
<th>Product</th>
<th>Description and source</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Wooden coffee table" /></td>
<td>Wooden coffee table joined by fasteners. Source: <a href="http://www.ikea.com">http://www.ikea.com</a></td>
<td><img src="image" alt="Copper water pipes" /></td>
<td>Copper water pipes joined together. Source: <a href="http://www.canadiantire.ca/">http://www.canadiantire.ca/</a></td>
</tr>
<tr>
<td><img src="image" alt="polyester sink" /></td>
<td>polyester sink joined to the wall. Source: <a href="http://www.ikea.com">http://www.ikea.com</a></td>
<td><img src="image" alt="Concrete Brick" /></td>
<td>Concrete Brick joined together. Source: <a href="http://www.homedepot.com/Building-Materials-Concrete-Cement-Masonry-Concrete-Blocks-Bricks-Lintels">http://www.homedepot.com/Building-Materials-Concrete-Cement-Masonry-Concrete-Blocks-Bricks-Lintels</a></td>
</tr>
<tr>
<td><img src="image" alt="Polypropylene cup" /></td>
<td>Polypropylene cup joined to polypropylene lid. Source: <a href="http://www.ikea.com/ca/en/catalog/products">http://www.ikea.com/ca/en/catalog/products</a></td>
<td><img src="image" alt="Nylon Manifolds" /></td>
<td>Nylon Manifolds two halves joined together. Source: <a href="http://www2.basf.us//PLASTICSWEB/displayanyfile?id=0901a5e1800489c">http://www2.basf.us//PLASTICSWEB/displayanyfile?id=0901a5e1800489c</a></td>
</tr>
<tr>
<td><img src="image" alt="Rattan" /></td>
<td>Rattan joined in a mesh shape. Source: <a href="http://www.ikea.com/ca/en/catalog/products">http://www.ikea.com/ca/en/catalog/products</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Cell stacking process of high-energy lithium-ion cells

J. Kurfer, M. Westermeier, G. Reinhart
Institute for Machine Tools and Industrial Management, Technische Universität München, Munich, Germany

Abstract: Electro mobility is a significant future trend justified by the efficiency as well as by the low emissions from electric drives compared to conventional internal combustion engines. High-Energy Lithium-Ion cells seem to be the most promising energy storage devices for electric vehicles. In order to reduce manufacturing costs, improvements for the production are necessary. This paper focuses on the process of cell stacking and presents a methodical correlation between process parameters and the stacking outcome as a result of various investigations on the process.

Keywords: Assembly, Handling, Electric vehicle

1. Introduction

In these days the reduction of CO2 emissions is a societal issue driven by the global discussion about the climate change and its impact on political decisions. In the point of individual mobility the trend to electric vehicles is a direct consequence of these political guidelines on the one hand and of the broad acceptance in society on the other hand. Although technologies for electric vehicles are mostly available, introducing the new key components into automotive mass production leads to fundamental changes of the established supply chains. Furthermore, huge challenges on the level of production systems and processes are to be solved. One of these new components, which adds significant value to an electric vehicle and comes up with a number of unsolved production engineering questions, is the traction battery. In case of lithium-ion cells its production can be separated into two main process chains: the cell production and the battery system assembly. Investigations in the field of assembly system design for battery manufacturing and its dependency of module design are addressed in [1, 2]. The Institute for Machine Tools and Industrial Management (iwb) focuses on systems and processes for the mass production of lithium-ion cells. For this purpose a laboratory for the applied research in this field was set up at iwb, funded by the German Federal Ministry of Education and Research (BMBF) [3]. In the current state of expansion this laboratory includes a climatic chamber, providing a realistic production environment, and two highly automated machines, one for the preprocessing of the electrode materials and one for the cell stacking process, which is examined in this paper.

2. Cell Stacking Process

The cell stacking is an essential process for the production of lithium-ion cells. Its task is to create an alternating arrangement of electrodes and separator layers in order to achieve an electrochemical functional cell at the end of the process chain. In this paper, cells with an anode-cathode-anode arrangement are investigated. If \( n \) is the uneven number of electrodes to be assembled in the cell, there are \( (n+1)/2 \) anodes, \( (n-1)/2 \) cathodes and \( n+1 \) required separator layers.

For the development of an automated stacking machine for lithium-ion cells, a methodical comparison of known stacking methods was conducted. Its aim was to identify the process promising the best trade-off between the cycle time and the resulting stack quality, see Table 1.

Table 1: Comparison of cell stacking processes

<table>
<thead>
<tr>
<th>Property</th>
<th>Flat winding</th>
<th>Single sheet stacking</th>
<th>Z-folding</th>
<th>Z-folding with single electrodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process time</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Electrode damages</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Separator handling</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Conventional z-folding [4] and flat winding [5] processes, well-established in the consumer electronics sector, have inherent drawbacks in processing electrodes for high energy cells. The coating thickness of such electrodes, which reaches from 50 to 60 microns per layer, results in a high critical bending radius. By folding, winding or pressing the electrodes as usually applied during z-folding or winding processes, the coating gets damaged in the bending area (-). The consequences are a loss of active surface of the electrode and the risk of particles that can penetrate the separator, where the latter defect can cause an internal shortcut. In contrast to conventional z-folding and flat winding, stacking of cut single electrodes and separator sheets [6] allows a very gentle material processing (+). Due to its sequential cycle (-), it is relatively slow compared to z-folding and winding where the materials are processed continuously (+). Additionally, the handling of cut separator sheets is challenging due to their limp material behavior (-). Besides that, high electrostatic charge and porosity inhibit a controlled handling with conventional grippers. Therefore, the z-folding process with continuous feeding of the separator (+) and single sheet handling of the electrodes [7], which combines the advantages of conventional z-folding and stacking, is set up at the iwb. This method allows a fast material processing as well as a gentle, bending-free handling of the electrodes.

Figure 1 shows the z-folding process. At the beginning of the stacking process the separator is fixed on a vacuum table (1), which is mounted on a horizontal linear axis. The alternating movement of the table combined with the fixed guiding roll above the table generates a folded separator (2). The electrodes are provided in magazines and placed in the fold by means of two 4-axis handling systems (4: anode, 5: cathode). Their position and orientation on the grippers are measured with two camera systems installed above the folding table (3). Step by step, a lithium-ion cell is built up through the alternating arrangement of separator, anode and cathode. After the stack is...
finished, it is transferred to the fixing station (left side, not in the figure) by a mechanical gripper, where the stack is fixed with tape.

3. Variables and influences on the z-folding process

The major challenge of the cell stacking process via z-folding is the correlation between quantity and quality of the stacking output and a multitude of different technological variations and process parameters. In this section, the most important output values are contrasted with the influencing variables of materials, machine parts and process parameters.

3.1. Output values

The output of the z-folding process is a stack of anodes, cathodes and separator layers (Figure 2).

![Figure 2. Scheme of a Z-Folded stack](image)

Regarding the target of a profitable cell production, the main quantitative value is the required time for stacking one cell ($t_{\text{stack}}$). If a continuous production of stacks in the arrangement anode-cathode-anode is supposed, the stacking process consists in general of three parts:

Preparation ($t_{\text{prep}}$): At the beginning of a new cell stack, the stacking table is located on the left side (anode side). Before the process can start, the table has to move to the middle of the stacking unit in order to enable the monitoring of the anode position on the gripper. After the investigation and correction of the electrode position the next period starts.

Alternated stacking ($t_{\text{stack}}$): This period starts when the table begins to move from the middle position towards the respective side of the stacking unit. It is divided in the movement towards the side ($t_{\text{side}}$), the period of placing an electrode on the stack ($t_{\text{place}}$) and the movement back to the middle position ($t_{\text{mid}}$). The investigation of the electrode position on the gripper happens parallel to the positioning of the electrode on the opposite side, so that this process has no influence on the cycle time.

Cutting ($t_{\text{cut}}$): After the stack has reached its requested number of electrodes it is covered with an additional separator layer by a last stacking movement. Then the table moves to the left, where the stack is picked up by a mechanical gripper and transferred to the fixation unit. The separator foil is cut off and fixed on the stacking table again.

According to this, the cycle time for one stack which consists of $n$ electrodes (anodes + cathodes) is defined as:

$$t_{\text{Cycle}} = t_{\text{prep}} + n \times (t_{\text{side}} + t_{\text{place}} + t_{\text{mid}}) + t_{\text{cut}}$$

In addition to the cycle time, the quality of the stack has to be evaluated. Therefore, two major quality aspects were considered. On the one hand, the positioning accuracy of the electrodes on the stack is a criterion for the cell quality. It is defined by three aspects (Figure 3):

- Deviation in the direction of movement of the stacking table ($\Delta x$) from the center point
- Deviation normal to the direction of movement ($\Delta y$) from the center point
- Deviation angle $\Delta \rho$ between the electrode edge and the x-y-coordinate system

![Figure 3. Criteria for the precision of the stack](image)

On the other hand, the separator and electrode materials can be damaged during the folding process, due to the mechanical contact between the materials and the machine components of the z-folder. Thereby, the following types of defects can occur (Figure 4, numbered):

- Cracks in the folding region of the separator (1) can cause short circuits in the cell and a complete rupture of the foil. In the latter case the folding process is interrupted completely.
- Scratches and cracks on the electrode and separator surfaces (2) can cause short circuits and a loss of electrochemical active surface, which leads to a lower energy density.
3.2. Influencing variables

The influencing variables (Figure 5) on the results of the z-folding process are divided in three groups. Firstly, there are variables that are settled in the product characteristics of the cell. This includes the anode, cathode and separator materials with their material properties (basic material, coating, thickness, minimum bending radius, etc.), the number of layers in the cell (stacking height) and the geometric shape of the electrodes, where the last two properties define the dimensions of the stack.

The second group of variables are the components of the z-folder that are directly involved in the folding process. In general these are the electrode handling units, the folding table with its components (e.g. the blank holders), the separator feeding system including the regulation of the separator web tension and the camera system for the investigation of the electrode positioning.

Third, the process is influenced by parameters that can be varied by the machine operator. These are mainly parameters concerning the motion of the folding table. The table has two axes, one in the x-direction to enable the separator folding, the other one in z-direction to compensate the growing stack height with each new layer. Each of these axes is influenced by the motion parameters position, velocity, acceleration and jerk. Additionally, the motion of the electrode handling and the tension of the blank holders on the stack must be considered.

4. Experimental investigations

Z-folding is a complex process influenced by the product specifications, the used machine components and the process parameters. In the following, the experimental conditions, the design of the experiments and the results are described and finally discussed.

4.1. Experimental conditions

In this first subsequence, the experimental conditions are described. The specification of the cell stack is shown in Table 2. In order to reduce the wastage of battery materials, a small number of electrodes has been selected. In the experiments, the influence of the folding table motion and the separator web tension on the cycle time and the separator damages (case 1) have been investigated.

The experiments have been executed in a climate chamber with a very low humidity of about 0.1 % (- 60 °C dew point). Thus, the experiments have been conducted at constant environmental conditions according to industrial standards. The influences of the electrode handling systems have been neglected in these tests.

<table>
<thead>
<tr>
<th>Table 2: Specification of the cell stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Length (x) (mm)</td>
</tr>
<tr>
<td>Width (y) (mm) (without conductor)</td>
</tr>
<tr>
<td>Thickness (mm)</td>
</tr>
<tr>
<td>Material</td>
</tr>
</tbody>
</table>

4.2. Design of Experiments (DoE)

The experiments were planned as central composite design [8]. Three parameters were taken into account: the velocity and acceleration of the folding table as well as the separator tension, represented by the force of the tensioning lever. The constraints have been defined by pre-investigations. Since high values of the table jerk lead to a complete rupture of the separator web, the experiments have been executed by a constant low value of 2000 mm/s³. With the method of central composite design, non-linear correlations can be identified by using only two levels for each parameter. This is enabled through the usage of center and star points (Figure 6). The set of experiments is defined by a full factorial analysis of the cube.

Figure 5. Variables influencing the z-folding process
points \((2^3 = 8)\), a separated investigation of each star point in combination with center values of the other parameters \((2 \times 3 = 6)\) and with a specific number of center point experiments. This number and the distance of the star points from the center define the statistic properties of the set. Here, the distance has been selected to \(\sqrt{2}\) and thus the number of center point experiments to 4, in order to get an orthographic set. The values of the parameters in the different points are shown in Table 3.

<table>
<thead>
<tr>
<th>Table 3: Experimental set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Star point</td>
</tr>
<tr>
<td>normed</td>
</tr>
<tr>
<td>Velocity (v) [mm/s]</td>
</tr>
<tr>
<td>Acceleration (a) [mm/s²]</td>
</tr>
<tr>
<td>Lever force (F) [N]</td>
</tr>
</tbody>
</table>

4.3. Experimental results and discussion

The results are divided in the effects on the separator damages and on the cycle time. The effects of the experimental parameters on the stacking precision can be neglected, because the positioning of the electrodes is predominantly influenced by the electrode handling system and the camera settings.

First of all, a method for the quantification of the damages must be developed. For that, the occurring damage in each of the seven folding zones is divided in front and back side (in \(y\) direction). Hence 14 areas are valuated with points from 0 (rupture) to 4 (no damage). The sum of the values is normalized by the best possible result so that every stack receives a quality value from 0 (rupture) to 1 (no damage). The damaged areas have been investigated with microscopic pictures. According to the experimental set, the influence of each parameter can be examined. The dominating parameters identified by a statistical test are \(v\), \(F\), \(v^2F\) and \(F^2\) (\(p\)-value = 0). The effect of the acceleration can be neglected (\(p = 0.70\)) because of the limited jerk which inhibits high values for the acceleration on the short distances in the \(z\)-Folder. Whereas the effect of the velocity on the separator damages is significant, low velocities enable high quality values because of the smooth folding procedure. The web tension also shows a significant correlation with the quality. Low lever forces lead to a non continuous tension in the separator, when the folding table moves towards the middle position. In this sequence, separator material is fed back into the buffer system. At the transition point the direction of the material motion changes, where a high separator tension leads to significant stresses in the separator material at the blank holders. Figure 7 points out the correlation between velocity, lever force and the quality of the folding. The measured values are wide spread because of the experimental design. In order to show the dependency of the stacking quality from only one parameter, a regression analysis was used.

![Figure 6](image)

**Figure 6.** DoE method with center and star points (according to [9])

The dominating parameters for the cycle time are \(v\) and \(v^2\) (\(p = 0\)). Also the lever force is characterized by a low \(p\)-value but its influence is very small (VIP (Variable Importance in Projection) = 0.69) compared to the velocity (VIP = 1.73). Analog to the results concerning the damage of the separator, the acceleration can be neglected because of the low jerk value. Obviously, high velocity leads to low stacking times. An interesting point is the development of the stacking time over an increasing table velocity, because a state of saturation is reached at a certain velocity. This is caused by the electrode handling, the period for the positioning of the electrodes on the stack and the investigation by the camera systems. The latter aspect leads to the fact that the increase of the table velocity is useful up to the saturation point. A further decrease of the stacking time can only be reached by the acceleration of the handling systems and the reduction of the hold-up periods. Figure 8 points out the correlation between the stacking time and the table velocity.

5. Conclusion and Outlook

This paper points out the various aspects concerning the \(z\)-folding process for the stacking of lithium-ion cells for automotive applications. \(Z\)-folding with single electrode sheets enables a very gentle, bending free handling of the electrodes as well as continuous separator feeding, which is an important point for the increase of the stacking speed. The investigated stacking process can be used independent of the electrode coating and the coating thickness. Thus, high-power as well as
high-energy cell stacks can be produced. Furthermore, the process can most likely be used for the production of lithium-sulfur cells.

However, the process is affected by a lot of influencing parameters, like the used machine parts, the product specifications and the process parameters. The correlations between the motion of the folding table, the separator tension and the outcome values of the cell stacking have been investigated.

Hereafter, more correlations will be analyzed by changing the separator materials from a coated non-woven material (Separion®) to porous polyolefin foils (e.g. Cellgard separator). Another aspect is the surface of the blank holders, which has a significant influence on the electrode damaging during the process. Not least, the relation between the table movement and the separator tension is one of the main reasons why the stacking speed cannot be further increased without reducing the quality of the folded separator.

In general, the cell stacking process is one key issue in the value chain for lithium-ion cells. To increase the stacking speed in order to enable an economic process, the system technology has to be improved. E.g. the decoupling of the table movement from the separator tension is one of the next steps. Also, the development of adapted gripping technologies is a key issue in this context. For a full evaluation of the z-folding process and its influence on the electrochemical properties of the cell, the completion of the cell production line at the iwb is necessary and will be realized in the near future.

Acknowledgements
This project was funded by the German Federal Ministry of Education and Research (BMBF) under the funding code 02PO2642 (DeLIZ) and managed by the Projektträger Karlsruhe (PTKA-PFT).

References
Interpretation of multiaxial gripper force sensors

K. Tracht\textsuperscript{a,c}, S. Hogreve\textsuperscript{a,c}, S. Bosse\textsuperscript{b,c}

\textsuperscript{a} Bremen Institute for Mechanical Engineering, University of Bremen, Bremen, Germany
\textsuperscript{b} University of Bremen, Department 3, Workgroup Robotics, Bremen, Germany
\textsuperscript{c} ISIS Sensorial Materials Scientific Centre, Bremen, Germany

Abstract: Mechanical grippers are key components of handling devices in automated assembly systems. For complex handling tasks these grippers can be equipped with additional force measuring modules. This paper presents the prototype of a gripper with fingers made of sensorial material. These gripper fingers contain six single force sensors and can measure forces along multiple axes. The results of the experimental investigations of the sensor performance are shown. It is also demonstrated how further information about the handling conditions can be derived by the computational combination of the sensor signals. A sensor network will enrich the capabilities of the gripper fingers.

Keywords: Handling, Measurement, Sensorial material

1. Introduction

Automated assembly systems are often equipped with mechanical grippers with two or more parallel fingers. They are one key component for the performance of handling devices. Force sensors, which are attached to the gripper system, enhance the functionality of the handling device by enabling the gripper system to grasp pressure sensitive objects and helping to pick up objects with unknown shape and orientation. Furthermore, the gripper gains the ability to perform force adaptive joining operations or force adaptive path corrections. Since the fingertips of grippers with force measurement systems do not need a work piece specific geometry, they can be adapted to new production setups more easily. Therefore, they help to overcome problems that arise from frequent product changes in flexible assembly systems. Commercially available force measurement systems for mechanical grippers are heavy and expensive. They are produced in the form of modules that can be added to the basic gripper. It can be distinguished between modules that are installed at the wrist of a robot system and modules that are joined to the finger of a gripper. Wrist sensors are usually multi-component sensors but they increase the size of the end-effector and add mass, so that the payload is decreased \cite{1}. Although wrist sensors are multi-component, they cannot measure the gripping force. Therefore finger sensors are suitable. They are lighter but harder to install. If both, gripping forces and external forces, should be measured two systems have to be installed, which results in a heavy and complex end-effector. The evaluation units of the sensor modules are usually placed in the central control cabinet of the handling device and require therefore several long cables, which makes it difficult to quickly exchange the end-effector.

Grippers with fingers made of “sensorial material” have the potential to overcome these drawbacks. They add only little mass to the gripper, allow force measurement along multiple axes and enable decentralized control strategies. Sensorial material is a new class of material that is able to retrieve information about the surrounding and its own condition and to communicate it \cite{2}. It is based on miniaturized sensor elements and chips and their embedment in a host material. The sensorial material allows the integration of the sensing elements directly into the gripper finger. Several sensors can be placed on one finger and allow multiaxial force measuring. The different sensor nodes will be connected to a sensor network and will perform fast and robust computing of the sensor signals \cite{3}. In a first development step a conceptual design based on strain gauge equipped gripper fingers has been introduced \cite{4}. These strain gauge equipped gripper fingers are a preliminary stage to a passive sensorial material. This paper presents the gripper finger design and the performance quality of the sensor system by experimental investigations. It is also demonstrated how further information about the gripper condition can be derived from the sensor data by intelligent interpretation. Finally it is explained how the transition from a passive to an active sensorial material will be reached in the future.

2. Integration of force sensors in gripper fingers

To evaluate the concept of a gripper with sensory gripper fingers, a prototype has been implemented at the Bremen Institute for Mechanical Engineering. This prototype consists of a commercial actuator which is equipped with two identical novel gripper fingers. Each gripper finger is cut out from a single aluminum block and contains three force sensors that are integrated in the basic finger structure. The force sensors are based on strain gauges that are connected to Wheatstone bridges. The three force sensors in every finger are arranged orthogonal to each other so that every sensor should ideally measure forces only along one basic axis (\(x, y, z\)). To support the physical separation of the basic force components into \(x\)-, \(y\)- and \(z\)-forces, special designed H-shaped cut outs have been milled into the gripper finger structure. These help to reduce the interferences between the different directions. This means, for instance, that the sensor for forces along the \(x\)-axis should not respond if the applied force only acts along the \(y\)- or \(z\)-axis.

Figure 1 shows the prototype of the gripper. The three sensors at the second finger are highlighted by red frames. It can be seen that the \(z\)-sensor is formed by two H-shaped cut outs. Each of these cut outs is equipped with four strain gauges, which results in eight strain gauges for the whole \(z\)-sensor. They are connected to a single Wheatstone bridge and can therefore compensate resistance changes that are caused by deformations from torque inside the \(z\)-sensor. Such a torque results from forces along the \(y\)-axis which cause bending around the \(x\)-axis.
inside the z-sensor. The design of the finger has been evaluated by finite element analysis as well as by practical experiments. A single finger was clamped to a fixture and static loads were applied to the fingertip in different directions to calibrate and test each single sensor. The sensors showed adequate measuring accuracy and a very good separation of the basic force components [4]. For evaluation of the sensor performance under varying load conditions and for investigation of the sensor behavior during real gripping situations, further experiments have been carried out. The results are presented in the following sections.

3. Evaluation of sensor performance

3.1. Description of experimental setup
The gripper prototype described in chapter 2 was mounted to a test stand like depicted in figure 2. A modular data acquisition system (DAQ) was used to measure and record all signals in real time. The DAQ also contains analogue output terminals that are used to drive an external amplifier which delivers the required voltage and current to power the DC motor inside the actuator of the gripper. An additional load cell is also connected to the DAQ and delivers the reference signal which is then compared with signals from the gripper fingers. The load cell has a rated capacity of 900 N, has a nonlinearity of 0.05 % and a hysteresis of 0.05 %. This measuring accuracy is higher than the expected measuring accuracy of the gripper fingers. Therefore the load cell can be used as reference signal during the investigation of the sensor performance. Different experiments were carried out with the single gripper fingers as well as with the whole gripper assembly. The fingers were tested on their behavior under varying loads and on the influence of the contact point on the measuring accuracy. To evaluate the behavior of the force sensors under dynamic conditions the load cell was pushed against the fingers from different directions. For each pushing all signals of the gripper finger and the load cell were recorded simultaneously.

This procedure was repeated for different contact points at the fingertip to survey the influence of the position of the contact point on the measuring accuracy.

3.2. Results and interpretation of experiments
Figure 3 shows exemplarily the performance of the x-sensor of finger 2 for a load that is applied to the tip of the finger along the x-axis. \( X_2 \) is the signal of the force sensor and \( F_r \) is the reference signal of the load cell. Since \( X_2 \) and \( F_r \) are so close to each other that they cannot be distinguished in the graph, the relative error \( \Delta X_2/F_r \) is also depicted. It can be seen that the mean relative error is below 1 %. Except for small forces the relative error stays nearly constant over the whole measuring range. For investigation of the influence of the contact point the absolute measuring error was recorded for loads at different contact points. Figure 4 shows the results for different points at the bottom of the finger. For better illustration only the best fit straight lines are depicted.

It can be seen that the measuring error varies with the contact point. Although this happens within a small range, it is recommended to apply loads during calibration at the most likely contact point, which is at the very tip of the finger. For contacts that occur away from the tip, the measuring error might then be greater. But this is acceptable, since this type of contact normally only occurs during collisions and for collision detection the measuring error is of minor importance.
were recorded from the DAQ. This experiment was repeated for the opened gripper. Then the gripper was closed while the signals were placed upright on a stand between the fingers of the reference load cell were carried out. Therefore the load cell information is the actual gripping force that acts on the grasped object. The gripping force is also called internal force since it only appears inside the gripper system and cannot be observed from the outside. If no external force acts on the gripper fingers, the absolute values of both x-sensors in fingers 1 and 2 should be equal to the gripping force. But if an additional force acts from the outside there might be a load shift from one finger to the other. Therefore it is assumed that the actual gripping force could be derived by calculating the arithmetic mean of the measured forces \(X_1\) and \(X_2\).

\[
F_g = \frac{X_2 - X_1}{2}
\]  

(1)

To evaluate this assumption, experiments with the gripper and the reference load cell were carried out. Therefore the load cell was placed upright on a stand between the fingers of the opened gripper. Then the gripper was closed while the signals were recorded from the DAQ. This experiment was repeated for different preset motor voltages. The graph in figure 5 compares exemplarily for one single run the calculated gripping force \(F_g\) with the measured forces \(X_1\), \(X_2\) and the measured force of the reference load cell \(F_r\). It can be seen that \(F_g\) is equal to \(F_r\), which proves that the assumptions have been correct and that equation (1) is suitable to calculate the gripping force.

Besides the internal gripping force, the value, direction and contact point of external forces are of interest during the control of handling processes. These external forces can be caused by unforeseen collisions with other objects or during the joining with another object. Figure 6 depicts a free-body diagram of two gripping fingers and a grasped object. From the equilibrium of forces along the x-axis the following equations for the external force \(F_{ext,x}\) can be derived.

\[
F_{ext,x} = X_1 + X_2
\]

(2)

Similar equations can also be derived for external forces along the y- and z-axis.

\[
F_{ext,y} = Y_1 + Y_2
\]

(3)

\[
F_{ext,z} = Z_1 + Z_2
\]

(4)

To get the direction and value of the overall external force the equations (2) to (4) can be combined to a vector equation.

\[
\begin{bmatrix}
F_{ext,x} \\
F_{ext,y} \\
F_{ext,z}
\end{bmatrix}
= \begin{bmatrix}
X_1 + X_2 \\
Y_1 + Y_2 \\
Z_1 + Z_2
\end{bmatrix}
\]

(5)

Experiments were carried out to verify the validity of the equations. A rectangular workpiece was grasped by the gripper and then the load cell was used to push against the workpiece from different directions. A ball-shaped tip that was screwed to the load cell reduced the contact area to nearly point contact and allowed the measurement of the contact point position. The graphs in figure 7 demonstrate the experiment results for an external force along the x-axis (a) and the z-axis (b). The experiments with y-sensors delivered comparable results.

The graph in figure 7 a) shows an interesting behavior of \(X_1\) and \(X_2\) for external forces over a certain limit (15 N in this example). If a positive external load is applied along the x-axis, the absolute value of \(X_1\) increases while the absolute value of \(X_2\) decreases by the same amount at the beginning. This is the expected behavior and results in a constant overall gripping force \(F_g\). But when the external force \(F_{ext,x}\) further increases, \(X_1\) starts to increase together with \(X_2\) which represents an increase of the gripping force, although the gripper actuator is not powered. Figure 8 depicts the reason of this phenomenon. Due to the leverage effect of the workpiece, the gripper fingers are spread apart. Since the fingers cannot draw aside due to the self-locking transmission of the gripper actuator, the spreading results in elastic deformation of the fingers and therefore increases the gripping force.

The leverage effect allows the determination of the contact point from the force sensor signals. The equilibrium of moments around the y-axis delivers equation (6). The distance \(w\) between the both finger tips is a function of the tilt angle \(\alpha\), but for the expected forces \(\alpha\) is very small and it is assumed that \(w(\alpha(F_{ext,y})) = w(0)\), which equals the width \(s\) of the workpiece.

\[
h = \frac{X_1 + Z_1 + s}{X_1 + X_2} = \frac{-X_1 - Z_1 + a}{X_1 + X_2} + a
\]

(6)

In this case the contact point is described by the distance \(h\) from the upper side of the finger tip, like depicted in the free-body diagram in figure 6. Figure 9 shows the result of an online calculation of \(h\) for the same experiment as in figure 7 a). Equation (6) is only valid when the leverage effect is in progress.
5. Sensor network for gripper sensors

For principle experimental investigations the evaluation of sensor signals was performed on an external central processing unit. The DAQ provided the necessary measuring and control hardware. This is not in the sense of a sensorial material, since a sensorial material should perform decentralized preprocessing of the sensor signals and should communicate ready-to-use information to high level control systems. The next generation of sensory gripper fingers should head towards an active sensorial material. The transition from a passive to an active sensorial material component will be performed using decentralized miniaturized low-power active sensor nodes that are arranged in a sensor network integrated in the finger structure. Each sensor node consists of a small printed circuit board and contains components for power regulation, signal acquisition, data processing, and communication. Every force sensor is connected to one sensor node, resulting in three sensor nodes for each gripper finger. The integrated sensor network leads to increasing functionality and robustness. Using a central processing system requires manual calibration of all six sensors. The proposed sensor nodes use a zooming ADC approach with auto-calibration. This allows self-calibration and adaption to the change of environmental conditions through the network of sensor nodes. The communication between the sensor nodes uses the Simple Local Intranet Protocol like proposed in [3].

6. Summary

This paper demonstrates the concept of a mechanical gripper with sensory gripper fingers made of passive sensorial material. The gripper fingers contain six single sensors and can perform multiaxial force sensing of internal and external forces. Experiments showed that the measuring accuracy is of suitable quality for handling process control, if the calibration is done at the tip of the gripper finger. Collisions can be detected at all points below the x-sensor section. Therefore the measuring accuracy is of minor importance. The six single sensors deliver data about the acting forces divided into basic force components along the x-, y- and z-axis. To get information about the effective gripping force or external load, further computation has to be performed. The gripping force can be calculated by the arithmetic mean of the both x-sensor signals. The direction and value of an external load can be derived from the combination of all six signals. Due to the leverage effect it is also possible to determine the contact point of an external load at the workpiece. The transition from a passive to an active sensorial material by the integration of decentralized sensor nodes connected via a network is part of the future research.

Acknowledgement

This work is part of the research carried out at ISIS Sensorial Materials Scientific Centre. The Bremen Institute of Mechanical Engineering and the Workgroup Robotics are members of ISIS.

References

Calibration of reconfigurable assembly cell using force feedback and vision

S. Dransfeld
Department of Production and Quality Engineering, NTNU, Trondheim, Norway

Abstract: One of the most time consuming operations in a reconfigurable assembly cell is the calibration of surrounding peripherals. Accurate calibration is crucial for a successful implementation for reconfigurable assembly systems, as flexible devices has larger tolerance areas, more degrees of freedom, and a dynamic work area. This paper describes how a reconfigurable assembly cell can be constructed, and by introducing a force sensor for the robot, an automated calibration system is implemented. The speed and accuracy of the implemented system is tested.

Keywords: Assembly, Calibration, Flexibility

1. Introduction

Systems for location of components by usage of vision have been a research topic since the early 1980s [1, 2, 3]. In flexible and reconfigurable assembly cells, vision has become the ubiquitous solution to locate randomly placed parts, driven by the need for flexible feeding solutions [4].

Even though there has been a tremendous development in technology for vision sensors, from single camera systems to multi camera stereo vision systems, and 3D vision, the calibration of machine vision solutions is usually done manually, even in research [5]. For stereo vision, or robot mounted cameras, automated calibration solutions have been developed [6-8].

For most applications, especially for high speed applications where SCARA robots are used, single camera solutions will be faster, simpler and cheaper to utilize. Single camera systems only provide coordinates in a plane, and the height of the plane must be calibrated separately. As robots can be dynamically positioned, it is important to create a setup where this advantage is utilized, and less time is spent on mechanical adjustment of the surroundings.

The main objective in calibration of machine vision systems is to co-locate points in both vision coordinates and robot coordinates, so parts located in vision coordinates can be transformed into robot coordinates. Because of the many different factors affecting the resulting accuracy, it is important to understand and handle the individual tolerances of each subsystem in the calibration chain.

Lens distortion is one of them, and the camera system should be geometrically calibrated to improve the accuracy of the object recognition. If absolute measurements are needed, the geometric calibration must be three dimensional, but for standard pick and place operations a two dimensional calibration should be sufficient.

The robot vision calibration should then be performed in 3D, to handle objects with different heights. A 3D calibration will drastically improve localization of parts which have a different picking plane than the base calibration plane of the pickable area.

The requirement to create a 3D calibration is a least 5 points, and if the components to be picked can be rotated, each calibration point should be taught from several angles to average any rotational errors in the calibration chain. This kind of inaccuracy is often caused by the relatively poor absolute accuracy of articulated robots, and more calibration points will increase the total accuracy of the machine vision system. Especially as the offset errors in an articulated robot design is non-linear.

The objective of the work presented in this paper is to demonstrate an effective method for automated calibration of a flexible and reconfigurable assembly cell. The cell is an industrial prototype of a reconfigurable assembly system, designed for pick and place assembly of small batch products. Some restrictions for the construction of the assembly cell are discussed, and the speed and accuracy of the implemented calibration solution is presented.

By introducing a force sensor to the robot, and constructing suitable calibration targets, it is possible for the robot to locate the calibration target itself, and thereby automating most of the calibration process.

2. System setup

The assembly cell, shown in figure 1, consists of one six-axis articulated robot equipped with a force sensor and a tool changer. Around the robot are several positions where flexible feeders or blisters can be placed for component supply. Each position has a single camera overlooking the reachable area.

![Figure 1. Image of work cell](image)

Each picking area is flat, because of the limitations of the single camera setup. The robot is equipped with a dedicated calibration tool during calibration, and uses a cone shaped calibration object with three levels to obtain 3D information from the vision analysis.
2.2 Calibration object

The calibration object, shown in figure 2, is a circular cone which is moved around in the calibration area by the robot. The calibration tool for the robot is pointed, and the calibration object has a centre hole so the robot can centre the calibration object by entering the calibration tool.

Figure 2. Calibration object

The calibration object can be located in three heights by the camera: the circular area at the bottom, the middle level, and at the top. By having three heights, the machine vision system can use the height information to 3D calibrate the vision field, and parts with varying heights can later be located with improved precision.

2.3 Force sensor

The robot is equipped with a 6-axis force/torque sensor, shown in figure 3, which is used to measure the position of the calibration object by pushing it down against the reference plane with a preset force.

Figure 3. Force sensor and tool changer at robot end effector

2.4 Vision system

The vision system consists of a CCD camera with resolution of 1024x768 pixels. A commercially available vision system is used for vision analysis. One important aspect of the system is that it can be geometrically calibrated, hence compensating the lens distortion. This makes it possible to remove the fisheye effect on images, and get distances measured in real units, not pixels.

For the experiments presented the system was geometrically calibrated in only one plane (2D). The pixel size of the resulting calibration was 0.5mm.

3. Calibration object accuracy

During calibration the robot centers the calibration object by entering the center hole, and uses the force sensor to detect contact. The force sensor is continuously read, and when the contact forces reach the preset threshold, a stop command is sent to the robot.

The robot controller has a cycle time of 16 ms, and after the stop signal is sent, the robot might use some cycles before it comes to a complete stop. Hence, the robot velocity during calibration movements is therefore important for overall system accuracy.

To test the centering accuracy of the calibration object, the robot moves the calibration object to a predefined X, Y position from 4 different perpendicular directions. After centering the object with the Z movement, the robot moves away from the calibration object, and a picture is taken. This is repeated for 9 different X, Y positions in the vision field, resulting in a total of 36 measuring points.

Figure 4. Plot of calibration object location deviation

<table>
<thead>
<tr>
<th>Table 1. Summary of object location deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>dx [mm]</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>Maximum</td>
</tr>
<tr>
<td>Std. deviation</td>
</tr>
</tbody>
</table>

The centering repeatability of the calibration object can be seen from table 1. All coordinates are measured by the vision system. “dr” is the deviation in radius of the calibration object. Figure 4 is a plot of all offsets. The variation is much smaller than the pixel size of the vision system, so the possible centering inaccuracy will have no major impact on the system accuracy.

4. Force sensing accuracy

As a single camera system can only locate objects in X, Y coordinates, the Z coordinate has to be calibrated independently. In an assembly cell, the plane of the pickable area is not necessarily normal to the Z axis of the robot, and needs to be calibrated. To achieve this, the X and Y coordinates
are used to calibrate the vision system, while the Z coordinates are found by using the force sensor.

Table 2. Summary of force sensing accuracy

<table>
<thead>
<tr>
<th></th>
<th>dx [mm]</th>
<th>dy [mm]</th>
<th>dz [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.02</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Std. deviation</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
</tbody>
</table>

To measure the force sensing repeatability the 9 X, Y positions from the calibration object centering test was reused, and the X, Y and Z positions of the robot was recorded and compared with the positions from centering test. The Z coordinate is found by force sensing. Table 2 shows the results from the test of force sensing repeatability.

The repeatability measured in this test is equal to the repeatability of the robot, meaning that the possible force sensing inaccuracy will have no major impact on the system accuracy.

5. Robot accuracy

The relatively poor absolute accuracy of an articulated robot is often the major contributor to system inaccuracy in a robot vision system. Hence, it is important that the machine vision calibration handles this inaccuracy. To test the robot accuracy the distance between the 9 X, Y positions was measured both by the vision system and the robot.

Table 3. Robot accuracy

<table>
<thead>
<tr>
<th></th>
<th>delta [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>-0.58</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.58</td>
</tr>
<tr>
<td>Average</td>
<td>0.04</td>
</tr>
<tr>
<td>Std. deviation</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Table 3 shows the measured difference between the robot distances and vision distances. The results show that the error is slightly bigger than the resolution of the vision, and also the major contributor to system inaccuracy compared to the calibration object centering system. Since the vision system has sub-pixel accuracy shown in the calibration object centering test, it can be concluded that the robot inaccuracy is the major contributor to system inaccuracy.

6. System accuracy

Finally the system accuracy was tested. The calibration object was placed in 9 randomly chosen locations in the vision field. The robot then centered the calibration object using the force sensor and the calibration tool, and the position was checked again.

The accuracy of the calibrated object location system is shown in table 4. Figure 5 shows a plot of the location offsets.

The last test that needs to be performed is rotational accuracy of the calibration tool. The calibration of the system has been done with the same tool orientation to suppress any rotational error in the calibration tool. Hence, any rotational error in the calibration tool will introduce a calibration offset between the center point of the robot flange and the object position found by the vision system. To measure the rotational error of the calibration tool the calibration object was centered by the calibration tool with 4 different orientations.

Table 4. Summary of force sensing accuracy

<table>
<thead>
<tr>
<th></th>
<th>dx [mm]</th>
<th>dy [mm]</th>
<th>dz [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>-0.02</td>
<td>-0.14</td>
<td>-0.30</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.28</td>
<td>0.15</td>
<td>0.23</td>
</tr>
<tr>
<td>Average</td>
<td>0.13</td>
<td>0.02</td>
<td>-0.01</td>
</tr>
<tr>
<td>Std. deviation</td>
<td>0.11</td>
<td>0.13</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Figure 6 is a plot of the variation in object location when the tool is rotated around the Z axis. All values are measured by the vision system. The red circle is a fitted circle through the measured points, which has a radius of 0.26mm. This rotational error in the calibration tool is also a major contributor to the system inaccuracy, and needs to be handled in the system calibration. This is solved by rotating the tool around the Z axis, and uses the average vision position value as the calibration value.
7. Automatic calibration

As the FOV from a camera is rectangular, we can describe this rectangle by three points. If the robot is taught these three points, a suitable number of calibration points within this rectangle can be calculated.

To calibrate the used vision system, 18 points are needed. Since the calibration object has three heights, we can get 2 Z points for each calibration position, needing 9 points the robot must move to.

To remove the rotational error of the calibration tool, the average of 4 angles for each point was taken, making the total number of needed points 36.

For the automatic calibration sequence, the operator must teach the three points of the area to calibrate. The system will calculate all needed points, and move between them.

For each point the robot finds the Z coordinate by force measurement, and then moves away from the calibration object, so the vision system can locate the calibration object.

Table 5. Calibration time

<table>
<thead>
<tr>
<th>Operation</th>
<th>Time [mm:ss]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search</td>
<td>4:25</td>
</tr>
<tr>
<td>Move</td>
<td>1:31</td>
</tr>
<tr>
<td>Vision</td>
<td>0:04</td>
</tr>
<tr>
<td>Total</td>
<td>6:00</td>
</tr>
</tbody>
</table>

From table 5 we can see the time usage on the different operations of the calibration routine.

8. Summary and further work

By introducing a force sensor on the robot, an automated machine vision calibration system was assembled. The automated system can quickly recalibrate the assembly cell, and thereby opening up possibilities for quick reconfiguration.

The current robot – vision calibration assumes the robot moves in correct Cartesian coordinates, but as the calibration results show, this assumption is wrong. A better algorithm to predict the robot position should be introduced in the system.

9. Acknowledgements

This work has been conducted as a part of the research project KUPP founded by the Norwegian Research Council, and in cooperation with the industrial partners Kongsberg Automotive AS and Teeness AS.

References
Picking precision in assembly using robots and flexible feeders

S. Dransfeld\textsuperscript{a}, L. E. Wetterwald\textsuperscript{b}
\textsuperscript{a}Department of Production and Quality Engineering, NTNU, Trondheim, Norway
\textsuperscript{b}SINTEF Raufoss Manufacturing, Trondheim, Norway

Abstract: As complexity and multitude of tasks in flexible and reconfigurable assembly systems, there is an increasing demand on precision for flexible feeding and picking systems. Vision, robots, and flexible feeders have become the cornerstones in realization of flexible and reconfigurable assembly. Each component has been through a continuous evolution in precision, usability and flexibility, but not as a complete system. This paper describes how the possibilities of these components can be exploited to improve the understanding of the system and how this can be used to improve the precision.

Keywords: Assembly, Precision, Flexibility

1. Introduction

Flexible feeding has been an important factor in the realization of flexible and reconfigurable assembly cells the last decade [1, 2, 3]. Even though it is possible to handle a variety of parts in one flexible feeding system, most systems will only handle one component at a time – utilizing the flexibility for efficient changeover.

One major issue with flexible assembly systems is the overall precision. Although robots have a high repeatability, the absolute accuracy is often too poor to handle high tolerance assembly without sensor guidance. Powertrain assembly is one area which requires high accuracy, which has been solved by using sensor fusion (force and vision) [4].

In common flexible feeding solutions, the absolute inaccuracies of the system are often solved by using specialized grippers to orientate the components and relocation devices which improve the component location.

As more complex assembly tasks are to be handled by flexible assembly cells, there is an increased demand for the manipulators to handle several distinct components at once; thus making it inefficient to use dedicated grippers as the tool change time will be too high.

Modern robots have repeatability better than 0.1mm, tool change system have accuracy better than 0.01mm, and high resolution cameras with high precision lenses and sub-pixel location algorithms are used in measurement tasks with 0.001mm precision. By utilizing these technologies an accurate flexible system should be achievable, but industrial implementations show that there is higher failure rate in flexible part feeding than in dedicated feeding systems like bowl feeders. This is due to the large tolerance chain of a complete flexible and reconfigurable assembly system.

This paper presents a series of tests done on an industrial assembly cell for assembly of air couplings in manifolds. The objective of the test was to identify and quantify the different causes for picking inaccuracy, and suggest methods for handling these inaccuracies. The assembly cell utilizes flexible part feeders and a 6-axis articulated robot for the assembly task. Part localization is done with a 3D calibrated single camera vision system.

Based on the findings in these test a set of requirements for designing a flexible and reconfigurable assembly cell is proposed.

2. System setup and part localization

The assembly cell consists of one six-axis articulated robot equipped with a tool changer and a gripper with two flat fingers (see figure 1). One flexible feeder with a single camera overlooking the reachable area is used for object location. To measure the picking variation another camera is used to take a close-up of the object situated in the robot gripper after picking.

The flexible feeder consists of a feeding surface and a hopper. To generate feeding movement, the feeding surface is placed on to elliptic axes which produce forward and backward feeding motions, and flipping. Flipping is used to reorient parts.

Figure 1. Image of workcell

There are several reasons for choosing a single camera setup. The main reason is the reduction of complexity of the whole setup, and the reduced cost of not needing more cameras. Calibration of single camera systems is also simpler, and therefore faster. The tasks the assembly cell is constructed for are high speed assembly operations (cycle time < 3s), so components have to be partly organized. Bin picking and stereo/3D vision is therefore not applicable to the requirements of the assembly cell.

The vision system consists of a CCD camera with a resolution of 1024x768 pixels and commercially available vision software. One important aspect with the system is that it can be geometrically calibrated. Geometrical calibration compensates for lens distortion and removes the fisheye effect on images. Hence, the distances are measured in real units, not pixels.
The camera is 2D geometrically calibrated in one plane. Based on this geometrical calibration a 3D robot calibration is done. The resulting pixel size is 0.35mm in the reference plane.

The assembly task includes several different parts, and one of the parts with highest failure rate regarding picking precision was chosen. One important and obvious reason for the relatively high failure rate is the uneven surface of the part, causing it to stabilize in many different poses.

![Figure 2. Inscribed circle](image)

To locate the part, the centre of each side of the middle flange is found. With these two points and the camera location, an inscribed circle (see figure 2) can be created, and thereby locating the centre of the part.

Figure 3 shows the two localization points on the flanges and the resulting centre point which the robot used to pick the part.

![Figure 3. Localization points](image)

Even though the camera is 1.2m above the picking surface and the picking surface is only 0.2x0.2m it is important to handle this projection error. If the projection error was not handled, an error up to 1.2 mm would occur based only on where the part is located in the feeder. The maximum distance from the camera-feeder centre point is 0.15m.

3. Measuring of picking precision

To test the locating precision the pick operation was repeated 1602 times. 10 parts were added to the feeder and each part was returned to the feeder hopper after inspection. Out of the 1602 pick operations, 3 resulted in a grip on the part which was extremely out of position relatively to the others. The rest of the picking operations were within the tolerance shown in figure 4.

![Figure 4. Box – Whiskers plot of picking tolerance](image)

To get a qualified result of the test it is important that parts are picked from the complete picking surface, as there can be variation in vision calibration and robot accuracy over the complete area. Figure 5 shows the density of located parts on the picking surface. As can be seen there is a concentration of located parts in the lower right corner. There are two reasons for this. The first is that the surface of the feeder is not 100% leveled and the parts tend to cluster in this area. The other reason is that the vision system starts searching for parts here. So if there are valid parts located over the complete surface, the vision system prefers parts located here.

Since a two fingered gripper is used, the part will be centered along one of the three axes in the gripper. The other two position axes are used to test the precision of the part in the gripper. Figure 6 shows the two poses used. The first measurement is from figure 7a and shows how the part moves longitudinal in the gripper; the second measurement is from figure 7b and shows how the part moves transversal in the gripper.

![Figure 5. Part location on picking surface](image)

![Figure 6. Part situated in gripper](image)
There can be two types of errors in the system, a systematic error and a random error. Figure 7 shows that the system has a static offset in one direction. If the part is rotated normally to this direction (0 or 180 degrees) the error is reset by the gripper centering. Otherwise it is constant in one direction. The red curve is a fitted sine to the data with amplitude of 1.3mm. This error was traced back to an eccentricity in the calibration tool.

Figure 7. Pick rotation vs. longitudinal offset

Figure 8 shows that there are two peaks in the position data, which can be traced back to the two main positions the part flange can be rotated at. This is shown in figure 9. One side of the flange is flat, so when situated on the flat side, the part will be lower.

For each peak, more than >99% of the parts are within an offset error of ±0.75mm.

Figure 8. Histogram of transversal offset

Even though a robot has a repeatability of 0.02 mm, and the sub-pixel accuracy of the vision system is < 0.1 mm, the resulting accuracy of the complete system is a magnitude larger.

Table 1. Camera position relative to robot

<table>
<thead>
<tr>
<th></th>
<th>X [mm]</th>
<th>Y [mm]</th>
<th>Z [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old position</td>
<td>-40.04</td>
<td>-674.57</td>
<td>1302.94</td>
</tr>
<tr>
<td>New position</td>
<td>-32.20</td>
<td>-681.76</td>
<td>1317.18</td>
</tr>
</tbody>
</table>

Figure 11 shows the longitudinal offset distribution before and after adjustment of the systematic error. After the adjustment 96.9% of the parts are within an offset error of ±1mm.

The recalibration moved the camera 15mm relative to the feeder, see table 1.

The other major reason for inaccuracy in the picking precision has been traced to the physical properties of the part. Two sides of the flange the part is resting on are flat, on two sides are rounded, causing the part to rest in two different
process has traditionally been a multi-step and time consuming approach lacking precision in flexible assembly systems. The calibration process, and poor calibration is one of the major contributors to inaccuracy causing downtime and quality problems, it will be a profitable trade-off.

By introducing a position correction camera, or using the picking camera, to take a picture after a component is picked, it is possible to analyze the performance of the flexible feeding system. This camera does not have to be accurately positioned, as we are only interested in relative movement of the component in the gripper.

In flexible PCB assembly systems there are often installed extra cameras which are used in each assembly to improve the component position. For manufacturing this is often not an option as it will be inefficient to always pass one point, components are not flat, and assemblies are rarely pure 2D.

5. Evaluation of picking precision

The tests conducted in this study are performed on a relatively small component; ~40mm long and ~15mm in diameter. The picking surface is 200mm x 200mm, resulting in a pixel size of 0.35mm in the reference plane. If we scale the surface to a standard pallet which is 1200mm x 800mm we would achieve a pixel size of ~1.5mm. If a smaller pixel size is needed there are cameras available with >4x the resolution of the camera used in this study. Pixel size is strongly correlated to the accuracy a vision system can locate a component. Most vision systems utilize sub-pixel algorithms which can increase the theoretical accuracy to 1/10 of the pixel size, but in practice an increase to more than 1/5 of the pixel size is unlikely.

The accuracy of locating a small or large component should in theory be the same, but often will a larger component have more features which can be located. Vision algorithms are most effective when geometrical primitives are to be located, like squares or circles which are distinctly defined. The component selected in this paper does not have any well defined geometrical primitives, which is often the case in the real world.

In general the results from this study should be comparable if the size of the part changes. A larger part needs are larger view area, usually resulting in a higher pixel size. But a larger part will probably have more features to locate and more pixels to analyze for feature locations.

6. Design of flexible feeding system

Accuracy is critical for the performance of flexible feeding systems, and poor calibration is one of the major contributors lacking precision in flexible assembly systems. The calibration process has traditionally been a multi-step and time consuming manual process. This often results in avoidance of recalibration and optimization of the part localization, and instead physical means like specially designed grippers and relocation devices are used to overcome problems which could be corrected in the localization system.

With a cycle time below 3s, more than 5 million components will be handled each year. If some thousand of these cycle times are increased to improve the overall performance of the system by measuring and correcting the inaccuracy causing downtime and quality problems, it will be a profitable trade-off.

7. Summary and further work

Most commercial systems do not 3D calibrate, and most do not do geometrical calibration either. The reason why it works is mostly larger parts, dedicated grippers and robust repositioning. If parts are to be picked by a universal gripper and placed directly in an assembly, these are not viable options.

By analyzing the resulting pose of parts gripped in a flexible and reconfigurable assembly system it is possible to deduce if the current system is optimally configured, and the contributors to inaccuracy can be identified, quantified and eliminated with relatively simple methods in most cases.

8. Acknowledgements

This work has been conducted as a part of the research project KUPP founded by the Norwegian Research Council, and in cooperation with the industrial partners Kongsberg Automotive AS and Teeness AS.

References
Influence of welding sequences on compliant assembly geometric variations in closure assemblies

A.F. Nagy-Sochackia, M.C. Doolanb, B.F. Rolfeb, M.J. Cardew-Halla

a Research School of Engineering, The Australian National University, ACT, Australia
b School of Engineering, Deakin University, Victoria, Australia

Abstract: The fit of closure assemblies, such as door and hatches, is an important aspect of automotive quality. Poor fit and increased closing efforts can be easily identified by consumers and affect customer perceptions of brand quality. The sheet metal assembly process involves complex interactions that can impact significantly on the dimensional quality and corresponding fit of closure assemblies. This work highlights the importance of weld sequence selection in the presence of part variations in controlling the process output. The sequence is shown to influence both the deviation and variation of dominant shape modes; furthermore, a link between clamp placement and sequence selection has been identified.

Keywords: Assembly, Sheet metal, Joining

1. Introduction

During manufacturing, process variations will occur that affect the accuracy of the final part; the acceptable limits for these variations are specified as tolerances. When the parts are assembled together, the variations stack-up to form the final assembly variation. A variety of tolerance stacking techniques are used to combine the part tolerances to determine the final assembly tolerance. Similarly, part tolerances can be determined based on the desired final assembly tolerance. However, these techniques only apply to rigid bodies. In 1980, Takezawa's research showed that flexible or compliant parts, such as sheet metal panels, do not follow the traditional rigid body assembly rules [1]. Moreover, part variations could be absorbed through the assembly process.

1.1. Sheet Metal Assembly Modelling

The process of sheet metal assembly was represented as the Place, Clamp, Fasten and Release cycle by Chang and Gossard [2]. During this process, components are placed into a holding fixture, clamped into place on control surfaces, fastened together (either by welding or rivets) and then the clamps are released. To further develop an understanding of this process, Liu, Hu and Woo [3] applied linear mechanics to study assembly level variation based on either parallel or series assembly. This work highlighted the importance of assembly order, either serial or parallel, and illustrated why sheet metal assembly tolerance analysis does not follow the conventional rigid body rules. Another approach that has proven useful to develop relationships in simple assembly designs is the beam based approach developed by Ceglarek and Shi [4]. This incorporates the influence of part and fixture variations into the assembly process to better understand variation propagation.

When complete designs are available, another technique known as the Method of Influence Coefficients (MIC) [5] has been useful in studying the assembly process. Under this model, finite element simulations are utilised to determine a sensitivity matrix that relates part deviations to assembly spring back deviations. The model is based on the assumptions of small deviations, linearity and contact interactions are not considered. To develop this model further, Dahlstrom and Lindkvist [6] incorporated a contact based search algorithm, which showed better representation of the assembly behaviour.

The importance of both the clamping and welding stages in assembly has been identified in a number of studies [7-10]. Shiu, Shi and Tse [7] used linear mechanics to derive a minimum stress criteria for weld sequencing. This work showed a reduction in internal stress and consequently dimensional variation by welding from fixed to free ends. Hoffman and Santosa [8] developed a simple mathematical model and used optimisation techniques to show the importance of sequence dependence in both clamping and welding. However, to capture the effects of fixture design and the process steps that occur during both the clamping and fastening stages more accurately, a more complex approach is required.

Studies have examined a range of more complex approaches. The influence of thermal effects and sequence on overall distortion was studied by Fan et al. [9]. Matuszyk, Cardew-Hall and Rolfe [10] applied nonlinear contact based finite element analysis to illustrate the importance of clamp sequences in controlling shape variations of a top hat section. Assembly fixture design is an important factor in the control of dimensional variation of assemblies; Camelio, Hu and Ceglarek [11] presented an optimisation algorithm to minimise dimensional variation given part and weld gun variation.

To analyse variation in large high dimensional data sets, a common approach is through the use of Principal Component Analysis. This technique has been used in fault identification and diagnosis of the sheet metal assembly process [12] and for identifying variation patterns in data sets [13].

A substantial body of work exists in sheet metal assembly modelling. The importance of clamping and welding sequences has been identified and the impact is known for certain situations, such as long slender bodies with fixed and free ends. However, the impact on some structures remains unknown. This paper analyses a closure assembly representation forming a loop. This represents a different type of structure than has typically been analysed, with no fixed or free ends identifiable to weld from and to, respectively.
2. The case study

Closure assemblies are a common element in many sheet metal industries, particularly automotive. In this work, an idealised shape that represents the structural properties of an un-joined window frame was used in a simulation-based study. This involved finite element simulations with Principal Component Analysis (PCA) used to analyse the results.

The frame profile is represented in Figure 1. The profile forms a 500mm square loop. This size is smaller than a typically automotive closure assembly; however, its structural properties pre and post welding are similar to a production assembly. Where the structure itself does not have significant torsional stiffness until assembly is completed.

The gap shown in Figure 1 was varied between two and eight millimetres in one millimetre increments; this was effectively 1.25 and 7.25 millimetres after the material thickness of 0.75 millimetres was taken into account. This gap range was selected such that it completely encompasses, and goes slightly beyond, the gap range observed in a production environment.

![Figure 1. Shape and profile of the idealised structure](image)

For each of these parts, a different weld sequence was performed, with 25 different sequences studied. The final unclamped geometry of each part for all 25 sequences and seven gaps were combined together and Principal Component Analysis performed.

2.1 Finite element simulations

The simulations were performed using ABAQUS/Standard static implicit solver. A multi-step approach was used with contact interactions incorporated. The mesh size was 10mm, and surface to surface contact interactions were utilised. Both the weld tips and clamps were represented by analytical ridged surfaces. The simulation steps were performed as follows:

1. Apply all clamps
2. Perform weld
   a. Move weld tips to position
   b. Squeeze weld tips together with sufficient force to completely close the gap.
   c. Add element representing spot weld nugget
   d. Release weld tips
3. Repeat step 2 for each weld in the sequence
4. Release clamps

The sequences that were chosen were based on a variety of conditions; including welding clockwise and anticlockwise around, alternating sides, welding towards the clamps and welding away from the clamps. A number of randomly generated sequences were also performed. The fixture and weld layout can be seen in Figure 2. Four example sequences used in this work can also be found in Table 1.

![Figure 2. Fixture and welding configuration](image)

<table>
<thead>
<tr>
<th>Sequence Number</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>22, 23, 24, 1, 2, 3, 21, 4, 20, 5, 19, 6, 18, 7, 17, 8, 16, 9, 15, 10, 14, 11, 13, 12</td>
</tr>
<tr>
<td>5</td>
<td>21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 24, 23, 22</td>
</tr>
<tr>
<td>11</td>
<td>4, 1, 14, 9, 23, 6, 5, 21, 20, 12, 2, 16, 11, 3, 18, 22, 19, 10, 7, 8, 13, 17, 15, 24</td>
</tr>
<tr>
<td>14</td>
<td>16, 4, 7, 3, 19, 15, 11, 2, 22, 5, 17, 18, 13, 23, 6, 20, 10, 12, 8, 24, 9, 14, 21, 1</td>
</tr>
</tbody>
</table>

2.2 Principal Component Analysis

Principal Component Analysis (PCA) is a mathematical technique that transforms a data set into uncorrelated linear relationships, known as the principal components. This technique is useful in identifying the greatest variation in data sets and for dimensional reduction. It can be performed through eigenvalue decomposition.

To perform PCA, the data mean is first calculated and it is then used to determine the covariance matrix of the data set.

\[ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \]

\[ S = \frac{1}{n - 1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T \]

From here, the eigenvalue and eigenvector pairs are calculated from the covariance matrix. The eigenvectors \( \Phi \) now represent the principal components of the data set. The eigenvalues are relative magnitudes of variation for each principal component. The eigenvector/eigenvalue pairs are sorted in descending order based on the magnitude of the eigenvalues. When used for dimensional reduction, principal components that contribute less to the variance of the data can be neglected. The data can be reconstructed as follows:

\[ x \approx \bar{x} + \Phi b \]

Where \( b \) represents a vector with the corresponding magnitudes of each principal component contributing to a data sample.
3. Results and Discussion

3.1. Shape modes

Principal Component Analysis was performed and the variation modes of interest are shown in Figure 3.

The most dominant variation mode was based on the welded edges bending into each other for the varying gaps, which was expected. This is a consequence of the structure design and a gap variation study. The different gap sizes require different amounts of deflection for complete closure of the welded edge, which PCA identified as a variation. For the purposes of this work, this variation mode has been removed. The percentage variance of each remaining variation mode for the complete data set can be found in Table 2. Additionally, the sixth variation mode contributed just 0.3% of the remaining variation. This mode has also been removed from future analysis.

The second greatest variation identified was assembly twist (See Figure 3). This variation mode accounts for 85.7% of the dimensional variation of interest in the assembly. This form of variation is significant in closure assemblies, where incorrect flushness and increased closing efforts can occur. This variation mode is the focus of this work.

Table 2. Percentage of variation extracted via PCA from all sequence simulations final shape geometry

<table>
<thead>
<tr>
<th>Variation Mode</th>
<th>Variation % (Disregarding Mode 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>-</td>
</tr>
<tr>
<td>2nd</td>
<td>85.70</td>
</tr>
<tr>
<td>3rd</td>
<td>8.97</td>
</tr>
<tr>
<td>4th</td>
<td>3.15</td>
</tr>
<tr>
<td>5th</td>
<td>1.93</td>
</tr>
<tr>
<td>6th</td>
<td>0.26</td>
</tr>
</tbody>
</table>

The third variation mode was the folding of the structure’s internal corners in opposing directions across a diagonal. This variation, like twist, is a naturally compliant direction of the structure design. The fourth variation mode was a consequence of the clamping and welding locations, with variations in the magnitude of the gap between welds occurring depending on the sequence. Finally, the fifth variation mode represents a similar variation to the third mode, although 90 degrees out of phase across the other diagonal. However, it also appeared to be influenced by additional variations in the lower right hand corner.

3.2. Structure twist

The projected data for the twist variation mode (mode 2) were then plotted against the gap. This was done for the selected sequences and is shown in Figure 4. Each of these curves represents the relationship to the twist variation mode for a specific sequence. The shape of this curve determines the mean and variation of the twist variation modes in response to gap variation.
3.3. Monte Carlo Simulations

To illustrate the impact of these curves, an indirect Monte Carlo simulation was employed. Since the nonlinear contact simulations employed in this work are computationally expensive, it is not plausible to perform them hundreds of times in a direct Monte Carlo simulation. Therefore, in this work, the mapping curves in Figure 4 were used.

Utilising these curves, a gap input distribution with a mean of four millimetres and a standard deviation of zero point five millimetres was mapped to the resulting twist. This was performed for the sequences illustrated in Figure 4; the results can be seen in Figure 5.

![Figure 5. Monte Carlo simulation results for selected weld sequences](image)

3.3. Welding sequence selection

From the sequences simulated, it was observed that sequences that alternate between opposing sides achieved a smaller deviation and variation in the measured twist than sequences that proceeded around the frame, clockwise or anticlockwise. The twist variation mode caused by a welding sequence was also found to depend on the starting location of the welding sequence. The welding sequences starting at the clamps and working towards less supported areas achieved the smallest mean shift with less variation of the dominant variation mode.

4. Conclusion and future work

This simulation based study has demonstrated the importance of welding sequence selection for assembly of closure assemblies. The stresses induced by closing the gap between the welded edges can result in different welding sequences inducing different magnitudes of deformation. Furthermore, when an indirect Monte Carlo simulation was performed, it was observed that both a mean shift towards zero and a reduction in the dominant variation mode occur together. This was observed in all the sequences that were studied.

In selecting the correct welding sequence for a closure assembly loop, as presented in this case study, starting at the most supported areas and working towards the less supported areas minimised both the deformation and variation in the assembly.

This work forms the foundation investigations into dimensional variation of closure assemblies, in particular weld sequencing in the presence of a gap variations. It also informs future laboratory and industry based trials in this field.

Acknowledgement

The work presented here is supported by funding from the Australian Research Council (Linkage Grant #LP0560908), and in-kind contributions in the form of staff and facilities from Ford Australia.

References

Estimation of the Weldability of Wingle-Sided Resistance Spot Welding

D. Kim\textsuperscript{a}, Y. Cho\textsuperscript{b}, Y. H. Jang\textsuperscript{a}

\textsuperscript{a} School of Mechanical Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-749, Korea
\textsuperscript{b} Hyundai Motor Company, 772-1 Jangduck-dong, Hwaseung-si, Gyeonggi-do, Korea

Abstract: A single-sided spot welding technique is investigated by numerical analysis using a commercial CAE package to estimate the weldability. The reliability of the method is verified by the welding experiment performed under conditions similar to those of a real product. Several conditions are analyzed to find an optimal condition set, including the boundary condition for the ground location and the electrical contact resistance. The resulting Lobe curve with respect to welding time and current is obtained. The weldability of the single-sided spot welding specimen is estimated through the tensile strength test. A certain level of tensile strength can be obtained at the range of welding variables for the optimal nugget sizes, which supports the reliability of the single-sided RSW.

Keywords: Single-sided spot welding; Numerical simulation; Lobe curve; Experimental validation

1. Introduction

Resistance spot welding (RSW) has been widely employed in sheet metal fabrication for several decades because it is a simple and cost-effective joining method [1-4]. The conventional RSW machine uses two electrodes, upper and lower, to increase the quality of the welding. However, the use of two electrodes limits the machine’s own adaptability for RSW of an automotive body assembly. For example, a large C-type gun for a weld on the auto-body floor hinders access to the weld spots; hydroforming tubes have no weld flanges, and welding is often limited to single-side accessibility [5-7].

A new RSW was introduced to overcome the drawback of traditional RSW [8]. This welding system was designed to create a weld using single-side access with low electrode force. Using the system, spot welds were made using only single-sided access with or without a backing plate. It is expected that the practical use of this method may be successful, but the basic characteristics of the welding method are not well understood. Finite element analysis is a powerful tool for obtaining detailed information about a welding procedure such as spot welding and has been used effectively to collect such data. Recently, a numerical analysis of single-sided RSW for hydroformed tubes was reported to explain the effects of welding parameters on a nugget [9]. It may be possible to numerically model the spot welding of hydroformed tubes because the current path can be captured in finite tube geometry. However, numerical analysis of single-sided RSW welding is questionable when the dimension of the real chassis of an autobody is large compared with the nugget or electrode geometry because the numerical model requires a considerable number of meshes and computing time.

In this paper, we extend a conventional FEM model for RSW with two-sided electrodes to a model for a single-sided electrode. A feasible numerical analysis of the single-sided RSW for a real chassis structure is proposed. Weldability lobe curves are also determined in terms of welding time, current, and electrode force. Several unique performances on single-sided RSW are also described. Experimental verification is performed using a specimen cut from a manufactured chassis.

2. Numerical analysis model for RSW with a single-sided electrode

FEM analysis has been performed using a customized RSW package called SORPAS[10], which accurately estimates the behavior of RSW and can determine the conditions of RSW. Since SORPAS is basically optimized for RSW with a two-sided electrode machine with upper and lower electrodes, several numerical techniques are required to incorporate the model of single-sided RSW.

Fig. 1 shows a possible current path of a two-sided electrode model (a) and a single-sided electrode model (b). In the two-sided electrode model, the current travels a shorter distance from the upper electrode to the lower one since the path has minimum electrical resistance. Thus, it is possible to simulate RSW for a model with a limited dimension without considering the full dimension of the model. On the other hand, the single-sided electrode model generates a current path between an electrode and the ground that is far distant from the electrode and which varies according to the shape of the weldment. Thus, the shape of a weldment is a key element in single-sided RSW. Thus, in order to properly simulate RSW, the weldments should be the same shape. However, since it is impossible to consider a full scale model in RSW simulation, an approximated model is sought to appropriately simulate the single-sided RSW.

![Figure 1. Two different RSW models: (a) two-sided electrode and (b) single-sided electrode and the corresponding possible current paths.](image-url)
merit of single-sided RSW. Thus, we conclude that the axisymmetric model is suitable for single-sided RSW.

The single-sided RSW forms various current paths according to the real shape of the weldment. Thus, it is difficult to obtain the same current density distribution as in the real case by using a limited analysis model which does not consider the full scale of the welding system. In order to obtain a current density distribution close to that of the real case, we have attempted to find an approximate ground connection. Fig. 3 shows two of several ground connections. The ground is connected to the lower weldment in Fig. 3(a) and is connected to both the lower and upper weldments in Fig. 3(b). Note that we use the ground connection of Fig. 3(a) for the experimental setup. Applying FEM analysis, the resulting current distributions of the two connections in Figs. 3(a) and (3b) are obtained and are shown in Figs. 4(a) and 4(b), respectively. While the current density distribution for the ground connection of Fig. 3(b) is distributed in both the upper and lower weldments, the current density of Fig. 3(a) concentrates in the lower weldment. This is because the FEM result of deformation for the ground connection of Fig. 3(a) shows a separation at the end of the weldment. Thus, most of the current passes through the contact region from the upper weldment to the lower one, leading to the concentration of current density in the lower weldment. This induces a significant difference in heat generation between the upper and lower weldments, which can hinder welding. In the production of a welding system, contact between weldments can occur at locations other than between the electrode and the weldments. This is why the ground connection in Fig. 3(b) is appropriate for single-sided RSW.

In the SORPAS program, electrical contact resistance is defined as

\[
\rho_{\text{contact}} = \frac{3}{2} \frac{\sigma_{\text{soft}}}{\sigma_1 + \sigma_2} \left( \rho_1 + \rho_2 + \gamma \rho_{\text{contaminants}} \right)
\]

where \( \sigma_{\text{soft}} \) is the flow stress of the softer of the two metals in contact, \( \sigma_i \) is the contact normal pressure at the interface, \( \rho_{1,2} \) are the resistivities (with subscripts 1 and 2 indicating the two base metals in contact, respectively), \( \rho_{\text{contaminants}} \) is the surface contaminant resistivity due to oxides, oil, water vapour, and dirt, and \( \gamma \) is a factor introduced to adjust and verify the contact resistance. Since this equation has been optimized for the two-sided electrode model that has symmetric pressure distribution between electrodes, the contact resistance should be changed for the single-sided RSW model. In addition, the applied force of the electrode in the single-sided model decreases to about 1/10 of the force of the two-sided model. Thus, this force variation increases the electrical contact resistance between weldments. In this software, the contact resistance can be changed by adjusting \( \gamma \). The effect of contact resistance is investigated in the section on verification.

![Figure 2](image)

**Figure 2.** Two different models for RSW with one-sided welding: (a) block model, (b) axisymmetric model.

In order to verify numerical simulation results with experimental results, welding was carried out with a medium frequency direct current (MFDC) welding machine controller. A single-sided welding gun with a pneumatic cylinder was used to apply the electrode force. The gun was designed for a maximum electrode force of 150 kgf with twin piston rods. A truncated dome-type electrode tip was used with an outer diameter of 16 mm and tip-end diameter of 6 mm. Various tip-end shapes were tested to generate a sound fusion area between faying surfaces of the sheets. Two kinds of materials were welded with the same material of two-sheet combination. One was a low-carbon steel called SPCC with a thickness of 0.7 mm and tensile strength of 270 MPa. The other was a high strength, low-alloy steel called SPRC with a thickness of 1.2 mm and tensile strength of 340 MPa.

The variation in nugget geometry from experiment results has been compared with that of the numerical results. In the case of the one-sided electrode model, the welding experiment must be performed using the same geometry as that of a mass-produced model to generate the same current distribution. For this purpose, a section from a mass-produced model was used in this research.

The geometry of the numerical analysis model differs from that of the experiment model, and the shunt effect cannot exist in the numerical model. However, a certain level of approximate current density generation is made possible by the previously presented methodology.

If the difference between experiment and numerical results is recognized, a more valuable numerical model can be
inferred. For example, from the results in Fig. 5, we predict that the experiment result generates a larger nugget than does the numerical analysis model.

![Figure 5. Variation of nugget size according to the gap distance between lower and upper grounds.](image)

4.1. Effect of variation of the ground connection

We have mentioned the ground connections located at both the upper and lower weldments. This connection method is contrived from the idea that the location of the ground connection affects the current density distribution of the weldment, which could be a substantial factor in optimizing the current distribution. According to the ground connection methods suggested in the numerical model, the formations of the nugget geometry are shown in Fig. 4. For the ground connections of Figs. 3(a) and 3(b), the nugget geometries of Figs. 4(a) and (b) are obtained, respectively. Each ground connection provides isosceles trapezoidal nugget formation, but the lower angles of the trapezoidal nugget for the cases of Figs. 4(a) and 4(b) are acute and obtuse, respectively. Through the nugget geometry obtained from the experiment shown in Fig. 8, we confirmed that the nugget formation of Fig. 4(b) is more appropriate and the ground connection of Fig. 3(b) is the appropriate one, as we anticipated in the modelling.

Another interesting result is that the nugget sizes differ according to the relative distance between the upper and lower grounds. Note that we only consider the case in which the upper ground is always situated on the right side of the lower ground. Fig. 5 shows the nugget size variations according to the two ground connections. These results indicate that, when the relative distance increases, the slope of nugget size variation in the direction of weldment thickness decreases. For the comparison of the nugget size obtained from the experiment, the optimal distance between the upper and lower grounds is chosen, as well as the electrical contact resistance.

The adjustment of electrical contact resistance to single-sided RSW is performed using the factor $\gamma$. Since the applied force of the electrode in the single-sided model is smaller than that of the two-sided model, the electrical contact resistance for the single-sided model is greater than that of the two-sided model. We find that when we change the factor $\gamma$ from 0.5 to 9.0, the corresponding nugget size in the single-sided model increases according to $\gamma$.

The optimal welding conditions for the numerical analysis were chosen by selecting the relative distance between the upper and lower electrodes and the factor of electrical contact resistance and comparing the nugget size and the angle of nugget geometry with the experimental results. Fig. 6 shows the nugget size variation according to the relative distance between the upper and lower electrodes and the factor of electrical contact resistance. Based on the results, the distance between the upper and lower electrodes was set at 1.2 mm, and the factor of electrical contact resistance was 9.0.

![Figure 6. Variation of the nugget geometry according to the relative distance between the upper and lower grounds and the factor of electrical contact resistance.](image)

5. Verification of numerical method

5.1. Comparison of experimental results

The simulation results for the variations of current and time are compared with the experimental results by applying the boundary conditions to the numerical model. Fig. 7 compares the nugget shape results between the experimental and simulation results according to current and welding time. Although the nugget shapes and sizes are similar to each other, the simulation does not reproduce the exact shape or size of the nugget. However, it allows the possibility of a proper boundary condition so that the selection of boundary conditions may require careful consideration at the initial stage of the selection. The results of the nugget indicate that we must determine the nugget shape using the simulation results.

5.2 Effect of the electrode geometry

In the two-sided RSW, two types of electrodes, round or flat tipped, are generally used. These electrode geometries induce different contact pressure distributions at the faying surface and the bulk region of a weldment, thus affecting current density distribution and nugget geometry. In particular, the two-sided electrodes generate sufficient contact between weldments. However, single-sided RSW uses just one electrode on one side, so the applied force does not produce a good contact between weldments. Thus, single-sided RSW depends on electrode shape.

The limitation of single-sided RSW is that there is no support on the opposite side of the applied force, inducing insufficient contact at the faying surface. Thus, single-sided RSW requires a higher contact pressure at the weld interface. The magnitude of contact pressure is higher for the round tip...
is greater than \( t_{\text{min}} \), where \( X_{\text{max}}, Y_{\text{max}} \) are the maximum dimensions of the nugget in the X and Y directions, respectively, and \( t_{\text{min}} \) is the minimum thickness of the weldment. This criterion is based on the nugget shape in the two-sided electrode RSW that generates symmetric nugget geometry. However, since the nugget shape of single-sided RSW is asymmetric, the optimal nugget criterion cannot be followed in the one-sided electrode model considering an asymmetric characteristic by replacing \( t_{\text{min}} \) with \( t_{\text{opt}} \), which is defined as the nugget size at the contact interface between two weldments. Adopting the new criterion, proper sizes of the nugget are estimated and shown in the green region of Fig. 9. The red region presents the expulsion region, which increases by 40% of the optimized nugget size.

5.3 Estimation of Lobe curve

The range of optimized variables has been estimated. The criterion for the optimal nugget is generally that \((X_{\text{max}} + Y_{\text{max}})/2 \) is greater than \( 4\sqrt{t_{\text{opt}}} \), where \( X_{\text{max}}, Y_{\text{max}} \) are the largest dimensions of the nugget in the X and Y directions, respectively, and \( t_{\text{opt}} \) is the minimum thickness of the weldment. This criterion is based on the nugget shape in the two-sided electrode RSW that generates symmetric nugget geometry. However, since the nugget shape of single-sided RSW is asymmetric, the optimal nugget criterion cannot be followed in the one-sided electrode model. Thus, we need to modify the criterion of optimum nugget geometry of a one-sided electrode model considering an asymmetric characteristic by replacing \( t_{\text{min}} \) with \( t_{\text{opt}} \), which is defined as the nugget size at the contact interface between two weldments. Adopting the new criterion, proper sizes of the nugget are estimated and shown in the green region of Fig. 9. The red region presents the expulsion region, which increases by 40% of the optimized nugget size.

5.4 Measurement of tensile strength

In order to estimate the weldability of the single-sided RSW specimen, tensile strength is measured by the tension test. Although the test shows that the breakage of the upper and lower plate is irregular, a consistent outcome for the tensile strength can be obtained. Table 1 shows the tensile strength of the welding specimen according to the welding time and current. The tensile strength recommended by the automotive industry is greater than 200 kgf. The tensile strength test result indicates that most combinations are satisfied with this criterion, assuring the reliability of the single-sided RSW. According to the welding variables in Table 1, a certain level of strength (360kgf) can be obtained at the range of welding variables for the optimal nugget sizes. This is another result that supports the reliability of the welding method and the merit of the single-sided RSW.

### Table 1 Tensile strength according to welding current and time.

<table>
<thead>
<tr>
<th>Current (kA)</th>
<th>Time (sec)</th>
<th>Tensile strength (kgf)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.6</td>
<td>0.10</td>
<td>321</td>
</tr>
<tr>
<td>6.6</td>
<td>0.15</td>
<td>351</td>
</tr>
<tr>
<td>6.6</td>
<td>0.20</td>
<td>381</td>
</tr>
<tr>
<td>7.2</td>
<td>0.10</td>
<td>353</td>
</tr>
<tr>
<td>7.2</td>
<td>0.15</td>
<td>397</td>
</tr>
<tr>
<td>7.2</td>
<td>0.20</td>
<td>426</td>
</tr>
<tr>
<td>7.8</td>
<td>0.15</td>
<td>424</td>
</tr>
<tr>
<td>8.4</td>
<td>0.10</td>
<td>398</td>
</tr>
</tbody>
</table>

6. Conclusions

A welding technique single-sided RSW is investigated by numerical analysis using a commercial CAE package to estimate the weldability. The reliability of this welding method is verified through the experiment performed under conditions similar to a real production, showing an asymmetrical nugget which contrasts with two-sided RSW. Several conditions are analyzed to find an optimal condition set, including the boundary condition for the ground location and the electrical contact resistance. The resulting Lobe curve with respect to welding time and current is obtained. Through the tensile strength test, the weldability of the single-sided RSW specimen is estimated. A certain level of tensile strength (360kgf) can be obtained at the range of welding variables for the optimal nugget sizes. This is another result that supports the reliability of the single-sided RSW.

References

RMS design methodology for automotive framing systems BIW
A. Al-Zaher, Z. J. Pasek, W. ElMaraghy
Industrial and Manufacturing Systems Engineering, University of Windsor

Abstract: Today, markets increasingly require more customized products, with shorter life cycles. In response, manufacturing systems have evolved from mass production techniques, through flexible automation and mass customization, to produce at mass production costs. Manufacturing facilities must incorporate more flexibility and intelligence, evolving towards reconfigurable manufacturing systems. RMS is amid to possess such flexibility and responsiveness and said to be the next generation of world class systems. RMS is designed for rapid change in structure and for a quickly adjustable production capacity [1]. This paper proposes a novel methodology (High level process) of framework using RMS principles for automotive framing systems as well as to provide a guideline to support the structure of different stages of the design methodology. The proposed methodology is presented through a case study using data based on actual production systems of three different styles; (process and design data) which supports the hypothesis of the research.

Keywords: Reconfigurable manufacturing systems, Automotive framing systems, Design methodology, Digital manufacturing

1. Introduction
In today’s global economy, with open and unpredictable market changes; new enablers are needed in both product development and manufacturing systems. Car makers, particularly those of North American origin, explore economies of scale and scope through globally shared features of their products, for example: (1) common vehicle development (2) collaborative engineering processes (3) unified manufacturing processes. Manufacturing enterprises are forced to reassess their production paradigms, so that their products can be designed to maximize potential achievable variants, while the manufacturing systems to make them can be designed to operate efficiently by robustly accommodating future product changes, minimizing time to market and providing a reliable production base. Manufacturing facilities have to possess a high degree of flexibility, enabling mass customization of production. Mass customization is the capability of a firm to offer high product variety, high overall volume and at the same time, low cost and fast delivery.

Multiple models can now be assembled on a single production line. Typically, more than 80 percent of the tooling and equipment in a body shop are not specific to an individual model but can be used for all models produced. The production line can be programmed to weld various models, such as sedans, or SUVs. In a paint shop, all equipment are programmable to cover all styles efficiently and cost effectively. In the final assembly, all major automakers are building more models, derived from global platforms, on the production lines that can simultaneously handle multiple products allowing for efficient utilization of people and equipment. Fig. 1 shows the automotive framing process, consisting of a common under-body complete (Body-In-White, BIW, platform), followed by the inner/outer framing. The inner framing, creating the car body structure, together with the outer metal panels (skin) define the vehicle styling, which in turn are tied to vehicles’ unique visual features.

This paper focuses on the production systems and attempts to identify flexible elements in product variants of a product family design and their map to the manufacturing process and to the production systems. The hypothesis is that if the right subset of car body elements is designed with proper care for future flexibility, then the proposed corresponding production system can then better accommodate body styling changes, enabling structural flexibility in the gate design supporting increased number of body variants produced without the need for tooling changeover, which improves throughput of the systems.

Fig. 1: U-Body BIW Platform and Vehicle Styling

Reconciling continuously evolving product designs and expanding styling variants, including new materials and new processes, with the throughput required for profitable car assembly plant operations (3-shift, mixed model production on 240 days a year, assuring annual throughput of 360,000 units or hourly rate of 62-70 units) poses a great challenge.

Section 2 discusses the current practices and trends in the automotive framing systems. Section 3 presents the (proposed) RMS framework (methodology) of Automotive framing systems. Section 4, a Case study of Automotive framing Systems Using the RMS Methodology is presented. Section 5 discusses the findings and results. Section 6 summarizes the findings and outlines the future work.

2. Current Practice of Automotive Framing Systems
‘Framing system’ is a process and the related infrastructure for a precise positioning and securing under-body platform with the upper body components [6]. The work discussed here is focused on the body framing systems, which are viewed as critical in terms of assuring proper geometry and structural integrity of the automotive vehicle body. Most passenger
vehicles made today have a (structural) body that comprises 100–150 stamped metal panels, assembled into 18-22 modules (sub-assemblies). These modules are assembled into 3 systems to create the vehicle body formed (prior to painting, is referred to as body-in-white) in a multi-step manufacturing process, during which the modules are joined together by welding. In order to assure the final body geometry according to design specifications, the panels to be welded are positioned (by pins and NC blocks), constrained by clamps and presented to the process by highly complex, automated fixture devices. The spot welds are typically laid out by robotic arms with multiple degrees of freedom, carrying the welding guns to execute the joining process [8]. Overall, a vehicle framing system can be divided into three subsystems, as shown in Fig. 2: (i) Under-body platform (consisting of up to 10-12 sub-assemblies), which needs more than 120 robots for assembly. (ii) Inner framing (5-9 sub-assemblies added prior to framing), which needs more than 30 robots, grouped into up to 4 assembly cells. (iii) Outer framing (3-5 sub-assemblies added prior to framing), which needs more than 30 robots, grouped into up to 4 assembly cells.

During the framing process, all modules have to be position-ally secured with respect to the under-body-platform and inner/outer skin, as only then the welds attaching the roof skin can be placed. In most production systems, the time required for all these activities is set between 45 to 56 s, depending on the vehicle size (smaller times for smaller bodies). The time required for framing also determines the cycle time (CT) of the whole line, and hence causes a bottleneck operation. Historically, all current framing systems have evolved from the Open-Gate and Robo-Gate system, initially developed in the 1980’s. The Open Gate Framing (see Fig. 3) accommodates the use of multiple, exchangeable, dedicated gates. While only a single gate (fixture) can be in operation at any given time.

When a body style change is required, the changeover can be performed in an average amount of time, 30 minutes (loss of production at about 40-50 job). Otherwise, the operation of the gate is the same as described previously for Robo-Gate. The cost of lost production due to the changeover is about 300,000 USD [13].

The key issues with the current framing systems can be summarized as follows: (1) A strongly coupled design; (resulting in complex coordination of motion for some work units moving after the gate in the working position; increases both mechanical complexity and overall weight). (2) Changeover times must be accommodated in the production plan, since the line has to stop running. (3) Gates storage systems require a significant amount of floor space. (4) Overall high cost, high lead times for engineering and build time. (5) Very expensive to manufacture and maintain. (6) High cost of lost production due to downtime (breakdown or retooling). (7) High risk in capital investment. Fig. 4, shows the proposed new vehicle framing systems using Reconfigurable Open Gate to address all or most of the issues mentioned above.

3. Life Cycle and Framework for the Reconfiguration of Framing Systems with TC Support (proposed methodology)

Manufacturing system life cycle: refers to the evolution of a manufacturing system from concept through development to production, operation and its ultimate disposal [14]. The manufacturing system design framework is proposed to support the reconfigurable manufacturing system design methodology execution and to clarify communication and collaboration among the design team. The framework provides a guideline to support and to structure the different stages of the design methodology. This framework is mainly based on system lifecycle concept. Fig. 5, shows the main 4 stages of proposed framework; first, 3 transitional stages and lastly, the parallel stage: (1) Mfg systems analysis, (2) Mfg systems design, (3) Mfg systems Op & maintenance, and (4) is the refine offline gate combined with manufacturing support center (Teamcenter), brief explanation for each stage in sec. 3.2.
3.1 Motivation for the proposed methodology

The newly proposed framing systems should address the issues with the current systems (state of practice) and fulfill the needs of the market dynamics: (a) Robustly accommodate growing number of product variants within a family without change of Gates, (b) Utilize standard components (platform) to reduce cost, improve maintainability and quality, (c) Extensively utilize math-based engineering concepts (to validate the production process prior to physical implementation) and PDM/PLM teamcenter to facilitate broad engineering collaboration, (d) Develop and enhance new manufacturing strategies and operational tactics.

3.2 The RMS framework of Automotive framing systems

The design process is considered a problem solving process approach, which is a widely accepted approach, performed by the Analysis, Synthesis and Evaluation [16] method. The first stage, diagnoses, defines and prepares the information about the problem to be solved; the second one synthesizes possible solutions; and lastly, the evaluation stage tests the possible solutions against the goals and requirements.

Stage 1: Manufacturing systems analysis

The manufacturing systems analysis is the first stage of the life cycle where the formulation and definition of the manufacturing system is performed to satisfy specific needs. The main constraints at this stage are the manufacturing strategy, the characteristics of the product and the process. Therefore, the output of this stage is the definition of an information model that represents and captures information describing the product characteristics, process, manufacturing resources and strategies of the manufacturing system. The automotive framing systems with complex product and processes, systems designer and product developer need to have a perfect knowledge of the decomposition and integration of all the modules and components of each module (interface components and their process) in order to upgrade to new vehicle styling.

Stage 2: Manufacturing Systems Design (PD)

At this stage, more focus is given to the manufacturing process and production systems than to the products. The manufacturing system design is the second stage of the manufacturing system life-cycle. The main elements of this stage are: (1) the function requirements translated to systems design variables (SDV), (2) Process design (conceptual design), (3) System design detailed design (DT), (4) the production and operation. The main inputs for this stage are the requirements of the manufacturing system in terms of reconfigurability, which is the result of the assessment.

Once the design parameters are linked to product’s key critical elements (PKECs) and grouped into key characteristic control (KCCs) with clustering process applied to establish the DOF of each group, flexibility and capabilities of manufacturing systems can be defined [15]. Thus, production and operation is a set of models where information of manufacturing systems represents how the manufacturing system will operate. The models are logic, information and virtual validated through the use of virtual manufacturing tools that support the design and reconfiguration of manufacturing systems. Virtual design of manufacturing systems with Teamcenter manufacturing supports is set to be the future hub of the framework. This set to provide, (a) Instant information, more detailed of production systems and the operation process, (b) Instant evaluation for different alternatives of joining processes, controls and layout to satisfy the best choice according to the defined objectives.

Stage 3: Manufacturing systems operation & maintenance

The third stage of the framework is the implementation or the launching of the manufacturing systems. Once the manufacturing systems are operating, it is important to establish operational matrices aligned with the design objective and the performance of the production line rate, breakpoint in the production by changes in demand, introduction of new products, and engineering changes in product, among others. Using DM with a platform base as to support and integration tools (Teamcenter) is the place where all the date synchronized from the real production line to the virtual manufacturing systems.

Stage 4: Refine offline gate

During the design activities, more detail is needed. The manufacturing characteristics such as product, operations, processes and alternatives of layout are designed. The control of the manufacturing system is designed and the human and technological resources are identified. More data is needed for the process transportation and zoning of robots during the execution of assembly process, however it is impossible to account for it in the modelling stage. Thus, using Offline Gate as a parallel process for testing new controls, a new process prior to implementing the real production once it is tested and proven can then be added to the virtual processes.

3.3 The RMS Design Methodology of Automotive framing systems

Finally, the methodology is the integration of these four stages by the RMS Design framework which is proposed to decompose each stage into activities to analyze, evaluate and synthesize the inputs and outputs of each stage in order to design/reconfigure a manufacturing system. Fig. 6 shows the main stages of the methodology with PDM/ DM (Teamcenter manufacturing support) base on the proposed framework. On the integrating side not only design, techniques and manufacturing are easily coordinated, but also imitation analysis and optimization of the manufacturing process in the virtual environment become more convenient than ever.

Research has shown that a prerequisite to innovate and to successfully develop complex systems is systems’ understanding, which can be increased through the use of representative models. In this methodology [5] DSM is used, which can highlight the interfaces or interactions between system elements by following these steps: (a) Decompose the system or sub-systems into main elements; (b) Understand the interactions between modules or elements of the systems; (c) Analyze potential reintegration of the elements via clustering (integration analysis).

Design Structure Matrix (DSM - Component-Based Architecture) is used in the automotive industry as a tool and technique. It provides a simple, compact, and visual representation of a complex system that supports innovative solutions to decomposition and integration problems. It is also an effective method for integrating low-level design processes based on physical design parameter relationships. Furthermore, it displays the relationship(s) between modules or components
models represent extensive system knowledge. Hence, DSMs can be difficult to build, especially initially, as they depict data that are not always at hand, easily gathered, or quickly assimilated.

4. A Case study of Automotive Framing Systems Using the RMS Designs Methodology:

The below case study is based on actual production data of three styles for different vehicles; the case study used 2 styles (Similar U-bodies for S1 and S2). The three workcells capable of running the production with changing gate using storage systems (see Fig. 3). The issues of current systems were discussed in Section 2, the proposed RMS with Open Gate to be used; one gate without need to change over (eliminate down time 30 minutes/per change over). The redesign of current systems (Current systems is coupled; dynamic interaction caused quality issues and unexpected breakdown. Analysis of tooling design lead to modification of selected units to make them identical for both gates systems (proposed system layout; Fig.4). Most automotive case studies started with existing production systems; the objective or activities at this stage is to (1) identify the states of current manufacturing systems for introducing a new product; (2) Perform Reconfigurability assessment for the manufacturing systems, (3) Decompose the manufacturing systems to define the structure of the systems and the capability of the manufacturing systems, (4) determine boundaries set of uncertainties more focus on manufacturing processes (Geo. Station; Gates, robots, End of arm tooling) to have the right manufacturing capabilities for the additional styling without switching equipments.

Motivation

High cost in platform leads to the increase of a variety of product family platform to justify the capital investment in production equipments. The frequency of product styling is shown in Fig. 7; body styling above the beltline every year and major change every five (5) years. Usually, equipment and tooling life cycle (with the exception of control updates) run up to 20 years.

Fig. 8 shows the addition style S2 (5c common styles current style in production), changes are shown in pillar A, C and RP. Table 1 shows the state of art decomposition of the automotive framing modules.

Analyzing the current manufacturing systems:

---

**Step 1:** Identify current manufacturing systems for the new styles (S2), the current manufacturing systems running mutable-
structures of Fig. 9 are presented in Fig. 11. Data and information of CPM and can be used as an input to construct the Fig. 9: Proposed modular structure of Open Gate

Evaluation chart Fig. 12. The data and information that are presented in the Evaluation Chart summarizes the processing design of current manufacturing systems prior to detailed design to estimate cost of the new styling.

Fig. 10: Change Propagation Matrix (CPM) for BIW Pillar B, C

Hybrid DSM is used to show the coupling between DT (design tasks) and DC (columns) represent design configuration which resulted from mapping KPCs. KPCs and KCGs. The 3rd half of shows the Mfg capabilities, spec and constraints of the production line.

Fig. 11: Formulation of Hybrid DSM for BIW Vehicle Styling

The next step is the design stage of the manufacturing systems; which is presenting the execution of activities that are outlined by the evaluation chart. Different scenarios can be evaluated at this stage with 3-D data modeling; also virtual simulation is used to validate Systems performance. Fig. 13 shows the detailed design for the Open Gate station (inner framing work cell) pin, CD locators and welding windows of the RH robots.

Stage 3: Manufacturing Systems Operation & maintenance; the third stage of the framework is the implementation of the manufacturing systems. Once the manufacturing system is operating, it is important to establish operational matrices aligned with the design objective and the performance of the

Stage 4: Refine offline-Gate

Using offline Gate as a parallel process for testing new controls, new process prior to implement in the real production line rate; by breakpoint in the production by changes in demand, introduction of new products, engineering changes in product, among others. In the automotive industries, on one hand, the time required for product developments is reduced significantly due to eliminating a number of uncertainty engineering changes. On the other hand, system integrations and debug taking longer time due to many reasons, such as a lack of synchronization between virtual productions and real production lines for product and production systems. Usually, a production system takes months to optimise equipment and to reach the steady state of production. All the changes are implemented back in the original model and one single source of future changes is kept. The challenges are how much flexibility is needed in the tooling and equipment of the manufacturing system (manufacturing capabilities), such as degree of freedom in the modules, weld guns forces, robot load and many more. The solution is knowledge and expertise, which can be presented in a low level of the new proposed framework (future work).

Stage 4: Refine offline-Gate

Using offline Gate as a parallel process for testing new controls, new process prior to implement in the real production
once the new process is tested and proved, then can be added to the virtual process. The offline-Gate serves as spare modules.

5. Results and Discussion

Running the simulation program Delmia IGRIP using the three modules structures (Preliminary 3D conceptual geometry models) shows 10% time saving in line rate; This results is 5 seconds (10% of the 52 CT) due to two components (1) dumping units eliminations and (2) reduction of gate size in heights. This time can be used to increase the time available for all robots in the three work cell. More results can be disconved in future work.

Once the initial data and information (CPM- Change propagation Matrix) DSM and HDSM model are built, they can serve as a knowledge base or platform for continued learning, improvement, and innovation. Teamcenter is used as manufacturing supporting center to quickly make data and information quickly accessible and available to all partners across the enterprise. The tools and equipment are built, and selected by finding experts knowledgeable about each activity and eliciting their expert opinions by filling in the rows and columns of the DSM and HDSM then propagation charts can be produced quickly. By completion of these matrices, the input and output of each stage’s activities becomes available. Fig. 15, shows and summarizes the high level input/output activities of four stage of the RMS framework.

The contents of these matrices (data and information) become a common and standard practice of the proposed framework. However, simply reaching this understanding can benefit a product development, systems manufacturing and the entire enterprise for future developments and training to be less depends on experienced leader in the field.

The new proposed design structure (modular approach) for the open-gate framing systems will eliminate the coupling design and enable increases of mixed variants within a family without the need for switching gates during production. This concept may lead to a new method of gate construction, allowing a reduction in the heavy tubing and unnecessary heavy dumping units, reducing at least 30% of the gate weight. PLM/PDM with team center support helps developers and designers to understand a product in its entirety, including the organizational processes to plan, develop, manufacture. The key tooling beltline concept was investigated for the first time [7]. Product and process are closely coordinated to achieve optimal matching of requirements for effective cost, quality, and delivery.

The future work will aim to: (a) Include simulation results. (b) Validate the flexibility with non-traditional techniques of joining process to achieve the design objectives, if needed. (c) Justify the initial cost of flexibility needed upfront to achieve production flexibility by using a reconfigurable line.

References

[13] Personal communication with Chrysler Canada Windsor plant manager, 2011, April 23
Assessing the structural complexity of manufacturing systems layout

V. Espinoza, H. ElMaraghy, T. AlGeddawy, S.N. Samy
Department of Industrial and Manufacturing Systems Engineering, University of Windsor, Windsor, On, Canada

Abstract: The layout of a manufacturing facility/system not only shapes material flow pattern and influence transportation cost, but also affects the decision making process on the shop floor. The layout of manufacturing systems determines its information content which determines its structural complexity inherent in the layout by virtue of its configuration design. This paper introduces a new method, which assesses the structural complexity of a manufacturing systems layout. Six complexity indices, based on the structural characteristics of the layout such as its density, paths, cycles, decision points, redundan
dy distribution, and magnitude are introduced. These indices are directly affected by the information content, inherent in the system layout configuration. Those individual complexity indices are combined into an overall measure using the radar chart technique. A case study, based on a facility layout from literature, is analyzed. The layout complexity index is a measure to compare system layout alternatives and make decisions to select the least complex layout.

Keywords: Structural complexity, Manufacturing system, Information content.

1. Motivation

Manufacturing companies operate in an uncertain and constantly changing environment. That uncertainty is driven by change in customer demands, product design and processing technology. Uncertainty in manufacturing systems also increases complexity. Different types of manufacturing systems complexity have been identified in literature, such as structural, dynamic, internal and external complexity [1] [2] [3]. Despite the attention received by researchers in measuring structural complexity [4] [5] [3], systems layout has not been included in structural complexity assessments.

The objective of this research is to assess the structural complexity of manufacturing system layouts by defining its characteristics and patterns that contribute to the information content caused by the decision making process in the system layout. A system layout is the arrangement of machines and the interconnections among them [6]. Additionally, system layout also considers the number and the type of modules (i.e. machines, workstations, transporters, etc.) and their locations and relationships, which defines the flow of work pieces between them [7].

This paper proposes a methodology which converts the physical system layout to a graph representation, in order to produce measurable complexity indices, based on the number and locations of decision making points within the layout. These indices are useful at the early design stage and facilitate comparing and evaluating layout alternatives, and identifying potential structural problems.

2. Complexity in manufacturing systems

Complexity in manufacturing system can be classified into static [1] or dynamic [2]. Measures for both classes can be found in literature. Frizelle [1] defined static complexity as the expected amount of information necessary to describe the state of a manufacturing system. It is based on the probability of a resource occupying a certain state. Dynamic complexity was defined as the expected amount of information necessary to describe the state of a system deviating from schedule due to uncertainty [2] [4]. The calculation involves the measurement of the difference between actual system performance and expected performance from the schedule.

Kim [5] used a heuristic approach to quantify the operational system complexity. He proposed a series of system complexity metrics, based on the relationships between system components, number of elements, and the complexity of each element. However, the relative importance of those individual metrics was not discussed nor was their combination into a single system complexity index. ElMaraghy [7] developed a structural classification and coding system that captures the structural complexity inherent in various types of equipment in manufacturing systems, as well as their layout. This chain type of structural classification code has been extended to include the assembly specific structural features of equipment used in assembly systems, such as assembly machines, transporters, buffers, part feeders, and handling equipment [8].

Gabriel [3] investigated internal static manufacturing complexity (ISMFC), based on product line complexity, product structure and process complexity components. However, his complexity measure did not consider layout, arguing that it is difficult to quantify layout complexity, because it does not have any evident quantifiable elements. A system configuration can become complex, due to the number of machines, and their arrangements and connections. Koren et al. [9] studied and represented system configurations, such as serial, parallel, and hybrid machines as shown in Figure 1, where squares correspond to machines. Koren analyzed the number of configurations and its effect on system performance in terms of reliability, productivity, quality, and scalability.

Hu et al., [10] introduced a measure of how product variety induces manufacturing complexity in assembly systems and supply chains, by developing models to characterize the propagation of complexity in multi-state mixed-model assembly systems and multi-echelon assembly supply chains. They defined a complexity model based on product variety including station level complexity, system level complexity, and supply chain complexity. These models of complexity were used to configure assembly systems and supply chain to ensure robust
performance by mitigating complexity, but did not model the system layout complexity.
Unlike previous studies of manufacturing systems complexity, this paper uses layout nodes to represent the points where decisions are made on the shop floor and their connections. The development of this representation will be explained in the next section.
The proposed model focuses on the structural characteristics of the manufacturing system layout at the early design stage, for instance, when deciding the flow of the system. The proposed complexity model does not assess complexities arising from the system dynamic behaviour during operation.

3. Systems structural complexity assessment
Manufacturing systems have plenty of components and subsystems, with many interactions and relationships which increase the complexity of the manufacturing systems and its information. The design of the manufacturing system does not only affect the performance (i.e. productivity, throughput, and quality) [11] [12], but also the structural complexity of the system (i.e. the number and type of machines or connections between them) [3].

In axiomatic design, information and complexity are defined only relative to what we are trying to achieve and/or want to know. Suh [13] defined complexity as a measure of uncertainty in achieving the specified functional requirements (FRs). Therefore, complexity is related to information content, which is defined as a logarithmic function of the probability of achieving the FRs. The greater the information required to achieve the FRs the higher is the information content, and thus the complexity.

The proposed layout complexity model incorporates the information content, represented by the characteristics of the manufacturing system layout. This ultimately, affects the decisions made and expressed in the complexity indices.

4. Methodology
This section presents the methodology to assess the structural complexity of manufacturing system layout. The IDEF0 model of the methodology is shown in Figure 2. The proposed methodology is described in three steps as follows:
1. Analysis of the layout based on workstations adjacencies and decision making points. According to the analysis, a diagram representation is built with nodes and arrows representing the decisions made, and the direction of the flow respectively.
2. Adjacency matrix is created to capture the relationships between nodes in the diagram representation.
3. Layout complexity assessment is based on the quantification of characteristics (i.e. connections, paths, cycles, number of nodes, and number of redundancies) captured in complexity indices. Consequently, an overall complexity index is calculated to reflect the information content of the system.

4.1 System layout analysis
The purpose of this step is to identify decision points and directions of the material flow, and represent them by nodes and arrows, respectively. Decision points where there is necessary decisions regarding the flow of the material are identified. For instance, a choice of one of several alternate routes for the next process is a decision to be made. Then, a decision point is identified and placed in the graph representation. The arrows represent the direction of the material flow (forward or backward) that exists among nodes.

Figure 1. Three machine configurations: a) serial, b) parallel and c) hybrid

Figure 2. IDEF0 model of proposed methodology for assessing the structural complexity of manufacturing systems layout

Figure 3. Physical layout of a product differentiation [14]

Systems input nodes symbolize the places where the material flow starts and output nodes symbolize exits for the material. The representation of input and output nodes in the diagram helps identify a path. Figure 4 shows an example of the
4.2 Adjacency matrix creation

The Adjacency Matrix ‘AM’ is created with the nodes of the diagram representation, corresponding to the column and row headings. The values of the matrix elements correspond to the arrows on the graph representation. If two nodes are connected, then the value is “1” otherwise it is “0”. The sequence of the node placement in the matrix commences with the input nodes and ends with the output nodes. A square matrix of size $n \times n$ is created, where $n$ is the number of nodes. An AM matrix is created for the diagram example illustrated in Figure 4 as follow.

4.3 Complexity indices

Six complexity indices related to nodes and arcs from the graph representation are defined. Subsequently, a method of quantification is proposed for each index individually. The complexity indices are defined and then integrated into an overall layout complexity index. The proposed complexity indices are: density, paths, cycles, decision points, redundancy distribution, and magnitude as illustrated in Figure 5. These complexity indices aim to measure information content, which increases or decreases the difficulty of making decisions regarding the flow of material in the system layout. The density index is calculated by equation 1.

$$\text{Density index}$$

$$D = \frac{k}{n(n-1)}$$

Where: $k$ is the actual number of connections and $n$ is the number of nodes.
\[ DS = 1 - \frac{SP}{LP} \]  \hspace{1cm} (5)

Where: \( SP \) is the number of nodes on shortest path, and \( LP \) is the number of nodes on longest path.

**Redundancy Distribution Index**

Redundancy refers to the repeated or duplicated connections that exist between two nodes. Distribution of redundancy refers to the occurrences of redundancy between adjacent nodes, regardless of the number of the redundancy branches. It increases the information content of the layout and consequently the complexity. The distribution of redundancy index is calculated by equation 6.

\[ RD = \frac{r}{a} \]  \hspace{1cm} (6)

Where: \( r \) is the number of occurrences of redundancies between 2 adjacent nodes, and \( a \) is the maximum theoretical number of occurrences of redundancies between all adjacent nodes.

**Redundancy Magnitude Index**

The redundancy magnitude index accounts for the number of redundant parallel arrows in the system layout. This is related to structural complexity because as the number of redundant parallel arrows increases the information content increases. The redundancy magnitude index is calculated by equation 7.

\[ RM = \frac{pr}{w} = \frac{w - a}{w} \]  \hspace{1cm} (7)

Where: \( pr \) is the total number of redundant parallel arrows; \( w \) is the total number of forward arrows and \( a \) is the number of adjacencies.

5. **Layout Complexity Index (LCI)**

The layout complexity index is the combination of the individual complexity indices into an overall complexity index. The radar chart technique is applied to integrate the complexity indices. The value of each index is marked on a radar chart. A shaded area is formed by these values as shown in figure 6.

Samy [18] developed equation 8, to calculate the shaded area formed in a radar chart.

\[ a = \frac{1}{2} \left[ (C_t + C_i) + \sum_{i=t-1}^{t} (C_i + C_{i+1}) \right] \sin \left( \frac{360}{t} \right) \]  \hspace{1cm} (8)

Where \( a \) is the shaded radar area, \( C_i \) is the normalized value on the radial axis of index \( i \) and \( t \) is the total number of indices in the radar chart.

6. **Case study**

This layout was presented by Kim [5]. The facility is a plant that produces anti-lock break system units (ABS) for automotive companies. The presented layout produces the ABS family 5.3 with two models. The initial ABS physical layout is depicted in Figure 7. Within the ABS family 5.3 different features vary according to the customer demand. Two variants are built: the anti-lock breaking system (ABS) or anti-lock breaking with traction control (ASR). This product variation requires flexibility. First, the raw material is stored in a warehouse, then parts are moved to the receiving area of the plant. Later, containers are pulled from the machining area for production. Automated Guided Vehicles (AGVs) are used to send the containers to one of the 7 machining cells.

![Figure 7. ABS production layout](image)

After being machined, the parts are moved to one of the 8 deburring machines by AGVs. The deburring operation, which removes burrs from the hydraulic circuits, is done automatically by using a high pressure water jet. Then, the parts are moved by a conveyor belt to the washing and drying machines. Parts are inspected manually after washing and are repacked into containers. These bins are moved to a buffer area by AGV, and bins are retrieved by assembly lines. The assembly tasks are performed in a U-shaped cell. Finished parts are manually taken off the conveyor and sent to the shipping area. Final inspection is performed to return products that do not meet specifications to the start of the assembly line.

The physical system layout is analyzed and the decision points and material flow directions are identified. Subsequently, a diagram representation is generated as shown in Figure 8. In the diagram, node A represents the decisions made to send the material to any of the seven machining cells, since there are no fixed routes.

![Figure 8. ABS system layout representation](image)

Node B denotes the decision made to send the material from the machining cells to one of the two deburring cells. Redundancy is represented in the diagram between adjacent nodes. Node C expresses the decision to send the material to either one of the two assembly lines. Nodes E and F represent the decision at the inspection point, to release accepted products or return them to the beginning of the assembly line for re-work.

The adjacency matrix is created according to the relationship between adjacent nodes in Figure 8.
The individual complexity indices of the ABS plant layout are shown in Table 1.

$$AM = \left( \begin{array}{cccccc} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right)$$

<table>
<thead>
<tr>
<th>Index</th>
<th>Density Index</th>
<th>Path Index</th>
<th>Cyclic effect Index</th>
<th>Decision sequence Index</th>
<th>Distribution Redundancy Index</th>
<th>Magnitude Redundancy Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABS layout</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>Result</td>
<td>0.30</td>
<td>0.08</td>
<td>0</td>
<td>0.50</td>
<td>0.83</td>
<td></td>
</tr>
</tbody>
</table>

The ABS system layout has a low density index, which indicates few connections between nodes; a zero path index indicates that there are no alternative paths in the graph beside the minimum paths. Low cyclic effect index indicates few cycles in the system layout. A zero decision sequence index indicates that the number of nodes in the shortest and longest path are equal. High values of redundancy exist. The values indicate that redundancy occurs in half of the connections between nodes, and there is a high number of redundant parallel connections between nodes in the system layout. In this layout is recommended to reduce redundancy to reduce the value of the redundancy index.

The layout complexity index using the radar chart (Figure 9) is calculated by applying equation (8).

$$a = \frac{1}{2} \left( C_{1} \cdot C_{1} + \sum_{i=1}^{i=n-1} (C_{i} \cdot C_{i+1}) \sin \left( \frac{360}{6} \right) \right) = 0.29$$

7. Conclusions

A new approach to assess the structural complexity of a manufacturing system layout was developed. Six individual complexity indices which include characteristics of the layout, such as connections, paths, cycles, decision points, and redundancies, were defined.

These indices provide an insight into the information content inherent in the system layout. They are useful at the early design stage when developing the work stations layout configuration and selecting connections, paths, cycles and redundancies in the process.
Abstract: Due to strong competition in the automotive market, the industry is under constant pressure to achieve cost savings and thus to optimise production processes. The optimisation of the assembly process chain in particular holds great potential. In order to tap this potential, the rear axle alignment process will be analysed within the framework of the presented research project. There are essentially two optimisation approaches. One involves the point of carrying out rear axle alignment during the assembly process. The other aims at achieving greater flexibility in rear axle setting stations.

Keywords: Assembly, Reconfiguration, Process Chain Optimisation, Rear Axle

1. Identifying the Optimum Rear Axle Alignment Option

There are various possibilities for adjusting the toe and/or camber of the rear axle in the assembly process chain, from pre-assembly of the rear axle to commissioning of the entire vehicle at the end of the line. Fig. 1 shows the assembly process chain of a rear axle. At first, the individual components of the rear axle are put together on a rear axle assembly line to form an axle module. On the loading line the entire chassis, consisting of the front and rear axle, is then loaded onto a module carrier, together with the power train. After this, the chassis components are joined and bolted to the car body during the so-called “marriage” process [1]. The alignment of the rear axle may be carried out on a setting station at the end of the rear axle assembly process, on the loading line prior to the marriage-process, on the main assembly line after the marriage process or at the end of the line on a wheel alignment stand [2]. Generally, the rear axle toe values are measured at the end of the line. These toe values are taken into account for the alignment of toe values of the front axle.

The various alignment options require the assembly planners to decide at which point the alignment may be carried out in an optimum way. In order to address this issue, the entire assembly process chain (from the pre-assembly of the rear axle module to the commissioning of the vehicle at the end of the assembly line) has been recorded and analysed during this project. Within the framework of this analysis, an interview has been carried out with experts from various German car manufacturers.

At OEMs and axle manufacturers all of the options are currently used, as shown in Fig. 1. This may be explained by the fact that the optimum alignment option is depending on the boundary conditions of each case. For this reason, a method has been developed for determining the optimum alignment option.

To achieve this, all relevant criteria influencing the decision where the rear axle alignment is carried out have been identified. Subsequently, an evaluation of the interaction between the criteria and the four alignment options, (i.e. 1. setting station in rear axle assembly, 2. loading line, 3. assembly line, 4. wheel alignment stand) has been conducted.

Figure 1. Possibilities of carrying out axle alignment

One example criterion is the “Building of variants in final assembly”. It must be noted that, in some cases, axles produced by an axle assembly plant supplying various final vehicle assembly plants are initially not specified to a certain vehicle. The specification is made only at the vehicle final assembly plant. In this case, it is recommendable to carry out the rear axle alignment during the final assembly of the vehicle, i.e. during the loading line, assembly line, etc. This refers to the fact that only at this point, the vehicle-specific parameters which are relevant for the alignment (such as the ride height) can be determined.
Figure 2. Procedure for determining the optimum rear axle alignment option

In order to determine the best point for carrying out the alignment, the assembly planner has to weight the different criteria according to the boundary conditions in his case. The weighting of the criteria is conducted like in a scoring model [3]. Through the given connection between the criteria (e.g. building of variants in final assembly) and the degree of fulfilment (e.g. fulfilled by setting station, loading line and assembly line but not fulfilled by wheel alignment stand) the weighting of the different criteria lead to a weighting of the different options for alignment of the rear axle (e.g. setting station). Based on this approach assembly planners can make a sound decision with respect to the best option for carrying out rear axle alignment in the assembly process (cf. Fig. 2).

Another example for a criterion from Fig. 2 is represented by the “accessibility of the setting mechanism”. In some production series these mechanisms are difficult to access once the rear axle has been assembled to the vehicle. In this case, it is advantageous to carry out the alignment during axle assembly or on the loading line. Parallel to this example a range of criteria favours carrying out rear axle alignment in the assembly process (cf. Fig. 2).

2. Achieving flexibility in rear axle setting

Due to the increasingly turbulent production environment, setting stations in rear axle assembly are particularly affected by demands resulting from the unpredictability of production volumes and the increasing number of product variants [6]. For this reason, an analysis of the current variety of axle variants is being used for developing a concept of a reconfigurable rear axle setting station.

2.1. Current setting stations and motivation

In order to adjust the toe and camber values in the pre-assembly line, typically, automatic setting stations are used (Fig. 3). After the axle module has been assembled, the wheel geometry (toe and camber) is measured and adjusted at that point.

The setting station mainly consists of the following units: wheel hub lifting devices, clamping units, spring load units, measuring units and bolting and setting units. All of these units are affixed to a basic frame. After the setting station has been loaded with the axle, the axle module is fixed to the later car body connection points (axle clamping points). The wheel hub support units are used to take up the wheel hubs. At first, the axle is sprung with a defined suspension load (pulses) in order to minimise later occurring phenomena of technical settlement. Force is conducted via the wheel hub support units and the spring load unit. Prior to the actual measurement and setting process, the axle is brought into the compression state that corresponds to the final assembled position in the vehicle. In this load state, the wheel geometry, i.e. the toe and camber of each wheel, is then aligned by the alignment bolts.

![Figure 3. Rear axle setting station designs](image)
Currently, table and portal stations are used for axle pre-setting (Fig. 4). The machine technology varies depending on the boundary conditions of the production, which are specified by the type of axle, the given cycle time and the transport technology, among others. Against the background of the changing automotive market, particularly during the past few years, it is necessary for automotive suppliers and OEMs to use automation solutions that enable them to quickly react to changes in the production environment [7],[8]. This results in a considerable need for reconfigurable machine concepts [9], which basically take two reconfiguration dimensions into account. The machine technology should be adaptable to different product variants. Furthermore, it should also have the ability to react to dynamic production volume changes. In order to meet the demands of reconfigurability, a consistent modularisation of the functional groups for future machine technology is necessary.

![Figure 4: Classification of rear axle setting station designs](Image)

Developing a reconfigurable rear axle setting station, certain product and process characteristics affecting the machine design need to be known. One significant characteristic of designing a machine is the way in which the axle is centred during the setting process. This may be achieved via drill holes in the axle’s sub-frame or via the car body’s connection points of the rear axle. This characteristic affects the centring unit of the machine. Whereas special centring units are needed in the setting station when centring is achieved via drill holes in the axle sub-frame, centring using the car body connection points is carried out with the existing clamping units of the setting station. In order to enable cost-efficient reconfigurability using a modular machine design, the typical combination of these axle characteristics needs to be identified. This may be achieved by classifying rear axles into different classes. For this purpose a classification method has been developed which allows modern rear axles to be summarised according to assembly characteristics.

In order to achieve this, 26 rear axles from different car manufacturers and different series have been analysed and compared with each other regarding assembly-relevant axle characteristics (cf. Fig. 5). As a result, five classes were identified. Approx. 70 % of the axles were able to be assigned to these five classes.

As the axles from one class are highly similar from an assembly aspect, assigning several axles to one class considerably simplifies the formation of the assembly modules. This allows the development of class-specific modules which may be used for all axles of each particular class.

![Figure 5: Set up of a rear axle station](Image)

### 2.2. Reconfigurable rear axle setting station

In the following a method for developing reconfigurable assembly systems is presented (Fig. 6). The structure of the rear axle setting station is broken down into individual modules based on the product and process analysis and thus on the demands of the assembly task. The foundation for this was provided by the above referenced axle classification, which was the result from the effected comprehensive analysis of current axle concepts. Decisive for the definition of the module boundaries is the fact that an assembly task – a responsibility – is clearly assigned to the individual modules.

The clear definition of each module ensures the functional independence of the assembly system [10]. The functional breakdown of the machine structure allows various technical applications to be achieved for the modules. For various scaling levels, for example, manual or automated technical characteristics may be envisioned for applying the modules, and these may be added with little effort when needed. Furthermore, by taking the product characteristics into account, variable interface concepts may be provided to allow for an adaptation to various axle classes.

The advantages of this procedure are shown in Fig. 7. The concept of the rear axle setting station developed in this project, which may be used in a largely universal manner, is shown by examples with two possible scaling levels. On the first scaling level, both the feeding of the axle into the machine and the actual setting process may be carried out manually by a machine operator. With increasing production capacity, an automated setting unit may be added to the assembly system and integrated on a line with minimal effort. The achieved reconfigurability of the innovative axle setting station increases...
the adaptability to unpredictable demands, such as volume fluctuations or new axle models.

1.1 Product analysis

1.2 Process analysis

2 Identification of the reconfiguration demands
- Adaptability to new axle variants
- Reuse of machine technology
- Adaptation to volume fluctuations

3 Functional organisation of the machine structure
- Assembly task
- Responsibilities
- Techn. characteristics

4 Design of the modules and the entire machine
- Modularity
- Scalability
- Standardisation

Figure 6. Design method for reconfigurable assembly systems

Scaling level 1

Scaling level 2

Images: Dürr

Adaptability to axle variants

Scalable volume

Figure 7. Scaling levels of a reconfigurable rear axle setting station

3. Summary

Through the integrated analysis of the entire assembly process chain, from axle pre-assembly to the end of the line during vehicle final assembly, the obtained results show the optimisation possibilities during the rear axle alignment. On the one hand, the issue addresses how to find the optimum way of wheel geometry alignment of the rear axle within the entire assembly process chain. On the other hand, a method has been presented which allowed the development of a reconfigurable rear axle setting station. The innovative concept of a rear axle setting station provides the possibility of being able to react to volume fluctuations and new axle variants with minimal efforts.

Acknowledgement

The research project “Reconfigurable Rear Axle Setting Station” is being carried out in collaboration between the Chair of Assembly Systems at RWTH Aachen University and the company Dürr Assembly Products GmbH. The project is funded by the Federal Ministry of Economics and Technology (BMWi) on the basis of a resolution passed by the German Federal Parliament within the framework of the central innovation programme for medium-sized enterprises (ZIM).

References


Configuration model for evolvable assembly systems

P. Ferreira, N. Lohse
The University of Nottingham; Manufacturing Division; Nottingham; UK

Abstract: The assembly systems domain in recent years has been pressed to provide highly adaptable and quickly deployable solutions and deal with unpredictable changes according to market trends. Furthermore, the current decreasing product lifecycles provided the need for more reusable assembly systems to distribute the system costs across different products. In this context a new concept of Evolvable Assembly Systems has been introduced in this research domain to address these issues. The introduction of this approach presents new challenges but also opportunities for current modular assembly system configuration methods. While the focus on standardised process models makes more automatic configuration approaches more feasible, challenges of how to effectively manage such standards arise at the same time. Furthermore, higher and lower levels of granularity in assembly system modularisation require hierarchical decomposition and consequently synthesis which are notoriously challenging topics for automatic configuration methods. This paper reports on a new configuration method which use a process focused semantic model, as basis for assembly system configuration. The paper identifies all the necessary concepts that enable evolvable assembly systems, namely the assembly system requirements, equipment definition, capability models and overall evolvable assembly system operation process. The potential of this approach is shown to significantly reduce the system configuration and control deployment effort for agent-based modular assembly systems.

Keywords: Adaptive control, Agent, Assembly, Automation, Conceptual design, Emergent synthesis, Flexibility, Intelligent, Knowledge based system, Lifecycle, Mechatronic, Model, Modelling, Modular design, Module, Open architecture, Process, Reconfiguration, System architecture.

1. Introduction

An analysis of the current context of the assembly system domain and its trends has identified serious challenges [1]. Systems require an increasingly high level of responsiveness due to the market demand for increasing product diversity, shorter product lifecycles and shorter times to market while maintaining the cost at a minimum and quality at a maximum [2]. Nowadays, markets are truly global and are characterized by an intensive global competition which is conditioned by socio-economic aspects that influence the assembly systems. In addition to this, market have become increasingly dynamic and unpredictable, requiring product changes and adjustments which emphasises the need more flexible assembly systems.

These challenges have led to several technological advances targeting more flexible, agile and adaptable systems. Assembly shop floors with higher degrees of automation must therefore become as adaptable as their competitors who rely on the techno-economic flexibility of human labour. The concept of flexibility needs to be extended to enable the structure and layout of complex assembly systems to be rapidly built, adapted and enhanced on demand, thus drastically cutting time-to-market and new product introduction times. Shop floors, besides flexibility in terms of product variance and capacity, need to be easily reconfigured, changed and adapted to dynamic conditions with a high rate of variability. Consequently there is a need for a new generation of innovative, rapid deployable, self-configuring plug-and-produce assembly systems, which can be configured with minimal effort to meet a wide range of individual assembly requirements and are enhance-able without disrupting production.

Evolvable Assembly Systems (EAS) provide a new paradigm that takes advantage of the recent advances in modular assembly systems, enhancing them with the concept of evolvability [3]. The vision behind these system it to take advantage of the “Plug & Produce” concept, building highly adaptable mechanical and control solutions that target reusability and interoperability of modules, both logical and physical, enabling the extension of the modules lifespan while providing shorter deployment times. This paradigm proposes the use of agent technology as the basis for controlling the assembly systems where each agent implements one or more assembly processes as so called skills. Skills represent the control equivalent to hardware modules and follow the object oriented approach of encapsulating reusable functionality. This combined with the agent technology provides a natural way for the devices to define clear sets of interchangeable and easy to configure control capabilities and allow their quick deployment into working systems.

This paper provides the definition of a skill concept, which allows for the description of the capabilities of different equipment and how these can be triggered and operated. Furthermore, the concept of skill recipes is introduced to provide the possibility to have deterministic systems using evolvable assembly systems. The paper provides an overall model that defines links to all the necessary concepts for the successful configuration of EAS.

2. Literature Review

An analysis of the current context of the assembly systems domain, quickly leads us to the growing need for higher levels of flexibility. The issue of flexibility in assembly systems is not new and has been one of the main research topics in the field of assembly, namely Flexible Assembly Systems (FAS), which provided the first concepts to introduce bigger flexibility through mainly the increase of the systems capabilities “just in case” and adding cost to the system[4, 5]. However, in realistic terms, these systems are still not common within industry. The fundamental reason behind this is, although technologically
interesting, these systems fail to provide cost-effective solutions for industry as flexibility itself requires added capabilities in the system. This means that the system has components which are not always in use, which makes the system have high degree of redundancy [6]. The concept does provide extra flexibility, but it is restricted to what can be predicted to be needed in the future. This raised other research questions on how to have a more flexible system that is able to deal with the market needs, without adding extensive amounts of redundant equipment that might never be used. It is in this context that notion of agility was introduced.

Traditionally, agility has been comprehended as the capability of an enterprise to operate in a “competitive environment of continually, and unpredictably, changing customer opportunities” [7]. It is important to note that being agile is different from being flexible. Agility defines change as a normal process, which is quite complex to predict and plan for. Thus, in order to have truly agile systems one needs to incorporate the ability to adapt to change. This concept led to the emergence of several production paradigms: Bionic Manufacturing Systems (BMS) [8], Holonic Manufacturing Systems (HMS) [9][13], Reconfigurable Manufacturing Systems (RMS) [10], Reconfigurable Assembly System (RAS) [11], Evolvable Assembly Systems (EAS) and Evolvable Production Systems (EPS) [3, 12, 13]. These paradigms denote the common concept of encapsulation of functionality in self-contained modules. These modules are then used as building blocks of the production system, thus taking advantage of the concept of “Plug & Produce” [14]. This modularisation of systems, provides significant advantages, namely adaptability for product changes, scalability for capacity changes, simplicity due to decoupled tasks, lead-time reduction, maintenance, repair and disposal, among others [15-17].

The modularisation of a system involves the analysis of the similarities among system components to establish modules, which should be kept as independent as possible from each other [11]. Once modules are defined under the context of a modular architecture, a finite set of modules can potentially deal with an almost infinite set of changes [18]. In the assembly domain there are two main types of modules, equipment modules and software modules. By definition modules are interchangeable and are connected by the flow of materials and information [19, 20]. Moreover, module combination and interaction provides the means for the emergence of new capabilities that result of module aggregations [6].

“Plug & Produce” targets the elimination of integration time. This significantly reduces the barrier for new system introductions and also together with interchangeability driven modularisation makes frequent system reconfiguration feasible. Hence the use of task-specific, process-oriented elements (modules) allows the continuous evolution of the assembly system. EAS paradigm is based on these concepts, defining process oriented modules as its basis for system adaptation. This paper reports on the configuration model for EAS paradigm developed under the research IDEAS research project [21].

3. Evolvable Assembly Systems Configuration Overview

The concept of evolvable assembly systems focuses on system adaptability. This adaptability will require an overall approach for the system lifecycle definition, which identifies fundamental concepts, their interrelationships and core characteristics. Very simplistically, the lifecycle of any assembly system can be divided into its design, its build and ramp-up, its operation, and its decommissioning phase. This basic lifecycle model is often extended by enabling a system to be reconfigured once its operational requirements change substantially enough to justify the required effort. EAS defines modularity as both physical and logical, thus establishing assembly processes as a basic logical block, the “skill”, which represents a capability.

The first step in the assembly process model definition is to understand the role and purpose of the model within the context of Mechatronic systems proposed by EAS. The Mechatronic agent concept proposed by the EAS paradigm,

Figure 1. Overview of Assembly Process centred Configuration Method
targets the reduction of the initial build and subsequent reconfiguration effort through the use of modular equipment with standardised interfaces and build-in control capabilities which allow modules (Mechatronic Agents) to be rapidly connected together and dynamically configured to achieve a wide range of assembly processes [13]. The Mechatronic Agent concept hence goes beyond the mere plug-ability of hardware building blocks, such as one would find in a LEGO system. The idea is to not only have the physical equipment modularity, but also create modules of the functional capabilities needed to execute an assembly process. These functional capabilities need to be directly related to the physical building blocks of the system. Hence when a Mechatronic Agent is plugged into the system is comes with its own process capabilities. These capabilities are the so called Skills.

Figure 1 shows an overview of the EAS methodology and how it impacts on the different lifecycle stages of an assembly system, identifying where and how the process model plays are role in this process. In the design stage the system requirements are defined and formalised. These are then input into the configuration stage along with the existing capabilities which are provided by the mechatronic agents. Once this process is completed the system is built using the blueprint that results from the configuration process. Then, the system is in operation with the agents (equipment units) collaborating in order to fulfill the requirements, under the constraints defined by the configuration process. In order to execute complex assembly processes it is expected that agents will create collaborations clusters. These collaborations will require a coalition leader that orchestrates the execution of the complex assembly process. The role of the assembly process (Skill) model is to formalise the concept of a Skill with all its characteristics and define how Skills are used to configure the process logic of Mechatronic assembly systems (MAS). In this role the model needs to support the planning/configuration of new or altered system as well as the execution of the assembly process within the agent control environment. Furthermore, the skill model should be defined such that it allows the capabilities of different agents to be defined by different equipment providers and yet still allow full interoperability.

4. Configuration Model for Evolvable Assembly Systems

The Skill concept is central to the Assembly System Process centred Configuration Method presented in this paper. Skills define the process capabilities offered by the Agents (equipment units) to complete the required assembly process steps. Process skills take a similar role as methods in programming or services in SOA systems. Those Skill capabilities will be used to select and configure a new or re-configured assembly system. From a configuration and design point of view this implies that the available skill capabilities will be compared to a set of process/skill requirements. A similar match process would have to take place if a control system wants to support real time resource (skill) allocation. This makes it evident that a process model will need to include a Skill and Skill Requirement concepts.

Furthermore, it is expected that Skills will have some parameters which can be set either fixed for the operation of an assembly system or dynamically based on the outputs from other skills. Hence it will be necessary to define which parameter settings a Skill should be executed with to achieve the desired result either in advance or during run-time with. These settings can be defined in the form of a Skill Recipe concept which should prescribe how a Skill Requirement can be achieved by a Skill.

Finally, one of the key advantages of the Mechatronic Agent concept is that modules can be developed in parallel by different module providers. Consequently it will be necessary to control the definition of Skills and Skill Requirements to ensure interoperability. One mechanism to achieve the consistency of the definitions is the use of predefined templates. If the same set of templates is used to define both the Skills and the Skill Requirements, they can be directly compared and matched. Those templates will need to be linked to predefined types of skills and parameters.

4.1. Evolvable Assembly System Skill Concept

The skill concept describes and represents assembly processes in a clear object oriented control structure, which can easily be interpreted within the context of the proposed agent environment. Moreover, the skill concept will take advantage of existing concepts within the assembly domain which provide a hierarchical structure of assembly processes. The concept will provide the means to structure and define higher and lower level processes with the ability to describe their composition, inter-dependencies and parametric constraints.

The skill definition will be required to follow the function block concept (approach) allowing for its use within the Mechatronic Agent concept, enhancing the concept beyond the mere plug-ability of hardware building blocks. The key innovation is the composite definition which incorporates both the functional description and the actual execution definition of the assembly process. The agents will possess the definition of the skills, which enables it to provide information on its capabilities has well as having the recipe for their execution. Figure 2 provides a schematic overview of this concept using IEC 61499 notations.

The main requirements for the definition of a skill can be summarized into four main characteristics, namely the assembly process type, the level of granularity (which establishes if a skill is composite or atomic), the skill control ports and the skill parameter ports. The Control Ports represent the control elements, both inputs and outputs, for operating the assembly process (namely: Start, Interrupt, Finished, etc.). These are the basis for the definition of the process sequence since they establish the means to connect between different assembly processes (Skills). The role of the Parameter Ports is to provide the ability to specify the exchange of process data between different assembly processes (skills). These parameters are not mandatory for all skills, but some assembly sequences might require an information flow between different assembly processes. A typical use of these is a force feedback loop, where the value of the force would be passed on to other processes via a parameter port. In other words, this provides the means for information flow of a given assembly process configuration.

The assembly process type intends to provide the information of the capability in terms of its classification within the assembly domain. It is important to note that the naming of capabilities does not necessarily need to be standardised; however without a common agreed terminology between all the
different actors across the whole configuration process, it would not be possible to use computer-based support mechanisms to ensure process consistency and allow automatic matching either during configuration or during dynamic skill allocation at runtime.

The first step in the configuration process is the specification of the assembly process requirements which will define the product work flow constraints (1). The product work flow needs to specify the required assembly process steps, their precedence constraints, and their required parameters. Once this is done, one can proceed to step two, which is the assignment of existing skills to a given requirements. Here there are two options, either these are left blank for a runtime assignment or one can assign one or more recipes for their execution (2). Therefore skill requirements model provides optional definition of recipes, which means that when these are not present the assignment of skills happens in runtime, otherwise the recipes are executed as defined. It is also important to note that one can have more than one recipe for each requirement, since it is easy to understand that depending on the available skills in a system a required skill could be executed in several ways while having the same output result.

5. Illustrative Example

This illustrative example is based on a real demonstrator form the IDEAS project. The demonstrator consists of a medical sector workstation that dispenses a liquid into a recipient. The process requirements are quite straightforward, the component for filling is retrieved, the liquid is handled into the component and the component is then stored. This provides an overview of the requirements; however one of these requirements is a composite skill, since it contains a set of requirements of its own. These consist on sucking the liquid, dispensing the liquid and testing to verify the properties of the liquid. Once these requirements are established and the physical system is built, a collation agent is created for managing the composite skill
execution. This agent will advertise the requirements and establish a cluster for their execution. Figure 4 depicts this and shows the links between the mechatronic agents and the assembly system requirements for this example.

Figure 4. Assembly Process Focused Overview of Configuration Process

The design and implementation of the above process logic and corresponding agent system has been tested with very stable results. The configuration effort of the system could be reduced significantly.

6. Conclusion

This paper presented a new process-based configuration model for rapidly adaptable assembly systems using the EAS concept. An outline of the model has been given which identifies the representational requirements for such a configuration model. The developed model provides a clear and formal basis for development and rapid deployment of modular assembly systems. The process centred approach ensures that the functional integrity of the system can be achieved during system build and executed during runtime.

Further work will focus on more automatic means to support both the design process of EAS and their runtime behaviour. Semantic technology combined with configuration and synthesis methods provides a very promising direction to further explore this work.

Acknowledgement

The reported work is partially funded by the European Commission as part of the CP-FP 246083-2 IDEAS project. The support is gratefully acknowledged.

References

Evolvable Assembly Systems: entering the second generation

M. Onori\textsuperscript{a}, J. Barata\textsuperscript{b}, F. Durand\textsuperscript{c}, J. Hoos\textsuperscript{d}
\textsuperscript{a}KTH Production Engineering, Stockholm, Sweden, \textsuperscript{b}UNINOVA, Caparica, Portugal
\textsuperscript{c}ELREST GmbH, Kirchheim, Germany, \textsuperscript{d}FESTO AG, Esslingen, Germany

Abstract: Current major roadmapping efforts have all clearly underlined that true industrial sustainability will require far higher levels of systems’ autonomy and adaptability. In accordance with these recommendations, the Evolvable Production Systems (EPS) has aimed at developing such technological solutions and support mechanisms. Since its inception in 2002 as a next generation of production systems, the concept is being further developed and tested to emerge as a production system paradigm. The essence of evolvability resides not only in the ability of system components to adapt to the changing conditions of operation, but also to assist in the evolution of these components in time. Characteristically, Evolvable systems have distributed control, and are composed of intelligent modules with embedded control. To assist the development and life cycle, a methodological framework is being developed. The evolution of these components in time. Characteristically, Evolvable systems have distributed control, and are composed of intelligent modules with embedded control. To assist the development and life cycle, a methodological framework is being developed.

Keywords: Evolvable Production Systems, Modularity, Distributed Control

1. Introduction

According to the results attained by many roadmaps \cite{1}, \cite{2}, \cite{3} one of the most important objectives to be met by European industry is sustainability, which is multi-faceted: including economical, social and ecological aspects. The obvious conclusion to this holistic problem is that future manufacturing solutions will have to deal with very complex scenarios. The truly interesting characteristic of this conclusion resides in the word “complex”. Although often used, the basic essence of complexity is that it may not be fully understood and determined in all its ruling parameters; however, we seem to continue to build production systems based on known functionalities and predicted operational scenarios. This, to the authors, remains a rather disturbing factor.

Albeit the enormous efforts made in the 1990’s by Flexible Assembly and Manufacturing Systems, followed by Holonic\textsuperscript{[4]} in the late 1990’s-early 2000s, Reconfigurable Systems\textsuperscript{[5]} and other approaches, the dream of cost-effective, high-variant assembly remains elusive. One of the reasons may lie in the fact that one cannot solve unpredictable scenarios with a focus on predictable functionalities. Nature does not work with predictability. Nature does not propose an evolutionary change based on a single factor, nor does it do so by selecting a single motivating factor. Living organisms evolve by proposing a variety of solutions, but this is done in ways that are not yet fully understood. Yet the adaptation is guaranteed.

Based on this pre-conception that it may be more realistic to assume that a production environment is not fully predictable, and that we should not focus entirely on the required functionalities alone, Evolvable Assembly Systems was proposed in 2002 and has, since then, been developed into a first-generation set of assembly systems. These were tested, enabling EPS to emerge as a production system paradigm (see EUPASS, A3 projects\textsuperscript{[6]}, as given by \cite{7} and the results exhibited at international fairs, Hannover 2008, fig.1). NOTE: the systems built for the Hannover Fair 2008 were followed up by 2 assembly systems at Windisch, Switzerland.

In January 2011, the EPS approach was tested in a second-generation assembly system at the FESTO premises in Germany. This 2\textsuperscript{nd} generation assembly system was re-configurable and exhibited self-organisation. Developed in the IDEAS FP7 project, the details will be given herewith.

2. Background

First of all, the solutions proposed by EAS are not intended to be understood as a general panacea for all assembly scenarios. At present they are a potentially cost-effective approach to large variant flora production and/or short product lifecycles. Once the methodology is completed, and the technology matures, EAS could become more generally viable and cost-effective.

EAS may be viewed as a development of reconfigurability and holonic manufacturing principles. It was initially developed in 2002 from the results of a European roadmapping effort (Assembly Net), and was subsequently further developed in a series of European projects (EUPASS, A3, IDEAS). Its objectives have all been drawn from roadmapping conclusions and are well elaborated in earlier publications\textsuperscript{[8], [9]}

As defined in \cite{10}) RMS incorporates principles of modularity, integrability, flexibility, scalability, convertibility, and diagnosability. These principles impose strong requirements to the control solution. In particular, centralized approaches...
become completely unsuitable due to their intrinsic rigidity. Decentralised solutions must be considered that take into account the fundamental requirements of plugability of components, which includes the aspects related to dynamic addition/removal of components, as well as adaptation in the sense that the system does not need to be reprogrammed whenever a new module is added/removed. This is a fundamental aspect behind any control solution approach to solve the defined requirements [11]. Therefore, the major challenge in the control solution is how to guarantee proper coordination and execution in a system in which both its components and working conditions can be dynamically changed. This is a challenge that needs a completely new approach and this is why in the context of EPS a solution based on concepts inspired from the Complexity Theory and Artificial Life is being developed. The next section covers what concepts from non-traditional manufacturing research domains are being used to create truly dynamic control solutions.

Hence, the control approach to be developed in the context of EPS wants to go back to the basics, that is to say relying stringy on the original idea of considering each component as a distributed intelligent unit that may aggregate in order to create a complex system. In this context, concepts such as emergence and self-organisation become more and more important to be applied to new generation control solutions. However, implementations of these new concepts within shop floor are still very few, as most applications have focussed on planning & scheduling [12].

Considering what was stated above, one may view Evolvable Production Systems (EPS) as a development of the Holonic Manufacturing Systems (HMS) approach; however, a closer looks reveals that, although there are similarities in the exploitation and implementation phases, the paradigms differ quite substantially in their perspective (or trigger issue), and that only EPS achieves fine granularity. By granularity it is considered the level of complexity of the component that compose a manufacturing system. For instance, when a line is composed of several cells and these cells are modules that can be plugged in and out, this is coarse granularity. If, on the other hand, the components that can be plugged in or out are grippers, sensors, or pneumatic cylinders, this is fine granularity. This issue is in fact a very important one in terms of distinguishing the paradigms. The target for EAS is the shop-floor control, which normally demands programming, re-programming and vast integration work.

This is where EAS plays a decisive role. The two fundamental aspects are:

- A methodology that allows the user to define modules at fine granularity level, from a control-point-of-view.
- The development of control boards capable of running the agent software and, simultaneously, be small enough to be embedded in the smaller modules.

The IDEAS project proved this to be viable as a shop-floor solution.

IDEAS-the basics

IDEAS stands for Instantly Deployable Evolvable Assembly Systems. This is an FP7 project that started in 2010 and will end in 2013. It is to develop EAS systems for two industrial customers, IVECO and ELECTROLUX.

The project took advantage of several developments that were done during the EUPASS (FP6...) project, such as:

- ontological descriptions of the assembly processes [9],
- equipment modules prepared for embedded control [13],
- data exchange protocols verified, [14],[15],
- basic methodological principles set [16],

IDEAS had as a main objective to implement the agent technology on commercially available control boards. This would enable distributed control at shop-floor level. What is being considered here is not the planning or logistics level but the actual operational level of the assembly system.

To this effect the ELREST company and FESTO research division set out to specify the exact requirements, based on the needs detailed by the industrial customers Electrolux and Centro Ricerche FIAT. MASMEC, Karlsruhe Institute of Technology and FESTO supported the effort by developing system modules, TEKS provided the simulation software, and UNINOVA and KTH developed the agent technology. Finally, the methodological framework upon which the whole project would base its work, was developed by University of Nottingham.

The project’s first objective was to prove the validity of the approach by running a medical assembly system at the FESTO facilities (see diagram below).

The system shown above ran the following processes:

- Glueing unit: Dispensing glue for assembly of small components
- Pick & Place unit Pick and place handling system
- Electrical testing unit Testing unit for quality/functional product test
- Stacker unit Pneumatic/Servopneumatic handling system

This assembly system, called the MiniProd, was finally demonstrated in January 2011. It ran with a multi-agent control setup, could be re-configured on-the-fly, and the modules self-configured. This was achieved thanks to the fact that the agent software could be run on commercial control boards (Combo, ELREST), which are shown in figure 3.

As this could probably be viewed as the first time an assembly system actually operated with a totally distributed control system, and self-configured, it was shown again for the European Commission in November 2011. The system performed flawlessly, confirming that multi-agent control can be used for truly reconfigurable assembly.
3. The IDEAS Drivers

In order to attain this success, IDEAS has relied on many years research (including the work done in RMS, etc.) and the following own developments:

- A simple and effective mechatronic architecture
- Control boards developed for multi-agent applications
- Am elaborate and well-structured methodology
- Industrial commitment

The mechatronic architecture is, first of all, an architecture that considers the control demands from an embedded-system point of view. That is, each assembly system module is an entity with its own control, hence the “mechatronic”. The difficulty was in creating an architecture out of which an effective control structure could be instantiated for any assembly system layout. As the demands on assembly are extremely diversified (see conveyor system in MiniProd-free-moving pallets!), this posed challenges. The final Mechatronic Architecture is based on four basic agents:

- Machine Resource Agent
- Coalition Leader Agent
- Transportation System Agent
- Human Machine Interface Agent

In order to implement this, the project developed several tools. The actual agent development environment, called IADE (IDEAS agent devt.env.) is based on an elaboration of JADE. The Java Agent DEvelopment framework is FIPA compliant and also provides basic development tools. The IDEAS project further developed these tools and included others to support the simulation of the agent control prior to its being downloaded into the modules. Experiments made at the simulation level and real module also indicated that the simulated module and real unit actually run the exact same code, rendering the simulation extremely accurate (1:1 relation).

The second main development has been the development of commercial control boards capable of running the multi-agent setup. The ELREST company provided the project with several alternatives, out of which the Combo211 was selected for use. This required quite some developments, amongst which:

- Combo200 series runs on WinCe6
- Implemented CrEme™, a Java Virtual Machine (NSI.com)
- Fits to the needs of the Agents and supports JADE
- Implementation of 24V I/Os, Ethernet, CAN and RS232/R5485 connections

The control boards function very well and have also been thoroughly tested at the other partners labs. The project currently intends to develop three variants of these control boards, depending on the required granularity and number of agents/module (from very small, cheap, to mid-size capable of running more than one agent).

Thirdly, the project would have never succeeded if the tools that are required to engineer such solutions were not specifically designed and integrated within the IDEAS methodology. This work, led by University of Nottingham, has brought together many partners (KTH, MASMEC, KIT, TEKS, ELREST, FESTO): the synchronisation and integration are sensitive aspects. The objectives included:

- Develop Semantic Representations for Devices and Skills
- Create Requirements and Target Specification Language
- Semantic Rules for Integration & Validation of Skills
- Develop a rapid System Configuration Environment
- Develop Visualisation and Transparency Tools

Note that this includes skill definition support, Workflow definition support, simulation tools and more. A very simplified view of how an IDEAS system takes shape is given below.

One of the most interesting outcomes of the work has been the link between simulated system and real system. Using commercial software (Visual Components) coupled to the multi-agent programs made it possible to run the exact run-time code prior to download. That means that the simulations represent exactly what will occur in reality (at control level).

All the developments, from EUPASS to IDEAS and beyond, would be quite superfluous if industry had not provided the critical mass and know-how to achieve such results. Industrial aspects are the key ingredient as the certification procedures, variation of hardware constraints, specific customer needs, market demands, etc., all play a decisive role in the effective deployment of a technology. IDEAS took this a step further as it set as an objective that one of the “missing links” had to be corrected: develop a control board for such applications. This was made possible by the industrial commitment, both at control development and requirements specification.

4. Future Steps

The project is now consolidating these results and developing them further. The next step will be to build two industrial systems, in order to verify the full-scale utilisation at customer-level. The two systems will be built at KTH (Stockholm) and MASMEC (Bari). The products to be assembled are an ECU (electronic control unit) from a commercial vehicle, and some specific washing-machine components. The figures below illustrate the schematic layouts.

Both solutions will be thoroughly validate and Life-Cycle Analyses performed. Finally, a new Business Model has just
begun to be developed in support of the more strategic decisions that will be encountered.

5. Conclusions

The article describes the first realistic developments for multi-agent control for assembly applications. The work is extremely valuable but there remains a fair amount of research and development work to be done.

First of all the human role in such automated systems needs to be studied such that people may become an integrated part of the work to be done. In addition, the need for a solid and robust development methodology (guidebook and set of tools) can be generated. This is a highly multi-disciplinary requirement as computer specialists will have to collaborate with production and system engineers at detailed level.

Industrially, these solutions seem to generate sufficient interest, especially as these first tests have clearly shown the viability. The show-stopper is, therefore, not particularly at industrial level but, rather, at academic: consensus as to which “paradigm” is chosen as the most promising is not being based on true industrial development results but on theoretical details. This attitude needs to change and closer, more practical collaboration is required in order to truly support industry. As Thomas Kuhn would possibly put it, we must abandon normal science and search for a true industrial breakthrough; the common comment is “what is new”? IDEAS developed commercial, industrially-certified controllers for multi-agent applications (not PLCs or gumstix). These have not been available to date. IDEAS has also shown self-configuring systems at industrial shop-floor level. Do contact FESTO for visits or videos (Dr.C.Hanisch, dha@de.festo.com). Whether these novelties are substantial or not is up to the individual. At least multi-agent applications have a commercial vehicle now.

Acknowledgements

The authors wish to thank the European Commision for the funding and, in particular, their Project Officer Jan Ramboer.

The IDEAS partners have shown incredible commitment and worked well beyond the budget limits and normal working hours, therefore our deepest thanks go to KTH, FESTO, Electrolux, UNINOVA, MASMEC, Univ.of Nottingham, ELREST, TEKS, Karlsruhe Inst.of Technology, and Centro Ricerche FIAT.

One of the final demonstrators, to be built at KTH, is also developed with the collaboration of XPREs (eXcellence in Production RESearch), a Swedish national R&D initiative.

References

[15]. L. Ribeiro, J. Barata, J. Ferreira; “An Agent-Based interaction-oriented shop floor to support emergent diagnosis”, Proceedings of 8th International Conference on Industrial Informatics (INDIN); Osaka,Japan; July 2010.
Operational characterization of evolvable production systems

H. Akillioglu, A. Maffei, P. Neves, J. Ferreira
Department of Production Engineering, The Royal Institute of Technology, Stockholm, Sweden

Abstract: On the way to achieve mass customization production systems have to obtain the capability of rapid reconfiguration of not only physical components but also from control point of view. Evolvable Production System targets highly adaptable mechanical and control solutions that can enhance reusability and interoperability of modules, enabling lifetime extension of the modules. The focus of EPS paradigm is to achieve overall system adaptability by autonomous modules which are dedicated to specific processes with the capability of short deployment time at shop floor without reprogramming effort. From the operational point of view EPS brings significant enhancements considering shop floor dynamics and performances therefore positioning of EPS principles and approaches in production system typology from different perspectives is essential. This has been done by two means which are process flow structure and customer order decoupling point location.

Keywords: Evolvable production system, invest to order, modularity, distributed control.

1. Introduction

Business environment has been developing and altering by increasing the amenability of the companies in order to satisfy every single customer. It became obligatory for the companies to produce in an efficient way due to increasing customer consciousness and powerful rivals. As far as the relation between customers and companies is concerned, there is a considerable change on the effect of customers on the companies unlike in the past when companies have produced and supplied goods to market as they designed and produced in big volumes. However in today’s environment the companies are targeting to obtain the capability of supplying specific products to every single customer in accordance to their wishes. This is not achievable unless the production system allows reconfiguration in a proficient way in response to changing product characteristics. Besides, production system should permit production volume to be changed without losing capability of rapid reconfiguration between products whose process requirements are diverse.

At the same time, optimality for different indicators is sought at systems in order to maximize economical benefits. Production systems that are designed to function close to optimum for a certain time interval and conditions (maximum efficiency for definite criteria) for a specific production scenario need to be reconfigured and re-planned for the next in order to remain optimal. This configuration and the following ramp-up will unquestionably occupy a certain amount of time, during which the system cannot produce and costly experts are involved to tune the system to the new requirements. Although optimality might be possible to achieve, for most products it is rarely advantageous to delay the production too long for the sake of running a system in optimum, especially in the light of shorter time-to-market and frequent changes of customer requirements. For this reason; there is discrepancy between system responsiveness and optimality of shop floor performance. Therefore, the balance between these two factors plays an essential role for companies to stay competitive. System responsiveness gained importance more than optimality in this balance with shortening product life cycles and demanding process requirements. On the other hand, the desire for optimality and high resource utilisation has resulted in systems that are composed of product specialised machines and equipments. The current trend towards adaptive and responsive systems, driven by ever more dynamical market conditions, is incompatible with the specialised machines and equipments for a product range. Traditional systems are lacking to achieve responsiveness with their current control and shop floor solutions. Dedicated equipment and centralized control approaches are in themselves restricting adaption to new circumstances. To overcome these problems there is a need for new rapidly deployable and affordable (economically sustainable) assembly systems based on reconfigurable and process oriented modular concepts that would allow continuous system evolution and seamless reconfiguration.

Evolvable Production System (EPS) approach was first introduced in 2002 [1]. The aim has not been to develop limited flexibility or barely physical reconfigurability with separate control mechanisms. Instead, it has been to achieve overall system adaptability by self-configuring, process oriented modules with short deployment time at the shop floor/component level based on agent oriented architecture. To be able to achieve system responsiveness (true agility/flexibility), the lowest building blocks of a system are required to exhibit highest rate of adaptability [2]. In the EPS approach, process oriented modules are the basic building blocks that enclose computational power, thereby being autonomous with the capability of interconnecting with other modules to create coalitions for the purpose of accomplishing specific system objectives.

From the operational point of view EPS brings significant enhancements considering shop floor dynamics and performances on certain domains therefore positioning of EPS principles and approaches in production system typology from different perspectives is essential. The well known classification in the literature is handled by 2 means, from the shop floor perspective with process flow structure and from the order management perspective with customer order decoupling point (CODP) location. This paper is targeting to enlighten the position of EPS and in relation to this, it illustrates how EPS relaxes the constraints on capacity planning which are imposed by the limitations of production system. Before all, EPS architecture and principles will be introduced.
2. Evolvable Production System

The focus of EPS paradigm is to achieve overall system adaptability by modules which are dedicated to specific processes with the capability of short deployment time at shop floor without reprogramming effort. Modules are the basic building blocks with the characteristics of being process oriented and embracing embedded computational power in order to enable autonomous behaviour based on multi agent system architecture. In EPS approach intelligent modules form a dynamic network for the purpose of dealing with production system requirements by means of collaborative behaviour. Main characteristics which are envisaged by Evolvable Production System are clarified below.

2.1. Process oriented modularity

A module can be described as the smallest building block of a system. Possibility of assigning varying functionalities to modules, their aggregation capabilities and contribution to control system grounds modular structure to have a wide application area in industry. Issues arising with modular systems include: accomplishing granularity at lower levels (fine granularity) where modular characteristics are shown; developing interface of the modules in order to communicate both to internal and external environment; defining skills and capabilities assigned to modules; and advancing abilities of cooperation and creation of coalitions in order to achieve a common target and their methodologies. EPS modules are designed according to the process needs rather than a specific product family. Figure 1 represents a traditional way of building system where the product requirements are imposed to production system by leaving freedom to product designers and assigning complexity to production systems. Furthermore, most of the product specific equipments become obsolete after the product life comes to end. In the case of new product requirements are introduced to system, the available equipments are missing to be reused due to being product specific.

![Figure 1. Traditional system design](image1)

However, the processes which are actually benefited can be repeated with new system design as in Figure 2.

![Figure 2. Process oriented system design](image2)

Process oriented modularity is a fundamental cornerstone of EPS. Process-oriented modules feature skills that can be mapped into the processes and therefore in the product requirements. Each task in a process can be exploded as a series of operations that involve a variable number of basic activities. This means that a process needs a rather big set of skills to be performed. The equipment owns these basic skills and when different modules with the right skills are put together EPS is formed [3]. Skills are in 2 types in EPS such as atomic skills and composite skills [4]. The atomic skill is the skill for what the module has been designed. They are directly associated with the manufacturing components. "Move" is the atomic skill of a robot. The composite skills are on the other hand the skills that emerge when more manufacturing modules are put together and cooperate [3]. To exemplify, the complex skill of pick and place is composed of 2 atomic skills, namely the skill of grasp which is settled into a gripper module and the skill of move which is settled into a robot arm.

2.2. Multiagent based distributed control

According to Barata et.al 2001, the multi-agent systems represent a suitable technology to support the distributed manufacturing environment [5] since the manufacturing applications present characteristics like being modular, decentralised, changeable, ill-structured and complex, for what the agents are best suited. Agent is defined as a computational system that is suited in a dynamic environment and is capable of exhibiting autonomous behavior [6]. Multiagent system is defined as an environment where the community of agents are interacting [6].

Analyzing the benefits of multi-agent technology it is possible to conclude that it fulfils some of main requirements of the actual distributed manufacturing systems: autonomy (an agent can operate without the direct intervention of external entities, and has some kind of control over their behavior), cooperation (agents interact with other agents in order to achieve a common goal), reactivity and pro-activity (agents perceive their environment and respond adaptively to changes that occur on it). Last, agents can be organized in a decentralized structure, and easily reorganized into different organizational structures [7].

At the shop floor of EPS, the flow of product through the system is following the process sequence located in workflow file which is specific to different product types. Workflow basically includes the list of necessary processes and skills to achieve the end product. It comprises also the dependencies between the required processes in order to give an option in the case of the resource which supplies the process in the queue is not idle. In this case, with the help of dependency information between the processes the product agent can ask for the other process from the available resources in the workflow as long as the busy process does not precede the next one. The main advantage comes with the plug and play modularity where the requirement of reprogramming every time the system configuration changes is avoided. That is to say the time to set up the system in response to produce a new set of products is a matter of seconds as long as the required modules are available in the module repository. Details and architectural characteristics of EPS are available in the literature [8], [9], [10], [11], [12], [13], [14].
3. Production System Typology and EPS Position

The typology of production system can be handled in different ways however two well known ways will be covered which are process flow structure and customer order decoupling point. It is possible to have different formations simultaneously in line with the changing characteristics of the product.

3.1. Process flow structure

In the following figure, different classes of systems are combined in comparison to each other according to various criteria.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Shop</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Batch Process</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Assembly Line</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Continuous Flow</td>
<td>Continuously</td>
<td>Continuously</td>
<td>Continuously</td>
<td>Continuously</td>
<td>Continuously</td>
<td>Continuously</td>
</tr>
</tbody>
</table>

Figure 3. Process Flow Structures & EPS position

Project works are appropriate in order to produce unique products at a fixed location through moving resources as needed and coordinating them using project management techniques. From the product point of view, the variety of products that can be produced is very high whereas the product volumes are very limited due to standardization of the processes are not possible with changing product requirements for each single product. Human contribution in a project work is extremely high while the investment for capital is low.

In a job shop environment general purpose machinery and equipment are utilized and there are a flow of products unlike a project work. Advantage is the capability of producing a high variety of products however efficiency is quite low. The equipments benefited in this kind of systems are general purpose and reliant very much on the knowledge of the workers. Batch process structure has higher standardization in processes in accordance with the product types in comparison to job shop process flow structure, at the same time, decreased capability of producing high variety of products. The essential prerequisite to set up an assembly line is necessity to produce sufficiently large volume, high quality requirements and at the same time low variety of products. Finally, continuous flow has a fixed sequence and fixed pace where the flow of products is continuous.

The characteristics and architecture of EPS does not fit any of the available class properties whereas it can be mapped in fact as a combination of different features of available classes. Figure 3 represents EPS position in comparison to other process structures in light of varying criteria. Red markers on the figure represent where EPS is located for those specified criteria. Each of them is pointing a direction which corresponds to enhancement direction of the system in response to evolution of it which brings not only structural benefits but also operational continuous enhancements.

Bearing in mind, at the shop floor multi agent architecture is benefited where each product agent has its own workflow and follows it according to the availability of the resources after several message exchanges, the volume and flow of the product passing through the system per unit time cannot be as high as an assembly system which is dedicated to a limited number of product varieties. As far as the product variety is concerned, it is one of the most important factors bringing competitive advantage to evolvable systems. Since system reconfiguration can be conducted according to the product needs through adding or removing modules in a very short time without the necessity of reprogramming effort, the number of product variety which can be handled with evolvable system can be so high as long as they are in the same domain. That is to say, you cannot change the system from assembling cell phones to watches however it might be possible to change from iphone to ipad if the proper modules are integrated to the system. That is why EPS might not be economically advantageous like a project work shop floor if you want every single product to be unique however it can produce as good results as a job shop process structure in view of the range of products. Although there is no cost analysis conducted, from the capital investment point of view, intelligent modules with embedded processor will bring extra cost. However it will provide superior return on investment figures as an outcome of the reusability of the modules in the long term since the modules are designed according to process characteristics which can be benefited for varying product requirements. Resource utilization on the other hand might not be as high as an assembly system considering the shop floor working and scheduling principles of multi agent system of EPS modules. However utilization figures can be as good as a assembly system depending on the product mix and also rapid integration of bottleneck process enables it. Variable cost in the figure, stand for unit cost of the product. The cost analysis is conducted on 2 parameters called fixed and variable costs. Fixed costs are not related to product sales volume whereas variable cost is dependent directly on the number of products sold. At figure 3, fixed cost has not been used as a criterion to avoid misleading comparison, since it encapsulate cost items which are changing to a large extent from industry to industry such as rental costs. As a replacement capital investment, one of the items of fixed cost, is used and it reflects the equipment cost. It is complicated to have a rough numeral for variable unit cost of EPS in comparison to other structures. Two main constituents of variable cost are; the cost of raw material and cost of labor, both are increasing with increasing rate of production in conventional systems. However, EPS, having autonomous shop floor control through intelligent modules, does not require increasing human workforce with increasing production. Growing demand is satisfied through increased investment to modules which is classified as fixed cost. Hence, variable cost becomes dependent mainly on raw material cost for EPS which can be considered similar as highly automated assembly systems. Human contribution should be handled at different levels of the whole system. For the figure 3, it corresponds to human effort on production at shop floor level. In the current architecture of EPS, human effort is benefited at the reconfiguration of the system in order to set the system.
and this is a typical make-to-stock production where demand is Afterwards big batches are stored at the end product warehouse order to reduce total setup time and system down time. This forces the company to produce in large batches in converting its production from one variant to another in a short variability. Main components and sub components are manufactured and stored and assembly operation starts with the introduction of the customer demand. This method is applicable if the company does not want to keep end product inventory for each kind of product variants and if the manufacturing of them are long and complicated. And make to stock systems have the risk of costly results coming from forecast errors, early production, inventory costs and such. All of these classifications are product specific and can change according to the forecasts. Decoupling point for Assemble-to-Order (ATO) systems are located between manufacturing operations and assembly. This means, products have required materials and components to be assembled in the case of a customer order arrives. ATO is prevalent for products having a large number of variants and long times for preparation of components. Comparing to a MTS system, this brings advantage on avoiding costly finished goods inventories. Producing and assembling in response to customer orders as in Manufacture-to-Order (MTO) systems denotes the decoupling point is located before manufacturing. MTO systems are feasible if the product is costly to keep in inventory and has a high number of variants, small quantity of demands.

CODP cannot be located in view of the characteristics of production system only. Market characteristics, specifications of the product, transportation issues, competition level etc. are all directly affecting factors to CODP decision. However the flexibility of choosing different places to locate CODP is directly related to production system capabilities since most of the constraints to force CODP to be located closer to the customer are imposed by production system incapability. That is to say, if the company is running in an environment where the product variety is quite high but the production system is not capable of converting its production from one variant to another in a short time, this forces the company to produce in large batches in order to reduce total setup time and system down time. Afterwards big batches are stored at end product warehouse and this is a typical make to stock production where demand is met by the products from the end product storage. It embraces both the risk of over production which can turn to be a waste if no demand appears and an opportunity cost rising from early start of the production.

A fully reconfigurable system can relax a considerable portion of the constraints imposed from manufacturing system on moving CODP away from the end product storage towards the procurement of raw material which enables to start value adding activities later in response the demand. Bearing this in mind, EPS brings radical enhancements on the CODP location flexibility. Evolvable system capabilities cannot exactly be benefited in one of MTS, ATO or MTO systems. To explain, make to order systems starts the manufacturing and assembly when the customer order arrives and the underlying logic is to customize the product according to the customer wishes. But the product is fixed, the options to choose by the customer are mostly predetermined and the production system is set according to the main product components. The order, in this case includes specific parameter settings which are chosen by the customer and it is used as an input to start the production and to specialize the product. Assemble to order systems on the other hand are suitable for companies dealing with high product variability. Main components and sub components are manufactured and stored and assembly operation starts with the introduction of the customer demand. This method is applicable if the company does not want to keep end product inventory for each kind of product variants and if the manufacturing of them are long and complicated. And make to stock systems have the risk of costly results coming from forecast errors, early production, inventory costs and such. All of these classifications are product specific and can change according to the shifting conditions of the production system, product or demand. They all can be used at an EPS environment and get superior results however the real benefit and breakthrough of evolvable system come into view with the fully reconfigurability which enables Invest to Order systems.

In Figure 3, there are 2 areas called forecast driven area and customer driven area which differs by the source of the input that trigger the flow of activities. The purpose is to enlarge the customer driven area to avoid the inefficiencies caused by uncertainties. Invest to order system goes one step further comparing to make to order by enabling the customer order to feed the new module and equipment investment for the production system to be reconfigured. This is possible only if you know what modules and processes to invest in order to meet the customer expectation. A sound process planning method is essential at this point and the approach to achieve this obstacle is available EPS methodology [2].

In industry today future market and targeted product scope are determined and investment to the production system is conducted accordingly in order to be prepared responding customer needs rapidly. However this is valid as long as the reality follows the plans. Long setup times and less reconfiguration capabilities of production system prevent to react to new and unexpected requirements coming from the market. Although the current production system might continue being profitable, the lack of rapid response to new requirements will end up high opportunity costs and loss of competitive advantage. Therefore the main challenge for an Invest to Order system is to obtain a production system which allows to be reconfigured not only physically but also from the control and
programming point of view. Evolvable Production System characteristics fulfill the fundamental requirements for an Invest to Order strategy. At the same time, to be able to conduct an invest to order system, a comprehensive planning reference architecture is compulsory which come into line with the working principles of not only EPS but also Invest to Order strategy.

In actual fact Invest to Order makes the capacity planning activity to be less dependent on long term forecasted data by enabling shorter capacity decisions to be taken. To give a better picture on capacity planning in the conventional systems figure 5-a elucidates two kinds of capacity decisions. Conventional system in this case represents the systems where the production system is product dedicated and does not have the competence of reconfiguring itself according to changing product needs in a short time. For a conventional system, the demand anticipation for the product life should be conducted at the beginning in order to determine production system capacity. If the inventory holding is not possible or the cost is very high in comparison to keeping excess capacity the production system capacity has to aim to highest demand of the product life (Figure 4(a), capacity level 1). The trade-off between excess capacity cost and inventory holding cost is the determinant factor on setting the intersection point of capacity and product life (Figure 4(b), capacity level 2).

Figure 5. Capacity decision of conventional systems & EPS

The common difficulty for capacity planning of conventional systems is requirement to anticipate demand before the investment to capacity. Evolvable production system, tackles this issue by providing fully reconfigurability through process oriented modular structure based on multi agent based control. This offers the possibility of delaying capacity investment (invest to order) since the system setup time is minimized or even eliminated. The obligation of forecasting product demand pattern for a long term can be relaxed and the risks involved in forecast errors can be avoided through EPS.

Figure 5-b shows how EPS can be utilized through incremental capacity planning. Instead of investing on capacity relying on long product life forecasts, the time to be forecasted can be minimized and the capacity can be extended in response to increasing demand incrementally. In the figure the parameters used are c, capacity increment; t, targeted time period; α, slope (product pattern parameter).

Capacity increment in this case is a multiplication of targeted time period and the product pattern slope. \( C = t^*\alpha \)

Capacity decisions for EPS needs to be supported by analysis to reach a optimal result on the trade-off between inventory holding cost, excess capacity, customer backlog cost etc. So that optimal intersection point of capacity and product life can also be found to balance the parameters for highest economical benefit.

4. Conclusion and future research

Operational characterization of EPS as presented in this paper enables to position EPS in comparison to other production system formations from 2 different perspectives. Architectural characteristics of EPS bring key advantages considering the shop floor performances where removal/minimization of setup time plays essential role in this which is enabled through multiagent based distributed control structure.

This study forms the base of the future research which targets to benefit the advantages brought by EPS by analysing the effects on planning system structures. That is to say, to generate actual benefit from evolvable production system, planning architecture and production system has to be compatible and consistent with each other and this will be the direction of future studies. The reported work is partially funded by the European Commission as part of the CP-FP 246083-2 IDEAS project. The support is gratefully acknowledged.

References

IADE – IDEAS agent development environment: lessons learned and research directions

L. Ribeiro, R. Rosa, A. Cavalcante, J. Barata


Abstract: The IDEAS Agent Development Environment (IADE) is a library of JADE-based Agents specifically designed towards rapid design and deployment of mechatronic systems. The environment covers fundamental shop floor activities such as hardware and process descriptions, transport and deployment aspects. The library follows the architectural principles of the Evolvable Assembly System paradigm whereby the system’s functional building blocks denote the capacity to self-organize in order to meet distinct production scenarios and disturbances. It does so by relying on a set of generic agents that abstract entities in a production context namely: products, resources, processes and transport equipment; and other agents that support deployment and yellow paging service.

This paper details the technical architecture of IADE. In particular, the core data model is presented, discussed and the main design decisions are explained.

Keywords: Agent, Manufacturing system, Automation, Mechatronic

1. Introduction

There is a growing interest from industry in the application of distributed computation concepts and paradigms at shop floor level. Despite a growing number of system architectures that aim to introduce plug and produce, graceful degradation, fault tolerance and resilience to disturbances, most of these proposals have been produced and tested inside the academia with little industrial impact.

Industrial efforts on their side have focused on the development of distributed information technologies (IT) platforms without any specific mechatronic model.

The main issues concerning academic prototypes are so that often powerful concepts such as self-organization and emergence, drawn from complexity sciences, are applied in an highly theoretical framework with a limited link to the mechatronic context.

The issue with focusing in IT is that the solution is frequently too generic requiring an architectural perspective that often does not fit a given technology.

The FP7 IDEAS project was created as a joint effort between industry and academia with the objective of developing an agent-based mechatronics framework that can stand as a proof concept of the applicability of distributed IT specifically harmonized with complexity science’s concepts (self-* properties) in order to render mechatronic system seamlessly reconfigurable and robust.

The IDEAS Agent Development Environment (IADE) is a library of JADE-based Agents specifically designed towards rapid design and deployment of Mechatronic Systems.

The IADE agents have been modelled to promote self-organization through careful interaction design and activation. The dynamic resource finding capabilities resulting from a full parametric model, whereby agents seeking resources do so using a best matching approach, which considers the signature of the required functionality, are additionally presented.

The current results build upon the earlier demonstrator [1] of the IDEAS project that highlighted the limits and opportunities of agent based production paradigms and motivated further developments and refinements of the IADE.

2. IADE Functional Architecture

As previously detailed the IADE architecture is agent-based. As a proof of concept, the architecture is developed upon the JAVA Agent DEvelopment Framework (JADE) sharing with it some basic functionalities such as: Foundation for Intelligent Physical Agents (FIPA) compliant communication support, basic agent architecture and description, IT management aspects and service publishing and subscribing functionalities.

The core of IADE, which shall be detailed in this paper, extends these basic functionalities and implements six mechatronic agents that support distinct aspects of modelling, design, deployment and execution of an IDEAS system. A fundamental aspect of IADE agents is that they enable a one to one relation between physical entities and processes in the production floor and its virtual counterparts.

The Resource Agent (RA) is the most basic entity in the architecture (Fig. 1) and it is responsible for the control of modular equipment that can be plugged and unplugged from the system. It directly interfaces with the equipment’s hardware and therefore is a true Mechatronic Agent (MA) in the sense that it harmonizes an equipment, its controller and the agent interface. The functionalities of a RA are indivisible from a system point of view and are directly linked with a specific hardware. Resource agents can be integrated in higher order agents to enable more complex functionalities and take part in system-wide reconfiguration processes.

The Coalition Leader Agent (CLA) is the entity responsible for the aggregation and orchestration of the agents(functionalities hosted by other agents, in particular RA’s and other CLA’s. A CLA supports the execution logic of processes which are designed by the user based on the available functionalities in the system. The CLA is able to react to changes that affect the composed functionality. In this context, any removal, or fault, in the modules used in a coalition forces the
CLA to negotiate a valid replacement to maintain that functionality.

The **Product Agent (PA)** is a special instance of a CLA. Indeed the PA is a CLA that does not respond to execution requests. The PA is the virtualization of a product in the system and therefore is responsible for managing the production details of each specific item.

The **Transportation System Agent (TSA)** abstracts components of the transportation system. It provides localization, transport and positioning functionalities. Each TSA keeps track of its own position which is typically associated with the position of RA’s, CLA’s or PA’s. The TSA defines an interface, that other agents can use, which hides the implementation details of different transport solutions. In this context the TSA does not enforce any particular kind of transportation mechanism. Using the interface approach is trade-off between increasing the complexity of the agent architecture (by incorporating a considerable number of transport specific agents and all the envisioned interactions between them) and pushing the complexity towards implementation specific details that necessarily need to be interfaced with the remaining agents and may change across distinct systems.

Agents publish their functionalities in the **Yellow Pages Agent (YPA)** so that other agents can consume them as well as subscribe the notification service that informs on the status of specific agents.

During execution the notification of a failure is directly received from the CLA or RA rendering the YPA an essential agent in any self-organizing process in the system. The YPA supports the discovery of functionalities based on parameterized queries that can return all the functionalities of a given agent or the agents that implement specific functionalities. These queries can be specified with the desired degree of detail including searching for a fully specified functionality or, otherwise, omitting irrelevant constraints enlarging the scope of response of the search.

Skills are the main executing construct in IADE and the concept that described the functionality of an IADE agent. Although at interface level (execution, querying and discovery) their description is harmonized, internally skills can be atomic or composite. **Atomic Skills (ASK)** are the RA implementation. Each atomic skill details the technical interaction and integration with the concrete equipment being controlled. In particular it encapsulates an instance of the low level library (I/O control level) and its mapping to higher order functions used at agent level. **Composite Skills (CSK)** support the design of processes in a workflow-like fashion. In this context it is possible to design parallel, sequential and conditional execution flows.

The **Deployment Agent (DA)** is an auxiliary agent that consumes a serialized description of a CLA or a RA, rebuilds the agent, and deploys it on a specific platform. The DA is the main agent to be bootstrapped on the controllers where the IADE platform is meant to be run.

### 3. IADE data model description

This section details IADE generic data model (Fig. 2).

From an implementation point of view the Mechatronic Agent Class is the main construct of IADE. It is an abstract class that encloses most of the behavioural logic that is common to the remaining IDEAS Agents. Although not all the fields in the data model have been depicted in Fig.2 the most relevant from a framework point of view include:

- **myType**: stores the mechatronic agent subtype (RA, CLA, TSA or PA). It is an implementation specific variable used in the agents serialization process during the system deployment phase.
- **OMACState**: state variable that keeps track of the machine status of a controller associated with that agent (if applicable or used). It is a conventional automation concept.
- **mySkills**: it is a list of skills regenerated during the bootstrap phase of the agent and contains their full definition for local execution purposes.
- **negotiator**: the engine responsible for negotiation with other agents to execute a skill (namely between PA’s and CLA’s, CLA’s and RA’s and CLA’s and CLA’s). This field is abstract and each mechatronic agent sub-class must provide its own implementation;
- **executor**: the engine responsible for skill execution. This field is abstract and each mechatronic agent subclass must provide its own engine;
- **yellowPages**: a library that enables the mechatronic agents access the Yellow Pages Agent services. The IADE YPA facilitates parametric querying for skills and agents in the environment.
• negotiatorBehaviour: is the server side of the negotiator. Enables all the IADE agents to react to negotiation requests. It is supported by the FIPA Contract Net protocol [2].

• executorBehaviour: is the server side of the executor. Enables all the IADE agents to react to execution requests. It supports the FIPA Request protocol [3].

• myPendingSkillMapping: the queue of skills to be locally executed. IADE agents execute their skills concurrently. In this context all the execution requests are listed here and executed as soon as possible.

• pendingBehavioursScheduler: the behaviour that controls the local skill execution order contained in myPendingSkillMapping queue.

• tempName: it is a user friendly name of the agent that facilitates its identification during the design and deployment phases.

• area: defines the area associated with this mechatronic agent. An area defines a physical delimitation of the system it is used to constrain the system’s self-organizing response ensuring that incompatible agents are never put together under a coalition.

As previously detailed agents expose their functionality as skills which work as a sort of lingua franca that enables the different agents to negotiate, execute, search and organize according. These classes take the following fields:

• name: is a domain specific name that may be used during the search procedure when CLA’s or PA’s are attempting to allocate resources. A skill can be overloaded inside each agent (i.e. an agent can have two skill with the same name provided that the parameters list is different).

• type: identifies the kind of skill: atomic, composite or decision. Type identification is fundamental so that the agents know how to use a specific skill (hardware access at RA level and process coordination at PA, CLA level).

• owner: defines a list mechatronic agents that may execute the skill. If the list is empty the system will negotiate the best agent to execute it.

• processingTime: keeps track of the average execution time of the skill.

• instantiated, executed, started and failed: these fields define the life-cycle of a skill and are used by the IADE scheduler to control the execution list.

• parameters: a skill can take different parameters which behave as I/O lines. Each parameter is defined by its name, native type and type (in, out or variable). Other auxiliary variables include the definition of upper and lower acceptable values (when applicable and currently restricted to numeric types) and a field reserved for enumerated values.

• mappingVars: table that does the mapping between local variables (input, output or input/output parameters) and global variables in a CSk inside a coalition leader agent.

• mappingValues: table that contains the global variables, that is, a pair name/value of that variable.
  
  An ASk is a type of a Skill that abstracts a low-level library method that does an action in the system:

• nativeClassName: the name of the Java class that will be instantiated and contains the method abstracted by this skill;

• nativeMethodName: the name inside that class that will be called for the purpose of executing the atomic skill.

• libClassInstance: stores a reference of the class instance to avoid re-instantiation of the class unless required due to a change in the library.

A CSk aggregates sub-skills of any type, allowing the execution of composed functionalities. CSk’s are the base element of coalition formation:

• executeAs: defines whether the sub-skills are executed as a sequence or concurrently.

• subSkills: a workflow of skills that define the CSk.

The execution of a composite skill is not trivial. In fact it starts by the analysis of the full execution tree until its leaves are identified. Once the analysis is finished a skill execution initiator behaviour is started (this behaviour is the client side of the executor behaviour described before) for each lower level skill.

There are no limits to the depth of the workflow description.

A Decision Skill (DSk) is a special type of skill that performs the conditional execution of sub-skills inside a composite skill. A DSk has an expression that can be evaluated in runtime and whose result branches the flow of execution:

• expression: stores the expression to be evaluated in runtime and which interpretation returns a boolean result.

• node0: stores the skill that is executed when the expression evaluates to true.

• node1: stores the skill that is executed when the expression evaluates to false.

DSk’s are only considered inside CSk’s and while there is an identity between a CSk and CLA and a ASk and an RA, DSk’s are not exposed by any IADE agent.

RA’s expose atomic skills and ensure that the interaction with a specific hardware is operational. RA’s have the ability of dynamically loading JAVA classes that implement the hardware libraries by using the JAVA Reflection API to guarantee the dynamic instantiation of both Java Classes and their methods associated with specific skills. For performance, the JAVA class is instantiated during the agent bootstrap, yet it can be re-loaded as required. All the parameters in the call are pre-screened to ensure correctness in both type and number.

From an implementation point of view the CLA supports the execution of composite skills. To improve the performance the CLA always maps the description of a CSk to a set of native JADE behaviours that can be directly executed by JADE’s scheduler. There is always a trade-off between speed of execution and system agility when developing a CLA since the higher the degree of composed functionalities in the system the higher the communication requirements is respect to message exchange. All the communications in IADE follow FIPA protocols.

TSA agents store information about positioning, neighbourhood and metrics of the transport system as previously detailed. From a system design point of view there is a one-to-one relationship between each element of the transport system and an agent. Three types of low level transport agents are considered

• Automated Guided Vehicle (AGV): any autonomous transport entity with the ability to perform free movement within some safety and space constraints.
• Conveyor: an element in a network of conveyors that conditions the possible path that parts can take in a production environment.
• Handover Device: a device that is able to transfer a part between distinct sections of a transport system.

The TSA provides other agents the ability to advertise the position where their skills are executed and to compute the cost of travelling from any other position of the system to its current position. This measurement can be used as a metric so that PA's can decide in runtime the best location to have a given skill executed.

The YPA controls the persistence of Yellow Page’s data which is currently assured by a relational database engine implemented in SQLITE [4]. The YPA also manages the IADE platform so that all the agents joining and leaving the platform are tracked and all the other agents consuming their skills are notified of that fact.

Yellow paging service is always a bottleneck in agent-based platforms. In fact, part of the rationale behind the development of the YPA is that the native service transfers part of the semantic processing load to the client side. This implies that RA’s and CLA’s would consume relevant processing time in analysing raw data or building search queries instead of handling production specific tasks. The YPA optimizes and speeds up this process by implementing IADE specific search and query templates.

4. Lessons Learned, Challenges and Concluding Remarks

Although the present paper details the functional and technical implementation of a agent-based framework for mechatronic systems it is inspired by the Evolvable Assembly Systems EAS paradigm [5, 6] which sets the main theoretical guidelines that should drive an evolvable, adaptable and self-organizing mechatronic system. The EAS paradigm itself was not an epiphany of its initiators, instead is descends from a line of emerging production paradigms namely: Bionic Manufacturing Systems [7, 8], Holonic Manufacturing Systems [9], Reconfigurable Manufacturing Systems [10] and the new interpretation [11, 12] of the relatively unsuccessful Flexible Manufacturing Systems. There is a common line along these paradigms that envisions the use of distributed, autonomous and self-contained shop-floor components that can be brought together in distinct ways to generate, in runtime, distinct functionalities and overcome emerging production disturbances.

The application of these production paradigms is still elusive in industry. Although a very significant part of the problem is the lack of supporting tools. Indeed these new approaches require a different set of tools. One of the main points of using IADE like systems is to eliminate the (re)programming effort during the life-cycle of a system. The main concept is that module providers may sell/rent mechatronic agents and, in particular, resources and this is the background for the IDEAS project. This shift from reprogramming to reconfiguration carries implications in respect to performance. The advantage of a modular system is plugability and its ability to react to disturbances at the cost of increased communication requirements with an associated time cost. This also implies some degree of redundancy so that the system is able to explore configuration alternatives. This is a fundamentally new way to build mechatronic systems and implies, to a considerable extent, a mindset shift whereby it is accepted that the system takes autonomous decisions, designed-constrained, rendering the overall system much more dynamic.

IDEAS like systems are indeed lively predictable entities but not manageable within a conventional framework. These systems mainly handle processes in the form of skills and, as a set of interacting building blocks, quite complex interaction patterns may form. The number of core entities in IADE has been purposely reduced to a minimum (four agents) to ensure that the collective response of the system is always consistent.

This is also part of a mindset shift that system builders must take: to model a system out of general purpose building blocks.

As a proof of concept IADE will not, under the scope of the IDEAS project, perform in real time industrial conditions. Yet in these early development stages it has allowed pointing out some the challenges that may lie ahead.

Although the preliminary version of IADE has already been tested in two industrial demonstrators as part of internal IDEAS activities there are still open technical challenges. As an extension of the JADE framework the mechanisms for handling agent failures from a software point of view do not meet industrial standards. In fact a new platform would have to be developed from scratch to accommodate, within the IADE principles of functioning, these requirements.

There are however other challenges within the scope of IDEAS and currently being pursued and integrate in IADE namely: the development of metrics that improve the self-organizing response of the system (in particular in material handling); the definition of learning mechanisms that improve the response of the system and the definition of well defined interfaces to a wide range of complementary tools to support users in the processes of creating and managing systems.

References
Distributed bayesian diagnosis for modular assembly systems – a case study

M. S. Sayed, N. Lohse
The University of Nottingham, Manufacturing Division, Nottingham, NG7 2RD, UK

Abstract: The growing interest in modular and distributed approaches for the design and control of intelligent assembly systems gives rise to new challenges. One of the major challenges that have not yet been well addressed is monitoring and diagnosis in distributed assembly systems. In this paper we propose the use of a multi-agent Bayesian framework known as MSBNs as the basis for multi-agent distributed diagnosis in modular assembly systems. We use a close-to-industry case study to demonstrate how MSBNs can be used to build component-based Bayesian sub-models, how to verify the resultant models, how to compile the multi-agent models into runtime structures to allow consistent multi-agent belief update and inference.

Keywords: Assembly, Modular design, Bayesian networks, Error Diagnosis

1. Introduction

New market conditions imply that manufacturing systems today need to be more flexible and responsive to enable quicker response to the changes in the market and the introduction of new products and product designs while being robust in the face of disturbances and unexpected failures [1]. To respond to these new challenges, new paradigms for the design and control of manufacturing systems have been proposed within the last two decades, the main drive being towards decentralized/distributed control combined with tendency to modular component-oriented system design approaches. The most notable of these paradigms being reconfigurable manufacturing systems [2], Holonic manufacturing systems [3] and more recently Evolvable manufacturing systems [4].

The Need for Modular Diagnosis

One of the main features that need to be properly addressed in order to realize the vision for robust modular manufacturing systems is the ability to diagnose faults quickly and hence allowing quicker recovery times and less down time and stoppage costs. The ability to monitor the system during the operation means the ability of the control system to adapt to changes and unexpected behaviour and hence resulting in more reliable systems. It has been reported that during a production system downtime, repairing the cause of the error does take a small fraction of the downtime. It is locating the source of the problem (i.e. diagnosis) that is responsible for the majority of downtime [5]. The difficulty in identifying causes of errors becomes even more relevant in complex production systems.

Another recent drive for diagnosis and prognosis efforts has been the growing interest in building integrated frameworks for condition-based maintenance management in industrial organizations where Diagnosis and prognosis models are used for the optimal planning and scheduling of maintenance activities [18].

Most of the available work on diagnosis is designed for centralized systems. Hence there is a need for new decentralised methods to address monitoring and diagnosis in modular, component-based physical systems. The drive towards component-based systems design, in fact brings new opportunities for more efficient affordable monitoring and diagnosis systems because specific component suppliers should then be able to design the monitoring and diagnosis models for their component as an integral part of the component design effort. Consequently, the diagnostic design effort, which represents the main obstacle in designing large scale diagnostic systems, will be distributed among specialized component suppliers allowing the people who know the most about each component to design its diagnostic sub-system.

However, in order for this vision to be materialized there is a crucial need for an underlying framework or architecture that would allow independently developed diagnostic sub-systems to be integrated into a wider diagnosis and monitoring system once the concerned physical system is built from its modular components. Since the use of multi-agent systems is becoming more prominent in the control of modular manufacturing systems (for example, evolvable production systems and Holonic manufacturing systems), it becomes very attractive option to embody the monitoring and diagnosis functionalities into the paradigm of multi-agent systems for manufacturing control.

Various approaches do exist for centralized fault diagnosis in industrial systems depending on the types of faults to be diagnosed. One prominent approach that has been studied in various domains including fault diagnosis recently is Bayesian Networks with promising results. However, all of the work on industrial diagnosis using Bayesian networks has been applied in a centralized manner using the classical notion of single Bayesian network.

A growing research interest has been directed to develop new approaches that would extend the use of Bayesian networks knowledge representation and inference techniques with their proven merits into multi-agent systems. A few notable works have been reported in literature suggesting new approaches to achieve this goal. It is our belief that there exists a unique opportunity to extend these new research endeavours into the field of monitoring and diagnosis of modular manufacturing systems in order to materialize the vision of component-based monitoring and diagnosis systems for modular manufacturing systems.

Figure 1 illustrates our proposed vision of the component-based Bayesian diagnostic framework.
This paper reports on our recent work in using multi-agent probabilistic reasoning for the problem of multi-station assembly systems diagnosis. The rest of this paper is organized as follows: section 2 gives a background review on the state of the art in manufacturing diagnosis. In section 3 the basic behind Bayesian networks and MSBNs are introduced. Section 4 describes our case study that is used to demonstrate the proposed approach, and finally section 5 finishes with concluding remarks.

2. Diagnosis in Manufacturing

A number of approaches have been reported in literature to address fault diagnosis problems in manufacturing. For example Discrete Event Systems (DES) have been used in [19] and [20]. Despite their theoretical soundness, however like all model-based approaches, DES-based approaches do suffer from major limitations when applied on complex systems due to state explosion and the need for accurate detailed modelling which is not always feasible for manufacturing systems [6].

On the other hand, data driven approaches with little modelling requirements have been investigated. These include artificial neural networks [6], fuzzy logic [7] and Bayesian networks [8] among others. The models here do not necessarily capture precise nominal system behaviour as in model-based approaches, rather a general understanding of how the system works, and the types of expected faults and associated symptoms can be used to obtain a basic diagnostic model.

Bayesian networks in particular have gained considerable interest as a mean for fault diagnosis in various domains with very promising results, for example in telecom networks[9], machining processes [10], manufacturing diagnosis [11] and electric power systems in aerospace applications[12].

3. Bayesian Networks

A Bayesian Network is a directed acyclic graph where random variables are represented by nodes and the arcs among the nodes represent the probabilistic causal dependency relationships between the random variables [13]. The diagnostic process in Bayesian Networks involves inferring the likelihood of the occurrence of an unobservable fault hypothesis based on the measured (observed) evidences (Symptoms).

A very important and powerful characteristic of Bayesian Networks is their ability of Belief Updating via bidirectional propagation of new evidences throughout the network. This allows for the conditional probability of each node to be updated as new evidences or observations become available.

One can distinguish between two components of Bayesian diagnostic framework, the Bayesian model and the inference method. The model represents the knowledge about the causal relationships that relate various errors to symptoms. These relationships can be elicited from human experts and available engineering knowledge about the system. The inference method is the algorithm used to infer error causes from observed symptoms.

A big body of literature exists on building the two elements. However, very few works address Bayesian modelling and inference in multi-agent environments. A prominent approach for multi-agent Bayesian modelling and reasoning is Multiply Sectioned Bayesian Networks (MSBNs) [14] which will be presented in more detail.

Multiply Sectioned Bayesian Networks (MSBNs) is a knowledge representation formalism that extends conventional Bayesian networks to enable cooperative multi-agent probabilistic reasoning. It can be used to model a domain using a set of Bayesian subnets, each subnet representing a subset of the domain.

The framework defines a number of constraints that need to be satisfied in order to ensure coherent distributed inference, which are presented in detail in [14], these constraints are:

i. All subnets must be organized in a hyper tree structure.
A Directed Acyclic Graph (DAG) must be used to represent each agent’s knowledge. This constraint is the same constraint for classical Bayesian networks which can be verified independently for each agent/subnet.

The set of variables shared by pairs of agents should form a d-sepset. This is to ensure that the shared interface between neighbouring agents allows sufficient information to be passed as necessary during belief updating (concise message passing).

**Definition:** Let \( G_i = (V_i, E_i) (i = 0, 1) \) be two DAGs such that \( G = G_0 \cup G_1 \) is a DAG. A node \( x \in I = V_0 \cap V_1 \) with its parents \( \pi(x) \) in \( G \) is a d-sep node between \( G_0 \) and \( G_1 \) if either \( \pi(x) \subseteq V_0 \) or \( \pi(x) \subseteq V_1 \). If every \( x \in I \) is a d-sep node, then \( I \) is a d-sepset.

The global network constructed from the union of all the subnets must be acyclic which extends the structural constraint of classical Bayesian networks. Verification of this constraint requires cooperation of all agents in the graph through message passing over the hyper-tree according to the algorithm detailed in [15].

Given a group of individual subnets that represent a common domain, in order to establish an MSBN that allows coherent distributed inference, a number of steps need to be followed which can be generally described as:

i. Through coordinated message passing among neighbouring agent, checking the graph consistency of neighbouring subnets.

ii. Through coordinated message passing among agents, moralization of the subnets.

iii. Triangulation of the moralized graph through message passing among the agents.

iv. The construction of the linked junction forest from the triangulated graph.

The resultant linked junction forest is the MSBN representation used for belief update and reasoning.

### 4. MSBN-based Monitoring and Diagnosis Case Study

MSBNs appear to be a very promising framework for monitoring and diagnosis in a multi-agent environment such as EAS for example. The monitored system can be modelled using a number of subnets each corresponding to an equipment component. The subnet will represent the local diagnostic knowledge of the autonomous agent monitoring the component. We will demonstrate the concept using a close-to-industry modular assembly system.

#### 4.1 Diagnosing HAS-200

HAS200 is the system under consideration of which one station is shown in Figure 2 is a modular multi-station educational assembly system that constitutes three production stations, quality check stations, in addition to packaging and warehousing stations. It employs a distributed control strategy as every station is controlled through a dedicated local controller. The production stations fill a box with different types of materials to produce a variety of products depending on the weight ratios among material types in the final produced box.

After the three process steps have finished the height of the final produced box is checked at the quality check station, and depending on preset height thresholds, the processed box is either accepted or rejected. When a box is rejected, it would be important to know the most probable cause for that.

To allow such queries to be made, a diagnostic Bayesian network is built for every station describing error-symptom causal relationships along with prior probabilities. The different error symptoms that indicate the presence of an error are related to possible underlying causes using the notions of Bayesian Networks. Symptoms and errors are all modelled using two state random variables that alternate between two states, present and not present. The prior probabilities of error are obtained from expert knowledge mainly based on the reliability information of the different system components. During runtime, observations are collected from across the system and are then used to update the prior probabilities of error in order to obtain the posterior probabilities of errors.

Figure 3 shows the overall Bayesian diagnostic model representing the three stations together. It relates the possible errors, namely, feeding errors, handling errors, measurement errors and material errors (box geometry in this case) to the possible observable error symptoms in the three stations. The three first errors are local to the each individual station and hence they are represented with local variables within the network. On the other hand the box geometry error is considered to be a global error since it affects all of the three stations unlike the rest of the errors which are confined within the station, therefore; it is represented using a shared variable. Other shared variables are the local height measurement variables, which affect the height measurements in the following station, and therefore a height variable is also shared between two consecutive stations.

To demonstrate the concept of modular component-based diagnosis, we will section this global network into three individual subnets, each of which will be assigned to an individual autonomous monitoring agent dedicated for one station only. This means that every agent will be observing a number of variables local to its station and will not be aware of any other variables within the scope of other agents.
The sectioning into separate networks should be in a way that guarantees maximum possible decoupling among individual agents by minimizing the number of shared variables where possible. The shared variables will be the means or the interface through which neighbouring agents communicate with each other, while the local variables will be hidden and only used for internal reasoning within individual agents.

4.2 Model Verification

Once the component-based subnets for individual agents have been constructed, they should then be verified for MSBN-consistency, i.e., checking that MSBN constraints are satisfied to guarantee efficient distributed inference.

The constraints presented in section 3 are firstly verified locally within each agent and then verified for the whole system through message passing among agents. This cooperative verification makes use only of the public variables shared among neighbouring subnets. This is a very important feature, as it allows the global verification to take place without the need for detailed knowledge about the internal structure of individual subnets. This feature makes the verification of individually developed subnets easy to perform and more importantly it accommodates for knowledge protection for component suppliers who might be unwilling to share proprietary knowledge about the internal functionalities of their components.

Certainly this entails the need for special care to be taken during the construction of individual subnets in terms of selecting shared variables that each agent makes accessible to the neighbouring agents. The model verification process here and all the subsequent steps for model compilation and inference are carried out using the software toolkit WEBWEAVR IV [16] which provides tools for multi-agent probabilistic modeling and inference.

4.3 Model Compilation

Once the individual subnets have been built and cooperatively verified for MSBN consistency, they should then be compiled into a runtime secondary representation to allow efficient and coherent unified reasoning and inference. This compilation involves three steps,

First, each subnet is independently moralized by connecting parents of each node together and dropping link directions. Then through message passing, neighbouring moral subnets are checked for consistency on the shared nodes.

Next step is triangulation on the moral subnets, where each subnet is locally converted into a chordal graph so that the set of consistent chordal sub-graphs defines the super-graph of the union of the set of sub-graphs. A chordal graph is defined by having at least one perfect elimination sequence. The set of chordal graphs is then converted into a corresponding set of junction trees. Moreover, a junction tree is established for each d-sepset between two sub-graphs, these junction trees are known as linkage trees. The conversion of sub-graphs into junction trees is to allow effective inference using message passing algorithms [17] and the linkage trees are to ensure concise message passing among agents through beliefs on d-sepsets. The resultant representation for the whole system is called linked junction forest which is then used for online MSBN inference.

The online reasoning is performed on the linked junction forest as each agent uses a junction tree for its internal representation and the associated linkage trees for communicating with the neighbouring agents on the hypertree structure. Figure 4 shows the model compilation steps on the agent representing station1.

To explain the need for multi-agent communication during the belief update process we will consider a case from this illustrative example. We can see that as every agent only has access to the states of the local variables in the station, it cannot always make appropriate belief updates on the local variables without communicating with neighbouring agents. For example in Station 3, assume the local variables are observed to confirm the absence of any errors within the station, i.e., both feeding and handling variables are observed and known to be error-free or in OK state. Assuming that Measurement variable is not directly observable in Station 3, Station 3 agent will not be able to update the state of Measurement variable. To do so it needs to communicate with the previous station in the line to update its belief on the previous height measurement and accordingly update its belief about the state of its measurement process. So assuming the final quality check rejected the box due to height
being unacceptable, then Station 3 agent will update its belief on the state of its measuring device depending on the belief of Station 2 about the height in Station 2. Neighbouring agents propagate their belief on the shared variables along the hypertree by responding to belief update requests. After the communication takes place, all agents’ beliefs will be consistent with all observations accumulated across the system.

5. Conclusions

We explained the need for new modular component-based diagnosis frameworks for modular manufacturing systems along with the associated challenges. To respond to these challenges we suggested the use of multi-agent Bayesian approaches as a distributed diagnostic framework. We demonstrated how MSBNs can be used to model a diagnostic system for an educational modular assembly system. MSBNs allow distributed autonomous and collaborative inference. Moreover, the processes of model verification, model compilation and inference are carried out while protecting agents’ privacy. We have used the Java-based software toolkit WEBWEAVR IV to implement the component-based diagnostic system and to demonstrate the process of distributed modelling, multi-agent model verification, multi-agent model compilation and belief update for consistent system level inference.

We will further investigate systematic approaches for automatic model generation in our future work with specific focus on model modularity. Follow up work will be focused on the performance evaluation of the proposed method in terms of accuracy of detection and computation efficiency to allow more understanding of the advantages and limitations of the method.

Acknowledgement

The reported work is partially funded by the European Commission as part of the CP-FP 246083-2 IDEAS project. The support is gratefully acknowledged.

References

Geometry assurance versus assembly ergonomics - comparative interview studies in five manufacturing companies

M. Rosenqvist, A. Falck, R. Söderberg
Department of Product and Production Development, Chalmers University of Technology, Göteborg, Sweden

Abstract: The objective in this study was to explore how assembly ergonomics and geometry assurance issues were regarded by manufacturing companies. Therefore, 85 persons in total of which 21 was geometry engineers in five manufacturing companies were interviewed. Their answers show good awareness of the implications of poor assembly ergonomics but this does not always reflect the system solutions created. Further ergonomic requirements cannot contend with geometrical requirements, geometry has priority over ergonomics when they are in conflict. The improvement focus should be in the early development phases, mainly through the use of better virtual tools including calculation of quality costs, by training, new cross functional working procedures and through lessons learned.

Keywords: Geometry; assembly; quality

1. Introduction
In today’s manufacturing industry there is a goal to achieve an excellent perceived quality of the products (Maxfield et al. [1], Wickman [2], Forslund [3] and Wickman et al. [4]). To facilitate this, much focus is directed towards geometry assurance (Shah et al., [5]). But how about the assembly of the products? Earlier research shows that bad assembly ergonomics influences the product quality (Axelsson [6], Eklund [7], Lin et al. [8], Yeow et al. [9] and Falck et al. [10]). Is there a conflict between the focus on geometry and assembly ergonomics? As much as 10-40% of a company’s total turnover is cost of poor quality (Harrington [11], Bank [12], and Booker et al. [13]). Both geometry and assembly problems often drive costs of poor quality and poor fitting directly relates to the geometrical quality (Wickman [2]). By making the correct decisions in early product development phases these costs can be reduced. In addition improved assembly ergonomics enables social sustainability in companies with manual assembly. Because of this, research is required in order to find out what the problems and needs are. This study is a continuation of a study performed by Falck [14] with the addition of more interviews and a different focus on geometry assurance.

Geometry engineer: Responsible for developing and verifying the geometrical system solutions (locating schemes, tolerances, fasteners etc).

Geometry system solution: Locating scheme, tolerances, fasteners, assembly method etc for a part.

2. Objectives
This study is done as part two of a large semi structured interview study performed in several companies, including professional roles both in product development and manufacturing engineering (Falck [14]). In addition to the interviews in Falck [14] (study 1) 21 geometry engineers in the two large companies were also interviewed. The main purpose is to find out what type of improvements are needed to achieve a better product design that enables both high geometrical quality and good assembly ergonomics. A further purpose in this study (study 2) is to compare the answers of geometry engineers with the answers of the respondents in study 1.

3. Methods
3.1. Interviews
The interview procedures in study 2 were similar to those in study 1 in which semi-structured interviews of project engineers in manufacturing engineering (ME), product development (PD) was performed. In this study additional interview of dedicated geometry engineers both in ME and PD was performed. Each interview was conducted in the workplace of each respondent. The questionnaire consisted of 37 semi-structured general questions that were asked to all respondents and a number of role specific questions only asked to the geometry engineers. In this study only the general questions will be analyzed. To make a comparison possible, answers from study 1 are added and discussed.

3.2. Participating companies and selection of respondents
The participating companies are the same as in study 1 in which five manufacturing companies were contacted for interview. Two of them are large companies in the automotive industry. Three of them are small and medium sized enterprises (SME) one of which is a supplier to the automotive industry, one produces lighting armatures and one locker modules for delivery vans. The selection of manufacturing engineers in the two large vehicle companies was made so that these represented main product areas such exterior, interior, chassis and electrical parts, components and geometry. In addition 21 geometry engineers in the two large companies were asked the same questions in study 2 to determine if their answers differ from other professional roles in study 1.

To be able to investigate the different circumstances and compare answers and comments the same questionnaire was used in all five companies but not all question were asked to all respondents. The questionnaire consisted of two parts, first part was general and asked to all respondents and second part was specific to geometry roles and only asked to geometry engineers. Except from the two last questions only the questions asked to all respondents are used in this analysis. Some questions not all respondents wanted to answer, therefore these have been omitted from the results.
4. Results

The interviews were conducted in 2010 (study 1) and 2011 (study 2). Each interview took between one hour (study 1) and three hours (study 2) to accomplish. All answers were carefully written down by the interviewer. Afterwards it was checked with the respondents that all answers were properly understood. Besides, the interviews were tape recorded. In those cases where there was questions about the written answers the recordings was used to determine the correct interpretation of answers.

In the two large companies 51 and 20 persons were interviewed and in the three minor companies altogether 14 persons. In total 85 persons participated. See Table 1.

<table>
<thead>
<tr>
<th>Company</th>
<th>Large A</th>
<th>Large B</th>
<th>SME</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD</td>
<td>18</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>ME</td>
<td>16</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Geo</td>
<td>17</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>20</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 1. Respondents.

Interpretation of diagrams:
- Total: All respondents
- Large PD/ME: both PD and ME from study 1
- Large PD: only PD from study 1
- Large ME: only ME from study 1
- SME: Respondents working at the three small companies, from study 1
- Large Geo: Respondents that are geometry engineers, added in this study (study 2)

Because of the extensive response materials only a selection of the questions and answers can be presented. These are:

- **“Do you consider ergonomics when you design your products?”**
  
  Large differences can be seen between the different professional roles within the companies. The group of geometry engineers stand out claiming that 90% does not consider ergonomics whereas in the other roles, both in large and SME almost 80% consider ergonomics. Amongst the geometry engineers a majority meant that “someone else” will consider this. See Figure 1.

- **“Do you think that poor ergonomic conditions could result in quality issues of the product?”**
  
  Regardless of role or company a majority of the respondents including the geometry engineers were convinced that poor ergonomics results in quality issues. In total 96% said yes with small differences between companies and roles.

- **“Do you know that poor assembly ergonomic conditions result in quality losses and high corrective action costs both in assembly and for products ready for delivery to the final customer?”**

  18 of 21 geometry engineers said that they knew this showing good awareness of the implications of poor ergonomics. Amongst other roles (Large and SME) 42 of 59 knew with no major differences between large and SME companies.

- **“If an ergonomics risk is costly to eliminate – what is required to solve it: is it sufficient with explicit requirements from manufacturing engineering/factory for acceptance of a product change?”**

  A majority agrees that it is not sufficient, with no significant differences between roles or size of company. Those who answered no/perhaps mostly stated that cost is the most relevant factor not ergonomics. See Figure 2 (No/P = No/Perhaps).

- **“Are geometric and ergonomic requirements often in conflict?”**

  A majority in total means that they are seldom in conflict but a majority of the PD respondents mean that they are often in conflict. Only 30% of the geometry engineers mean that requirements often are conflicts but this is mainly due to the fact that they don’t consider ergonomics. See Figure 3.
• “Can ergonomic requirements contend with geometrical requirements?”

In this question the results vary both from role and company size, but in total it is clear that ergonomic requirements cannot contend with geometrical requirements, only 26% of all respondents answered yes. See Figure 4.

Figure 4. Can ergonomic requirements contend with geometrical requirements?

• “Are assembly ergonomics considered when choosing geometrical system solution?”

Only 20% of the geometry engineers answered yes in contrast to 60% of the other roles that answered yes, which is a strange result since it is the geometry engineers that are responsible for choosing and developing the geometrical system solution.

• “Are the assembly ergonomics investigated enough when new geometrical system solutions are selected?”

In total a majority claims that assembly ergonomics are not investigated enough. A clear difference can also be observed between PD, which means that they investigate enough, and ME, SME and Geo that means that they do not investigate enough. This difference is partly already discovered in the previous question and confirmed here. See Figure 5.

Figure 5. Assembly ergonomics investigated enough?

• “Are there an interaction between assembly ergonomics and geometry?”

The difference discussed in the two previous questions is again obvious. Amongst geometry engineers only 33% claim that there is an interaction whereas the other roles in general claim that there is. See Figure 6.

Figure 6. Interaction between assembly ergonomics and geometry?

• “Can you define the concept of Robust Design?”

This question was not asked to the Geometry engineers since it is part of their daily work. In total 64% understands the concept of robust design, however there are large differences between different departments. In PD less than half (45%) of the respondents understand versus almost all (85%) in ME. An acceptable answer would include insensitivity to variation. See Figure 7.

Figure 7. Can you define the concept of Robust Design?

• “What support is needed for better decisions in the choice of assembly concepts that concerns ergonomics?”

This question was not asked to the geometry engineers. Several different suggestions were given in different levels of abstraction. But five main groups of proposed improvements can be distinguished based on the answers from the respondents in study 1:

1. Better calculation models and facts for ergonomics: 20 persons
2. Experience from production included in PD “lessons learned”: 19 persons
3. Earlier consideration of assembly ergonomics when selecting solutions: 15 persons
4. Better/more simulation of requirements and solutions: 14 persons
5. Better requirements and communication: 14 persons

5. Discussion
The interviews have been performed in different companies and in different conditions but several interesting common results can be seen. In all companies, and roles, there is a high
ergonomics, however, this does not always reflect the system understanding of the problems created by poor assembly ergonomics. The results imply this is not the case, assembly ergonomics comes in too late in the product development. The gap continues when evaluating the relation between ergonomic and geometrical requirements, geometrical requirements clearly has a higher rank in the companies, especially if it will cost money. The geometry engineers are aware that the assembly ergonomics of their system solutions are not considered and/or investigated enough, but PD thinks that they are OK. ME on the other hand that observes the solutions in practice also understand that they have poor assembly ergonomics. This difference in roles within the companies continue, geometry engineers claim that there is a low interaction between geometry and assembly ergonomics whereas the rest of the company seems to think that there is. The question is how can this be? Some reasons include; lack of training in ergonomics, lack of ergonomic requirements, lack of tools to assess ergonomics, lack of cross-functional communication, lack of knowledge about the “real-world” production etc.

The concept of robust design (Taguchi et al. [15]) is not well communicated within the PD part of the companies, but well known in the ME part, here is a definitive need of training. This can also be the effect of lacking cross-functional communication.

6. Conclusions

The results show that there is big improvement potential in this area. Although most of the answers showed a good understanding of the problems created by poor assembly ergonomics however this does not always reflect the system solutions created. Further can ergonomic requirements not contend with geometrical requirements, geometry has priority over ergonomics when they are in conflict. The improvement focus should be in the early development phases, mainly through the use of better virtual tools including calculation of quality costs, by training, new cross functional working procedures and through lessons learned.

7. References

Compensation of shape deviations for the automated assembly of space frame structures

J. Fleischer and M. Otter
wbk Institute of Production Science, Karlsruhe Institute of Technology (KIT)

Abstract: During the production process of three dimensionally-curved aluminium profiles which are used for the assembly of closed space frame structures, there are always deviations between the intended and the actual contour. These production-related deviations can add up during assembly in a way that an automated “closing” of these structures is only possible to a limited extent. This paper presents an approach to increasing the accuracy of space frame structures during assembly by additionally machining single profile ends. The derivation of the processing parameters and therefore, the optimization of structures, is based on the mathematical modeling of single three-dimensional profiles and all relevant processing steps, which are represented in this paper.

Keywords: Assembly, Flexibility, Quality

1. Introduction
    The use of frame structures, e.g. space frame structures (SFS) made of aluminum extrusion profiles, which are increasingly used in the automotive and aerospace industries, significantly contribute to weight reduction and therefore, to the increase of energy efficiency [1-3]. These SFS are made of single profiles, which are assembled adjacent to close SFS. Due to rising energy prices, lightweight concepts will play a decisive role in the reduction of fuel consumption [4].

    Simultaneously, there is a tendency of product customization which leads to an extended product variety, the distribution of units to different variants and therefore, to a modification of production processes into small-batch series processes [5]. Flexible and automated small-batch production requires a high degree of facility adaptability for the respective product that is to be manufactured [6]. Especially within assembly, which is still one of the most cost effective approaches within high product variety [5], the demands for more flexibility and the adaptation of the assembly process increases steadily [7].

    For economic reasons, design-for-assembly approaches usually prefer a lower number of single components. Concerning SFS, the number of joints should also be minimized as they usually represent the mechanical weak spot [8]. Therefore the geometry of single profiles becomes more and more complex. The parts can be produced using the process “curved profile extrusion”, which allows the flexible and economic production of complex three-dimensionally curved extrusion profiles [9-10].

    In general, product manufacturing processes are subject to variations and can therefore never be perfect [11]. New production processes such as “curved profile extrusion” and the tendency towards small batch production, which usually does not allow the leveling of processes, additional challenges arise concerning accuracy and therefore, there are always deviations between the intended and the actual contour of single profiles.

    In general the small-batch assembly of SFS includes a high degree of manual labor. However, automated assembly is preferred because of quality and economic aspects [12]. The existing deviations of single profiles can add up during assembly in such a way that the “closing” of a frame structure is only possible to a limited extent. As regards the assembly process, two fundamental issues arise. If a flexible, automated assembly without any specific devices is taken into consideration, it becomes evident that assembly will not be possible without any measures to increase the accuracy of SFS. In a conventional assembly process which is nowadays usually performed with specific rigid devices it can also happen that the “forced closing” of a SFS leads to the insertion of residual stress in the assembled frame structure. The result is a significant negative impact on the fatigue strength and thus, a decrease of the duration of life.

    Given that accuracy problems are often not recognized until the components are assembled or until the final product is completed [13], it is evident that advanced measures have to be implemented to ensure that assembly processes can be automated and meet high accuracy requirements while keeping the creation of residual stress to a minimum.

2. General approach and boundary conditions
    The machining of the profile end sections of single profiles is a flexible and economical way to increase the accuracy of the SFS. By shortening the start and end sections, as well as by inserting chamfers, the profile contour can be varied almost arbitrarily in all directions and thus, the accuracy of the SFS can be improved (see Fig. 1). The following procedural method is applied.

    The measurement data of the actual contour are available for all profiles which are assembled into a SFS.

    Based on these data the mathematical modeling is performed. Related to a virtual assembly, the contour deviation of single profiles and therefore of the SFS can be represented. This allows the error that occurs when “closing” the SFS to be determined. This error is minimized by a subsequent optimization with respect to basic conditions, and the processing...
2.1 Basic characteristics of different joints

If a profile is considered idealized and able to be moved in space freely, two translational and three rotational degrees of freedom can be adjusted. However, the conditions that are effective during real assembly joining processes limit the variation of linear and angular parameters. The scope of the research activities presented in this paper includes direct and indirect joining connections. Direct (profile / profile) connections can be realized through the use of different welding techniques, whereas indirect (profile / joint) connections can be realized with hydroforming or electromagnetic compression [1, 14-15] processes. This leads to the definition of the following boundary conditions:

Profile / Profile Connections:

i. No axial offset in any of the three spatial directions.
ii. Geometric coincidence of the surfaces of both joining partners. If angles in the form of chamfers are machined into the part, the change has to be made to both joining partners.

Profile / Joint Connections:

i. Within the joint element, rotation in two directions is impossible.
ii. A compensation of length can only be provided until the profile contour starts to bend.
iii. The desired joint surface which represents the strength of the connection as well as the flush "closing" section within the joint limit the length adjustment in the opposite direction.

The geometric design of the cross section of the profile also imposes restrictions for the possible solution. For square profile representing all non-circular geometries, it is evident that rotation about the profile-axis is limited for both connection types. The above-described boundary conditions have to be adhered to for the mathematical modeling and for the subsequent optimization process.

3. Mathematical Modeling

3.1 Modeling of profiles and joints

Regardless of the test set, there are often measurement data of actual profiles in the form of a scatter-plot in a local contour does not have to be considered. The relevant points of a profile are merely start point $S$ and end point $E$. The two normal vectors $n$ can be used (see Fig. 2) to identify the direction of extrusion of the profile. Consequently, a basic profile-vector can be described for any curved profile.

$$\vec{v} = \vec{E} - \vec{S}$$

Furthermore it must be possible to describe joint elements adequately, given that closed frame structures must be represented. The same description with a basic vector can be used (see Fig. 2). With start point $S_j$ and end point $E_j$, the joint-vector can be described as

$$\vec{v}_j = \vec{E}_j - \vec{S}_j$$
The alignment of the bore holes of the joint are defined with normal vectors \( n_{1,i} \) and \( n_{2,i} \).

### 3.2 Modeling of the machining process

The purely translational shortening of the profile start is shown in Figure 3. By shortening the profile start by the length \( l_s \) in the opposite direction of the surface normal vector, the end point \( E \) is moved into the direction of \( n_s \) by \( l_s \). By shortening the profile end in the same way, the end point \( E \) is moved in the negative direction of \( n_e \) by \( l_e \). Thus, the equation of the profile vector can be added for pure translational machining and is as follows:

\[
\vec{v} = \vec{E} - \vec{S} + l_s \cdot \vec{n}_s - l_e \cdot \vec{n}_e
\]

Figure 3. Modelling of machining process and itemization

In the equation shown above, the possibility of machining chamfers into the profile end and thus, angle variation, are not taken into account. Therefore the equation has to be expanded.

As mentioned above, profiles are usually defined in a local coordinate system \( B \) (see Fig. 2). By defining a global coordinate system \( B_e \), a profile can be defined completely in all three dimensions. By introducing a transformation matrix \( T \), the profile vector can always be described in the global coordinate system as

\[
\vec{v} = s \sum_{i=1}^{n} T_{B_e B_i} \vec{v}_{B_i}
\]

The transformation matrix \( T \) is represented by orthonormal basis vectors [16] in the following form:

\[
T = \begin{bmatrix}
\hat{b}_{1}^{(1)} & \hat{b}_{1}^{(2)} & \hat{b}_{1}^{(3)} \\
\hat{b}_{2}^{(1)} & \hat{b}_{2}^{(2)} & \hat{b}_{2}^{(3)} \\
\hat{b}_{3}^{(1)} & \hat{b}_{3}^{(2)} & \hat{b}_{3}^{(3)}
\end{bmatrix}
\]

The descriptions of the basis vectors of local system \( B \) with respect to global system \( B_e \) are represented within the columns of the transformation matrix. There is an advantage especially in terms of flexibility because special rotation sequences, for example the well-known Euler-Angles [16], can be omitted. With this type of representation, the spatial position of the profile can be varied during the subsequent optimization process. Therefore, it is possible to determine the processing parameters concerning the angles at the profile ends through the use of the entries in the transformation matrix.

### 3.3 Detailing of the machining process

The examination of the mathematical coherences is based on the profile-guideline. Therefore, the profile end is shortened again if the angles are adjusted (see Fig. 3). This reduction in length can be generally represented using geometric coherence:

\[
\Delta l = \frac{d}{2} \tan \nu
\]

Therefore, the profile vector has to be expanded to the following expression:

\[
\vec{v}_{B_e} = s \sum_{i=1}^{n} T_{B_e B_i} \left( \vec{E} - \vec{S} + \left( l_s + \frac{d_s}{2} \tan \nu \right) \cdot \vec{n}_s - \left( l_e + \frac{d_e}{2} \tan \nu \right) \cdot \vec{n}_e \right)
\]

Given that the starting angle is known from the measurement data of the profile, it is guaranteed that any profile shortening resulting from angular adjustments is taken into consideration.

### 3.4 Closed space frame structure

In order to optimize the “closing dimension” for an assembled SFS, the objective function has to be established. The SFS is virtually assembled from single profiles and joint elements, which is mathematically carried out by the vector addition of all elements. Thus, the equation can be presented as:

\[
E_r = \sum_{i=1}^{n} T_{B_e B_i} \vec{v}_{B_i}
\]

Figure 4. Objective function of virtually assembled frame structure

\( E_r \) is the “closing dimension” of the entire SFS and therefore also the objective function. The example shown in Fig. 4 contains the addition of three profile vectors transformed in a global coordinate system with \( E_r \) being the closing gap during assembly.

### 3.5 Constraints of the assembly process

For the assembly of two profiles with a joint, two constraints must be taken into consideration which can be described as
This formula indicates that both profiles have to match the respective bore hole of the joint (see Fig. 2).

If there are any cut-outs on the structure that cannot be varied spatially, further constraints must be taken into account. This for example may be a bore hole for a connection of another mechanical part. By way of example, point \( P_i \) from Fig. 4 is considered. The theoretical position of the point is known. Since the measured data of the third profile is available, all points of the third profile are known respectively to base \( B_3 \). For

\[
\min \left( \| P_3 - \alpha \cdot B_3 \| \right) \leq d_{\text{max}}
\]

it is guaranteed that at least one of the points of the third profile does not exceed a defined distance to point \( P_i \). It becomes evident that the coordinates of all points of the third profile have to be specified with respect to global basis \( B \). This is achieved by using transformation matrices. Because of a possible length reduction of the profile beginning, the constraint achieved by using transformation matrices. Because of a possible length reduction of the profile beginning, the constraint achieved by using transformation matrices. Because of a possible length reduction of the profile beginning, the constraint achieved by using transformation matrices. Because of a possible length reduction of the profile beginning, the constraint achieved by using transformation matrices. Because of a possible length reduction of the profile beginning, the constraint achieved by using transformation matrices. Because of a possible length reduction of the profile beginning, the constraint achieved by using transformation matrices. Because of a possible length reduction of the profile beginning, the constraint achieved by using transformation matrices. Because of a possible length reduction of the profile beginning, the constraint achieved by using transformation matrices. Because of a possible length reduction of the profile beginning, the constraint achieved by using transformation matrices. Because of a possible length reduction of the profile beginning, the constraint achieved by using transformation matrices. Because of a possible length reduction of the profile beginning, the constraint achieved by using transformation matrices.

By using this form of constraints all points of interest that must not be varied spatially, or that alternatively must be located in a definite position, can be taken into consideration.

4. Conclusion

The approach introduced in this paper shows a flexible option for the comprehensive characterization of frame structures and for the characterization of possible machining processes. Based on the results of the optimization – length of single profiles and spatial orientation – the processing parameters for minimizing the closing dimension can be derived. Initial trials with basic structures with only a few boundary conditions have shown that the developed modeling approach for these kinds of structures is suitable for the optimization process.

Future research will focus on more detailed real process modeling. For example, this includes the machining of profile ends within the bent section of a profile. The activities will also cover the constraints that are required to ensure that the optimized structure stays within a spatial tolerance defined by the reference contour. As part of the optimization, algorithms have to be advanced in order to retain a ceiling of flexibility and more important achieve the required optimization accuracy.

Further interesting research may also be how profiles have to be designed to avoid possible deviations directly in the manufacturing process, and therefore minimize refining operations, or how profiles have to be designed for the specific approach shown in this paper. Possible synergetic effects may occur, so that both tasks, minimizing the closing gaps and the difference of the intended contour can be achieved.

Acknowledgement

This paper is based on investigations of the subproject C4 - "Combined application and processing kinematics" - of the Collaborative Research Center/ Transregio10, which is kindly supported by the German Research Foundation (DFG).

References


Abstract: Automotive stampings and sub-assemblies are compliant, and as a result dimensional variation can be absorbed or further compounded by assembly interactions. Furthermore, different assembly approaches, including the order of welding and clamping sequences, can influence dimensional outcomes. While there are examples of experimental or industrial trials that assess variation propagation in a physical process, further studies in this area provide invaluable context to the body of research into compliant assemblies. This paper seeks to demonstrate key process characteristics of a typical single-station production assembly from the automotive industry, specifically highlighting the potential influence of differing welding and clamping sequences on dimensional outcomes.

Keywords: Assembly, Joining, Quality

1. Introduction

Variation is an unavoidable, natural part of manufacturing processes, but it can be controlled to differing levels depending on the particular process and the amount of resources manufacturers are willing to use in this effort. The process of deciding what levels of variation are allowable in a finished product is based upon functionality, desired quality level, manufacturing capability, and cost of manufacture, and is often an iterative process requiring ongoing revisions. In terms of dimensional variation, tolerances are the criteria against which products are assessed; they are the dimensional variation limits stating the maximum allowable variation of a product if it is to meet design specifications.

The study of how variation accumulates through the assembly process is necessary in order to accurately assign tolerances. Tolerance stack-up in rigid component assemblies can be calculated through the additive theorem of variance. This law does not generally apply to flexible component assemblies, such as stamped sheet metal assemblies, due to the ability of components to distort into conforming or non-conforming shapes [1]. Tolerances for sheet metal assemblies are therefore often based on poor estimations that can lead to either unnecessary quality control efforts, or the manufacture of components that are not capable of meeting tolerance limits. More accurate methods are therefore required to predict tolerance stack-up in sheet metal assemblies.

Research in the field of compliant assembly spans a wide range of areas. Several categories of mechanistic process models have been proposed to better capture compliant behaviour in assemblies, including direct finite element modeling (FEM) methods [2], and linearized approaches for determining an assembly's response to part variation [3]. Modeling methods can be combined with a range of optimisation approaches, with examples including determining optimal fixture configurations [4] and adaptive corrections [5] for maximising dimensional quality. Compliant assembly models have also been integrated within quality visualization and assessment tools [6]. Multivariate statistical methods have been used to identify variation patterns and faults in correlated compliant assembly measurement data [7]. Knowledge-based approaches aimed at generalising and simplifying compliant assembly design have also been proposed, including a minimum stress criterion for reduced dimensional variation [8], and a framework for integrating process knowledge and virtual analysis techniques for more efficient tolerance allocation [9].

Work that includes experimental case studies are of particular importance, as they assess the performance of actual processes [10] and allow for comparison with (and validation of) theoretical approaches for the analysis and assessment of compliant assembly. In terms of the investigation of progressive assembly constraints, such as clamping and welding sequences, several experimental studies have been presented. Hu et al [11] developed a compliant assembly simulation approach using commercial FEM software, with experimental results from a dashboard assembly showing strong correspondence with the proposed simulation method. Three different welding sequences were shown to produce differing assembly outcomes. The thermo-mechanical effects of spot welding processes have also been investigated [12]. Here, both an experimental study, and simulation approach that accounted for localised thermal expansion, demonstrated the potential influence on dimensional outcomes of sheet metal assemblies. Liao et al [13] proposed a nonlinear contact simulation approach, where simulations were compared to a simplified experimental assembly for a range of welding sequences. This comparison demonstrated the improved accuracy of contact modeling over previous approaches. Matuszyk [14] explored the influence of clamping sequence on the dimensional outcomes of an experimental two-piece channel-section assembly. It was shown that different sequences could result in considerably different assembly deviations, and that the optimal clamping sequence was largely dependent upon the specific variation modes of the input components.

The studies to be presented in this paper focus solely on the experimental investigation and observation of a production case study assembly: the actual assembly fixture, robots, work-cell, stampings and sub-assemblies were used for assembly fabrication. This presents a unique perspective in relation to previous experimental efforts, whereby simplified experimental assemblies and offline fixtures have been used for investigation
of compliant assembly welding and clamping sequences. A
description of the case study is presented in the next section,
followed by the results and observations of the clamping and
welding sequence studies.

2. Case study production assembly

The case study production assembly of interest is a front
cross member common across sedan and sports utility vehicle
models (see Fig. 1). The assembly locates critical visible features
in the final vehicle, including fenders, hood, and headlamps.
Excessive dimensional variation in this assembly can result in
flushness and fit issues, and impact customer quality
perceptions. For the purpose of this study, only the final sub-
assembly operation was investigated. Here, the front cross top-
plate is welded to the lower sub-assembly (which is comprised
of the remaining components).

Two studies were undertaken on this assembly, the first
observing the influence of assembly fixture clamping sequence,
and the second observing the impact of welding sequence on
dimensional outcomes. These will be presented in the following
sections.

3. Clamp sequence

For the clamping sequence study, the assembly fixture was
taken offline and placed on a coordinate measuring machine
(CMM). The top-plate and lower sub-assembly were then
clampered using a pre-defined sequence, and measured at a set of
critical measurement points (MPs) before the welding stage of
the assembly process. Through using the same top-plate and
lower sub-assembly for the study, part-to-part variation
contributions were eliminated, allowing for isolation of the
effect of clamp sequence on dimensional outcomes. Clamp
locations for the assembly fixture are illustrated in Fig. 2. There
are four clamps labelled A through D. Clamps B and C have two
clamp contact points (actuated by the same hydraulic cylinder),
one on the front flange and the other on the back flange. The
MPs consisted of 6 surface points on the front and rear weld
flanges, and the center-point of 12 locating holes, which
importantly position visible mating assemblies and components.

These observations suggest that clamp sequence should be
considered during the process design or tryout phases due to
the potential for dimensional process shifts. Valuable extensions
to this clamp sequence study would include larger sample sizes
to assess the impact on process variability rather than mean
shifts alone, and the investigation of simultaneous clamping
which was not considered here.

3. Weld sequence

To illustrate the effects of different weld sequences on the
dimensional outcomes of assemblies, two weld sequences were
investigated and compared. There are two weld robots which
perform welds in the last sub-assembly operation. Robot 10
welds the front edge of the assembly, and Robot 9 the back edge of the assembly as indicated in Fig. 4.

The two weld sequences investigated involved running one robot after the other. For the purposes of this paper, the weld sequence that corresponds to welding the front edge of the assembly followed by the back edge will be referred to as R9, and the weld sequence that corresponds to welding the back edge followed by the front as R10. The top plates and lower sub-assemblies were sampled from two 16 piece production runs (32 in total). The experimental assembly details are listed in Table 1, where weld sequences R9 and R10 were evenly split over production batches 1 and 2. This was to ensure that any mean shifts between the weld sequences or production batches could be identified. There are two key sets of results, mean shifts, and variation propagation, which will be presented in the following sections.

Table 1  Weld sequence and production batch for each assembly sample.

<table>
<thead>
<tr>
<th>Samples</th>
<th>Weld sequence</th>
<th>Production batch</th>
<th>Plot notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-8</td>
<td>R10</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>7-16</td>
<td>R9</td>
<td>1</td>
<td>x</td>
</tr>
<tr>
<td>17-24</td>
<td>R10</td>
<td>2</td>
<td>*</td>
</tr>
<tr>
<td>25-32</td>
<td>R9</td>
<td>2</td>
<td>o</td>
</tr>
</tbody>
</table>

3.1 Principal Component Analysis (PCA)

A traditional approach to analysis of this data set in terms of mean shifts would involve the univariate analysis and comparison of each of the four sub-groups for each of the 6 surface points and 12 locating hole center-points. This equates to 42 individual MPs, as hole center-point observations consist of measurements in the x, y, and z axis. Conducting and distilling important information from such an analysis can be a challenging task due to the number of MPs, sub-groups, and paired comparisons that need to be considered. An alternative approach to gaining an overall impression of the data set is to use a dimensional reduction technique known as principal component analysis (PCA), which would allow a multi-dimensional data set to be represented by a smaller set of factors that describe most of the information or variation in the data. Applications of PCA in automotive body assembly research have appeared often as a tool for identifying major fault modes [15], and more recently to assist with the visualization and understanding of variation patterns [16, 17]. To transform a data set X into a reduced PCA space, a covariance matrix is first formed, from which eigenvalues $\lambda_i$ and eigenvectors $\Phi$ are computed. Eigenvectors are then sorted in descending order according to their corresponding eigenvalue. The original data set can then be represented as follows:

\[ b = \Phi (X - \bar{X}) \]

where, \( \Phi = (\Phi_1, \ldots, \Phi_n) \) and \( b \) represents observation (or sample) scores in the reduced t-dimensional PCA space.

3.2 Mean shifts

In Fig. 5 the original data set consisting of 42 MPs and 32 samples is represented by a 2-dimensional PCA plot. There are two factors or principal components (instead of 42 MPs) describing the data, PC1 and PC2, and each point indicated on the plot represents a measured sample. Referring back to Table 1, it can be seen that each of the 4 sub-groups occupy separate and distinct areas within the plot. This indicates that there are clear mean shifts between each sub-group: the PCA approach therefore provides a compact and informative representation of the overall trends in the data set within a single plot. Also of interest is that the PC1 axis appears to correspond to the differences between weld sequences R9 and R10, and the PC2 axis to differences between production batches 1 and 2. It should be noted that a common linear classifying technique, linear discriminant analysis [18], was applied to successfully partition the sub-groups in Fig. 5.

While PCA provides an overall impression of what clusters (or mean shifts) exist in the data set, it is also of interest to see how these mean shifts look in the original body co-ordinate space. Mean shifts between welding sequences in the original body coordinate systems are highlighted in Fig. 6. Here, a silhouette of the assembly is overlaid with vector plots showing the direction and relative magnitude of the mean shifts: the maximum mean shift here was 0.85 mm. Again, an asymmetric distribution of the mean shifts between assembly approaches can be seen. The extent of the shift also confirms the influence of welding sequence on the dimensional outcomes of compliant assemblies. While not indicated in Fig. 5, the maximum mean shift resulting from the different production batches was 0.4 mm.
3.3 Variation propagation

Variation in rigid component assemblies accumulates according to the additive theorem of variance: for rigid assemblies, variation can therefore be expected to increase throughout the assembly process. It has already been stated that this law does not hold for flexible component assemblies, such as sheet metal assemblies, and that variation can be absorbable. To illustrate this behaviour in a production setting, the variance of MPs measured on the sub-assemblies was compared against the standard deviations of the same MPs from the corresponding 8 sample assemblies. It can be seen that more MPs exhibit a decrease in variance than an increase, which again supports the characteristic of absorbable variation in flexible component assemblies.

Table 2 The percentage of MPs displaying statistically significant changes in variance (95% certainty) over the single station assembly process.

<table>
<thead>
<tr>
<th>Sub-group</th>
<th>Decrease</th>
<th>Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weld Sequence R10, Batch 1</td>
<td>17%</td>
<td>1%</td>
</tr>
<tr>
<td>Weld Sequence R9, Batch 1</td>
<td>17%</td>
<td>2%</td>
</tr>
<tr>
<td>Weld Sequence R10, Batch 2</td>
<td>15%</td>
<td>1%</td>
</tr>
<tr>
<td>Weld Sequence R9, Batch 2</td>
<td>17%</td>
<td>1%</td>
</tr>
</tbody>
</table>

5. Conclusion

This fundamental study serves to highlight the extent to which clamping and welding sequences can influence the dimensional outcome of a typical single-station production assembly, with significant observed mean shifts between different sequences. The ability of compliant assemblies to absorb variation is also observed. Given the demonstrated influence on dimensional outcomes of compliant assemblies, the study also provides important experimental context to research aimed at identifying optimal assembly processes for improved dimensional quality. Future experimental studies on the ability of welding and clamping sequences to better contain or counteract inherent process variation, such as the observed production batch mean shift, would be a valuable extension to the study presented.

Acknowledgement

The work presented here is supported by funding from the Australian Research Council (Linkage Grant #LP0560908), and kind contributions from Ford Australia.

References


Figure 6. Front cross assembly with measurement points and relative mean shifts indicated as vectors.
Non-nominal path planning for increased robustness of robotized assembling

D. Spensieri, J. S. Carlson, R. Söderberg, R. Bohlin, L. Lindkvist

Abstract: In the manufacturing and assembling processes it is important, in terms of time and money, to verify the feasibility of the operations at the design stage and at early production planning. To achieve that, verification in a virtual environment is often performed, by using methods as path planning and simulation of dimensional variation. These areas have gained interest both in industry and academia, however, they are almost always treated as separate activities, leading to unnecessary tight tolerances and online adjustments.

To resolve this, we present a novel procedure and its corresponding tool, which are able to compute the assembling path for an industrial robot in presence of geometrical variation in the robot’s links, in its degrees of freedom (dof), and in the environment. The method presented requires information about the distribution of the geometrical variation in the individual parts, or about how they are built together by their locating schemes. The kinematical model of the robot together with the 3D models of the robot and environment are also needed. The main idea to achieve assembly process robustness is to combine automatic collision-free robot path planning with variation simulation, enabling robots to avoid motions in areas with high variation and preferring instead low variation zones. The method is able to preliminarily deal with the possibly different geometrical variation due to the different robot kinematic configurations. An industrial test case from the automotive industry is successfully studied and the results are presented.

Keywords: Assembly - Computer Aided Manufacturing - Dimensional Control - Path Planning - Robotics - Quality Assurance

1. Introduction

A common scenario when simulating manufacturing and assembling processes includes an engineer trying to simulate the process by manipulating objects in a digital mock-up software. In highly geometrical restricted assembly situations, this procedure is often sensible to errors and is time consuming. Furthermore, it is common that such manufacturing and assembling tasks are performed by robots, whose motions are difficult to plan and control by manual programming. If we also consider that, in reality, every physical object is subject to geometrical variation due to its manufacturing process, then it appears prohibitive for an engineer to verify the feasibility of the assembling procedure at an early stage.

An automated verification is therefore helpful, since it can decrease the enormous costs that arise when realizing the infeasibility of an assembling plan late in the production phase, and the following need to re-design the process and/or the products.

Robots performing assembling operations are subject to variation as any other assembly system. Another source comes from robot resolution, mainly due to the precision of the computing system and to sensors and actuators sensitivity. Resolution affects the accuracy of the robot, which is a measure of how close to the nominal value the robot can reach. The accuracy, then, influences the repeatability, which is the ability of a robot to perform the same task in the same manner, see [1].

1.1. Motivation

One way to substantially improve positional accuracy is by online teaching the robot the poses it will assume during its assigned operations: in this way the robot controller stores its internal state and the data on how to perform the same task in the future, see [2] for advances in online programming environments. Anyway, programming robots online in order to perform tens or hundreds of tasks, with their respective paths and via-points, can be prohibitive.

Another way to achieve more accurate programs is by robot calibration: during this operation the mechanical parameters of the robot model are identified, see [1][3], e.g. by measuring the differences between the estimated position of the Tool Center Point (TCP) and the actual one. Anyway, the results may vary depending on in which area of the workspace the calibration is done, on the robot’s load, as well as on speed and acceleration of the motion. Moreover, it is difficult to compensate for uncertainties in the robot’s dofs due to sensors and actuators resolution. Furthermore, even if the TCP behaves in the virtual world exactly as in the real world, the robot might in reality collide with the environment, whereas its nominal model does not: this might be caused by geometrical variation introduced during the manufacturing and the assembling of the robot links.

However, in many cases, even after the identification of the mechanical parameters, it is not possible to update the models in the simulation software, since these are provided by third party actors or because the offline programming (OLP) is done only once, before calibration, or because the motions are programmed for a broad class of robots with similar geometric characteristics.

All these considerations motivate the fact that OLP software has to take into account geometric variation both in the robot parts and in the environment. Modeling and dealing with these aspects helps in producing robust and more reliable solutions, not so sensitive to the uncertainties described above. In this paper we have such a goal and propose a method for path planning of robots performing assembling tasks in presence of variation.

1.2. Contribution and related work
Our main contribution is to adapt and extend the ideas in [4] and [5], in order to handle non-nominal path planning for robot in presence of geometrical variation. First, a robot path is generated by state-of-the-art automatic robot path planning, then the path is post-processed by a smoothing algorithm trying to minimize the probability of collision. Variation analysis is performed in order to modify the distance measure needed by the smoothing algorithm.

Two main research areas can be identified related to this work:

- variation analysis and robust design;
- path planning in presence of uncertainties.

Many articles are dealing with geometrical variation, see [6], which have also resulted in software tools, see [7]. In this work the variation analysis is done according to [4] and is used to compute variation for the positions of the TCP along the robot path resulting from the path planning algorithm. For an overview of modeling methods for positioning rigid bodies by locating schemes, see Section 2.

The research area of path planning under uncertainties is wide, dealing with uncertainties in the models, in the sensors, and in static and dynamic environments, see [8] and [9] for an overview. In our application, the most relevant uncertainties come from variation in the environment and from the geometrical parts of the robot. In [10] the uncertainty in the robot is translated into a probability ellipsoid which is geometrically added to the moving object. The disadvantage with this method is that a required likelihood often leads to unfeasible paths. In our work, instead, we do not fix a priori the wished probability of a path being collision-free and we do not explicitly incorporate the probability ellipsoids in the geometrical models: this allows skipping computationally expensive steps. In [11] the path planning algorithm tries to minimize an objective function trading between the path length and the probability of collision, by computing lower and upper bounds for the collision probability and refining the bounds during the search, when needed. The problem is the computational complexity of the approach for complex 3D models. At the same way, the approaches in [12] and [13] show only results for simple landmark obstacles or for 2D models. Our work, on the other hand, based on [5], overcomes this limit by providing a general method to handle uncertainties using bounding volumes hierarchies. It also decouples the collision query phase from the sampling based search, by allowing the reuse of existing search methods. Furthermore, in this work, by integrating variation analysis with path planning, we can perform variation analysis on the robot configurations we are interested in (along the nominal path), and not on the entire robot configuration space: this allows flexibility in obtaining good variation information without prohibitive computational effort.

The outline of the paper is the following: in Section 2 the locating scheme and variation simulation are briefly described; in Section 3 an introduction to robot path planning is presented; Section 4 covers proximity queries in presence of variation in the models and how robot path planning can handle uncertainties. Section 5 illustrates an industrial case where the method presented is successfully applied.

2. Locating scheme and variation simulation

In order to determine the position and orientation of a rigid body six dofs must be locked. Different approaches are used, see [14] and [15], for example. Here, we consider the locating scheme method which consists in defining six points and then forcing the object to be in contact with them, see [6]. Given a deterministic locating scheme the position of the object is univocally identified. This model can be used to simulate the fixturing process for workpieces and for assemblies, as a robot for example. An industrial robot consists of a number of links joined with actuated joints. By considering the links as rigid bodies, it is possible to build an assembly model of the robot, see [4]. Depending on the geometry of the joint this can be done in a number of ways. Here we consider the most common type of joint: the revolute joint.

![Figure 1. Revolute joint with its locating scheme.](image)

One type of revolute joint consists of an axis that can rotate relative to two holes in a yoke, see Figure 1. The position of the part connected to the axis is mainly determined by the contacts between the axis and the holes. In case the locating scheme can be defined in the following way:

- the first three points are located at the center of one of the holes; they lock the axial motion of the axis and the radial motion at that hole.
- the fourth and fifth points are located at the center of the other hole; they lock the radial motion at that hole.
- the sixth point locks the remaining degree of freedom which is rotation around the axis; it is placed at the radius of the hole.

The variation in the first five points can be determined from the tolerances on the mating surfaces, the axis and the holes. The last point does not represent a physical contact but can be used to introduce variation in the angle of the joint, which is determined by the actuators.

An industrial robot usually consists of six links and joints, determining 36 locating points, 6 for each locating scheme. By varying a locating point \( i \) by \( \Delta \epsilon_i \), it is possible to determine the orientation and the position of each robot link and of the TCP. Usually, loads and welding guns are attached to the TCP and the accuracy for the operations that the robots are designed to perform, e.g. sealing, stud, and spot welding, heavily depend on the accuracy of the TCP. From these considerations, in this work, we limit the variation analysis to the orientation \( R_{TCP} \) of the TCP and of its position, \( t_{TCP} \), and to the attached loads and tools.

3. Automatic robot path planning

In the last decades, with the growth of autonomous robots and virtual manufacturing, the path planning problem has received enormous interest. Many articles have been published and software tools addressing industrial problems have been generated. A general principle for path planning algorithms is to search for a connected path between a start and a goal node in the graph modeling the configuration space of the moving object (a robot in our case), while avoiding collisions with the
environment, or maintaining a clearance above some critical threshold, see [16][17][18].

Figure 2. An illustrative configuration space, with collision areas (in black), start and goal configurations (graph’s red vertices), samples (graph’s black vertices), and connecting paths (graph’s edges).

The configuration space is often sampled, see Figure 2, and collision and distance queries are performed. Then, in order to find a good trade-off between completeness and computing time, often lazy approaches are used, which try to delay collision queries, until needed, see [19].

4. Non nominal path planning

By using the methods in Section 3 we will compute a collision-free path for the robot, only considering the nominal models. Afterwards the path produced will be smoothed in order to deal with the variation in the robot and in the environment, see Section 4.3.

4.1. Non-nominal proximity queries

A class of powerful algorithms doing distance queries is based on bounding volumes hierarchies: a bounding volume (BV) of a set of objects is a volume that encloses all the objects in the set and a hierarchy of BVS is a tree whose nodes represent BVS. Different shapes for BVS are possible (box, cylinder, swept sphere, polytope, etc.), anyway the common idea is to search the tree until the proximity query is answered.

The leaves of the tree are usually triangles. In [5] it has been shown how to modify the distance between triangles in order to handle non nominal models. In that work, the measure of distance between triangles \( T_i \) and \( T_j \), taking into account the variation is defined as

\[
\alpha(T_i, T_j) = P\left( X^2 \leq \frac{d}{\sigma_i + \sigma_j} \right) \quad (1)
\]

where \( d \) is the actual distance between the triangles, \( X^2 \) represents the chi-squared distribution for 3 dofs, \( \sigma_i \) and \( \sigma_j \) are the triangles respective standard deviations. For the adaptation of BVS to this distance, see [5]. The disadvantage of this distance measure is represented by plateaus over large part of the working range, since the ratio \( \frac{d}{\sigma_i + \sigma_j} \) stays in the domain of the chi-squared cumulative distribution, where its value is close to 1. To avoid that, one can use, instead of (1), another monotone function with the property of having more sensitive values around the critical distance range. Our choice is:

\[
\alpha(h_i, b_i) = \beta \frac{d}{\sigma_i + \sigma_j} \quad (2)
\]

where \( \beta \) depends on the particular application.

4.2. TCP and tool variation

In Section 2 we have seen how the orientation and position of the TCP are computed when small displacements are applied to the locating schemes. Each point \( P \) belonging to the tool or to the load attached to the TCP is therefore influenced by the variation \( \varepsilon \). By considering a vector valued function for \( P \), \( p(\varepsilon) : R^3 \rightarrow R^3 \), with \( N = 36 \), and approximating it by a first order expression, then the following holds:

\[
p(\varepsilon) = p(0) + \sum \frac{\partial p(0)}{\partial \varepsilon_i} \varepsilon_i \quad (3)
\]

The partial derivatives for each point \( P \) are evaluated numerically by applying small displacements \( \Delta \varepsilon_i \) to the variable \( \varepsilon \) and analyzing the behaviour of the TCP. Therefore, we have

\[
\frac{\partial p}{\partial \varepsilon_i} \Delta \varepsilon_i = \frac{p(\Delta \varepsilon_i \hat{\varepsilon}_i) - p(0)}{\Delta \varepsilon_i} = \frac{R_{\text{TCP}}(\Delta \varepsilon_i \hat{\varepsilon}_i) p_i(0) + t_{\text{TCP}}(\Delta \varepsilon_i \hat{\varepsilon}_i) - R_{\text{TCP}}(0) p_i(0) - t_{\text{TCP}}(0)}{\Delta \varepsilon_i}
\]

where \( p_i(0) \) is \( P \) in the coordinate system of the TCP. Note that \( R_{\text{TCP}}(\Delta \varepsilon_i \hat{\varepsilon}_i) \) and \( t_{\text{TCP}}(\Delta \varepsilon_i \hat{\varepsilon}_i) \) are computed by forcing the points in the locating scheme to be in contact with their attached surfaces, as described in Section 2. Thus, (3) may be rewritten in matrix form as

\[
p(\varepsilon) = p(0) + A_p \varepsilon \quad (4)
\]

with \( A_p = \left[ \frac{\Delta p_1}{\Delta \varepsilon_1}, \ldots, \frac{\Delta p_p}{\Delta \varepsilon_p}, \ldots, \frac{\Delta p_{36}}{\Delta \varepsilon_{36}} \right] \). By assuming \( \varepsilon \) with mean zero and covariance \( \Sigma_p \), the mean vector \( \mu_p \) for \( P \) and its covariance matrix \( \Sigma_p \), are given by

\[
\mu_p = p(0)
\]

\[
\Sigma_p = A_p \Sigma_p A_p^T
\]

Note that the Jacobian \( A_p \) depends on which point \( P \), attached to the TCP, is considered. All these computations can be done when a specific robot configuration is fixed. In order to gather more information about the covariance along the robot motion computed as in Section 3, a number of robot configurations are sampled along the path and the analysis described above is carried on for each of them.

4.3. Path smoothing

Our goal now is to modify the existing assembling path in order to make it more robust to variation existing in the robot links and in the environment. To achieve that, we use the distance measure described in Equation (2). Its effect is to decrease the probability of collision, given a predefined order expression, then the following holds:

\[
p(\varepsilon) = p(0) + \sum \frac{\partial p(0)}{\partial \varepsilon_i} \varepsilon_i \quad (3)
\]

To smooth the path it is necessary to provide good approximations of the clearance function for robot configurations that have not been sampled. One way to do that is to first interpolate the standard deviation for each triangle for that particular robot pose, then updating the bounding volumes hierarchies, and eventually computing the clearance according to Equation (2). An easy way to interpolate the variation for a
triangle $T$, when the robot assumes the joint configuration $j^*$, is to take a convex combination of the variations for the closest robot samples 1 and 2 along the path:

$$\sigma_r(j^*) = \frac{d_1\sigma_r^1 + d_2\sigma_r^2}{d_1 + d_2}$$  \hspace{1cm} (6)

where $d_i = d(j^*, j_i)$ is the distance induced by some norm in the robot configuration space. In this work, the max norm is used: $d(j^*, j_i) = \|j^* - j_i\|_\infty$. It might be worth to increase the number of samples along the robot path to achieve better accuracy.

5. Industrial test case

We illustrate the results of applying the algorithms described in the paper on an industrial test case, consisting of a six dofs robot assembling a driving unit into a car body, see Fig. 3.

![Industrial robot assembling a driving unit into a car body](screenshot from Industrial Path Solutions, IPS, see [20]).

The environment is very cluttered due to the geometrical restrictions of the door opening. The workspace is further restricted if one considers that a very large load is attached to the tool, which is among the parts of the robot mainly influenced by variation in the model. Indeed, in this case a large gripper holding a driving unit has to be placed in the car: the path can both be seen as assembling/disassembling operation. The graphs in Figures 4 and 5 show the modified clearance function (2) along the path. The original path has several peaks under 6 (note dimensionless unit), whereas after applying the algorithm described in this work one can see several plateaus around 6. This increased robustness is directly translated into higher manufacturing quality.

![Graph for the robot clearance, modified as in (2), along the robot path, after smoothing.](image)

6. Conclusion and future work

In this article we have described a novel method that combines robot path planning and variation simulation in order to increase robustness for automatic assembling tasks performed by robots in presence of geometrical variation in the robot and the environment. The study on an industrial scenario is promising, showing that the probability of collision can be decreased with respect to a path computed by a standard approach. Anyway, the model can be further improved to better estimate variation outside the pre-computed path and to decouple the errors due to the joints and the ones resulting from the assembling.

Acknowledgements

We thank the whole Geometry and Motion Planning Group for useful discussions and help in the implementation. This work was carried out at the Wingquist Laboratory VINN Excellence Centre, and is part of the Sustainable Production Initiative and the Production Area of Advance at Chalmers University of Technology. It was supported by the Swedish Governmental Agency for Innovation Systems.

References


International Seminar on Computer Aided Tolerancing, Enschede, the Netherlands, March 22-24, pp. 231-240.


Statistical shape modeling in virtual assembly using PCA-technique

B. Lindaua,b, L. Lindkvista, A. Anderssona,c, and R. Söderberga

aDepartment Product and Production Development, Chalmers University of Technology, SE-412 96 Göteborg, Sweden
bDepartment 81210, TVG, Geopl. PVÖE 101, Volvo Car Corporation, Göteborg, Sweden
cDepartment 81280, TVG, Geopl. 26GO1, Volvo Car Corporation, Olofström, Sweden

Abstract: The use of virtual assembly tools is one way to understand and improve the geometric product tolerance setting and the conditions for successful manufacturing. Recent developments enable consideration to be given to the deformability of parts when joined. However, in order to produce reliable results, the geometric faults of the mating surfaces must be correctly assumed. In this paper, statistical shape models built on the Principal Component Analysis-technique are proposed to be used to describe the part variation. A generalized model is presented. Furthermore, the underlying intentions and implications are discussed. The method is exemplified using the tool RD&T. In the presented case, a non-rigid sheet metal assembly is modeled and distorted to create a set of sample shapes from which a statistical model is built. In the result, the statistic representation bears a good resemblance to the nominal distorted model when the two are compared.

Keywords: sheet metal, simulation, geometric modeling, and analysis.

1. Introduction

In Automotive car body assembly shops, non-rigid sheet metal parts and subassemblies are joined to build bigger structures into a final car body. In these factories, or so called Body in White (BiW) shops, the prevailing joining method is spot welding. Depending upon the achieved degree of robustness in part and tool design, the produced items tend to deviate more or less from their nominal specification. Robust part and tool designs sets the prerequisites for success, as improved quality, reduced lead times and development costs are essential factors to be competitive in the fierce global car market [1].

The use of virtual assembly tools aims to catch eventual non-robust solutions early in the development phases, thereby minimizing time-consuming, expensive testing and trimming activities late in the development- and industrialization phases. This is one way to understand and improve the geometric product tolerance setting and the conditions for successful manufacturing, fulfilling esthetical and functional demands. The same tools have the potential to be used in production, for welding. Depending upon the achieved degree of robustness in part and tool design, the produced items tend to deviate more or less from their nominal specification. Robust part and tool designs sets the prerequisites for success, as improved quality, reduced lead times and development costs are essential factors to be competitive in the fierce global car market [1].

The use of virtual assembly tools aims to catch eventual non-robust solutions early in the development phases, thereby minimizing time-consuming, expensive testing and trimming activities late in the development- and industrialization phases. This is one way to understand and improve the geometric product tolerance setting and the conditions for successful manufacturing, fulfilling esthetical and functional demands. The same tools have the potential to be used in production, for deeper analysis and a possible better understanding of the part and process behavior (see Figure 1).

Figure 1 RD&T Geometry Assurance Tool-box (Söderberg).

Rigid modeling is still much in use in practice. However, in order to take into account the non-rigid behavior, there is work to be done to verify the model accuracy and validate the usefulness of predictions made [2-5].

1.1. Problem description

Recent developments (within virtual assembly) enable consideration to be given to the deformability of parts when joined. However, in order to produce reliable results, the geometric faults of the mating surfaces must be correctly assumed. There is a need for a generalized method for how to represent the part deviation (offset and variation) in the detailed assembly simulation of non-rigid parts, representing the whole shape of the details. Demands on concurrent engineering, the amount of remodeling, reuse in later phases and the fact that the available detailed information differs as the projects evolves are aspects that ought to be taken into account.

1.2. Scope of the paper

A generalized method using a straightforward Principal Component Analysis (PCA) technique is presented and discussed. The thoughts are exemplified by a case, studying a distorted nominal sheet metal part assembly.

The paper is organized the following way. A more detailed background is provided in Section 2. In Section 3, the proposed method/model is presented. A case that exemplifies the ideas is presented in Section 4, followed by discussion in Section 5 and finally conclusions in Section 6.

The work was influenced by previous research contributions in non-rigid sensitivity and root cause analysis (see Section 2.2.). Furthermore, the work of I.T. Jolliffe [6] and the use of PCA in other research areas, such as atmospheric [7] and medical science [8], have made an impact on the choice to use this well-known technique in the study of multivariate data.

2. Background

The virtual tool RD&T has been used in the study as a demonstrator. The tool is primarily used for assembly simulations focusing on the geometric stability, sensitivity and variation analysis of complex products, taking into account both product and tooling design. The functionality to perform non-rigid simulations is implemented and verified on small sub-assemblies. There is still a need for research to propose and validate efficient working methods and to verify the accuracy of predictions made for large assemblies in a real industrial setting.

RD&T uses the "Method of Influence Coefficient" MIC-technique, based on linear "mechanistic variation models" in combination with the Monte Carlo Simulation [9]. Contact
modeling has been introduced [10], and made more efficient [11], hindering the penetration of adjacent parts. Joining order studies have been performed with promising results [12].

2.1. A Simplified RD&T, description

When analyzing a compliant sheet metal assembly process in RD&T, meshes for each included part are imported (on an ABAQUS file format, for instance) and the meshes’ material properties, such as Young’s modulus (E), poison ratio (v), material thickness (t) and density (ρ), are given as input. Each part or assembly is positioned according to the fixture locating scheme used in the joining process to be simulated. One example of such a scheme is the N-2-1-system. Weld points are defined, identifying the node pairs to be joined and the type of gun, fixed or balanced. For contact modeling, automatic contact search can be performed, and reductions can be made to cope with any calculations that are too heavy. Finally, the locating scheme in which the resulting springback calculation is performed is defined. Tolerances and dependencies for specific nodes on part or points on fixture devices can be specified in the system. Analysis measures such as gap, flush, point distance and so forth are used, depending upon what to be analyzed. The sensitivity matrices are created in the tool using the unit displacement method (MIC) and an internal FEM-solver to define the stiffness matrices. Once stiffness and sensitivity matrices are in place, sensitivity analysis can be performed and variation analysis can be conducted using Monte Carlo Simulation techniques (see Figure 2).

However, in order to produce reliable predictions of offset and variation, the geometric faults of the part deviations must be correctly assumed (see Section 3). Detailed descriptions of mating surfaces are vital for contact modeling. Even if RD&T is based on the MIC-technique, the whole shapes of the mating surfaces are vital for contact modeling. Even if RD&T is correctly assumed (see Section 3.), Detailed descriptions of and variation, the geometric faults of the part deviations must be represented in the meshes can be studied.

2.2. Previous research

As mentioned, the ideas proposed in this paper were influenced by previous research contributions in non-rigid sensitivity and root cause analysis using PCA and Covariance techniques.

To identify the sources of dimensional variation in automotive body assembly, Hu and Wu et al. [13] proposed the use of the PCA-technique as a systematic approach. Their study focused upon the root cause analysis of measurement data in body-in-white, BIW.

Camelio, Hu et al. [14] proposed a method to use “Component Geometric Covariance” to predict the assembly variation of non-rigid parts. They combined PCA with finite element analysis to estimate the part or component contribution in assembly variation analysis. The PCA-technique was used to extract deformation patterns from production measurement data. The component covariance was decomposed into individual contributions of these deformation patterns. The effects of each pattern were then calculated using the finite element methods and the method of influence coefficients. The aim was to find computational effective methods with a reduced number of variables, with a focus on sensitivity analysis.

Liu and Hu [15] proposed the use of Designated Component Analysis (DCA) for fixture fault analysis. They focus on the difficulty to interpret the variation patterns due to the “compounding” effects of PCA as it is a totally data driven process. Unlike PCA, the DCA method starts with a definition of orthogonal variation patterns with known physical interpretation. Using this characteristic, the different components can be analyzed and monitored individually, as the variation can be decomposed into the components and removed from the original data domain. They also point out that correlation analysis among the DCA components can be performed for diagnostic purposes.

Camelio and Hu [16] proposed a methodology that “…integrates on-line measurement data, part geometry, fixture layout and sensor layout in detecting simultaneous multiple fixture faults.” In this work, they continue the study using DCA for the multiple fault diagnosis of sheet metal fixtures. Here, however, they extend the methodology using patterns that are not necessarily orthogonal, a condition for DCA. To resolve this, they use a Gram Schmidt algorithm to find an orthonormal basis for the subspace spanned by the designated patterns.

Ungemach and Mantwill [17] used the MIC and extended it with a retroactive contact consideration-procedure. A classic linear springback calculation is first carried out. Thereafter, the contact calculations are performed, taking into account the deformation state. They conclude that the approach enables the complex nonlinear problem to be approximated with little additional computing effort. Furthermore they state that the methodology can be extended to multi-station assembly processes by integrating it with the methodology developed by Camelio and Ceglarek [18].

Yue et al. [19] focus on Product-Oriented Sensitivity Analysis in a multi-station non-rigid assembly situation, introducing indices (pattern-, component- and station sensitivity index) to measure the influence on the variation. They proposed methods for how to obtain ranges (max, min) of these indices, which they use to estimate the sensitivities without any information about incoming variation. A characteristic useful in early design stages when the knowledge of the components variation is limited.

However, in this paper we look at the use of PCA as a descriptive tool, creating a statistical model describing the shape deviation of the whole detail as a single part or as it behaves within a subassembly.

3. Proposed method/model

The model in Figure 3 was primarily created to visualize a method for how to represent the part deviation in detailed variation analysis, in tools like RD&T based on the MIC-technique.
Sets of part shapes, here called shape training set (N-shapes), can be used to describe the deviations. Shape training sets can be generated from several sources:

- **σF, σM**: with today’s increased use of forming- and molding simulation tools.
- **Scanning**: often a small number of parts or subassemblies are scanned. The details are often coming from the same production batches. Today in production, scanning is mostly performed for problem solving purposes.
- **CMM**: from production environment, geometric measurement data describe the deviations in certain control key points, mainly acquired with CMM or other point measurement methods. These measurements provide the main geometric historical documentation of produced parts.
- **Result**: from the assembly simulations of lower level structures.

Holes and edges are proposed to be handled separately, not as part of the actual detail shape. That is because these features can have been produced in separate processes, e.g. stamped, drilled, cut or bended. In assembly studies of sheet metals, the variation of hole positions can be added to the location scheme (Fixture & References) or be coupled to the hole centers. Edges can be generalized as N-contours, and a similar method as proposed to handle the N-shapes can be applied.

### 3.2. Nominal modeling

Residual forces are not taken into account in this model; only the pure geometrical deviations are considered. The main reason for this is that today in practice we know little about the internal residual forces from the geometric measurement and scanned data available from production environment. Furthermore, the part deviations are assumed to be small and, therefore, only elastic deformations are considered.

The actual changes in stiffness due to the assumed small shape deviations are also neglected. A linear “mechanistic variation model” are built, starting from meshes describing the nominal shapes [9]. In Figure 3 above, the square named ‘Nom FEM’ represents these nominal meshes, in combination with an internal FEM-solver and additional material properties, such as material thickness (t), Young’s modulus (E), and poison ratio (ν), for stiffness calculations (see Figure 2).

When modeling an assembly process, the different parts are positioned in relation to each other by their part locating schemes. The reference and support points (RSp) and their attending steering directions define a N-2-1 system for each part. The RSp-points are coupled to the corresponding nodes on the part meshes. Joining points (spot weld points (Wp), for example) are defined by the pair of nodes to be connected. For contact modeling, the associated pairs of nodes (Cp) are defined to hinder adjacent parts to penetrate each other. RSp points are also defined for the springback calculation. For variation and contribution analysis, measurements (Mp) can be defined in the model (see Figure 2 & 3). All these definitions are couplings made to the nominal meshed models. With this nominal input, stiffness and sensitivity matrices are calculated, which is later used in the MCS-calculation.

### 3.3. Part deviation modeling

The shape deviations are continuous and correlated by their nature. Therefore, the deviations in each node must be coupled, not independently varied in the MCS calculations (see Figure 4a).

One way is to treat each of the N-shapes’ deviation from nominal as a separate observation and to use these sequential sets in the following MCS calculation, as shown by the dotted line in Figure 4b. To each node are then coupled N-distances, in nominal normal direction, representing the deviation to each shape in the training set. Using this method the model sizes will grow fast.

Another way is to find a statistical model, keeping the correlation between the nodes. It is for this descriptive purpose the “Principal Component Analysis”-technique comes in handy. By calculating the mean deviations and by performing PCA-analysis of the centered observations in the nodes, the data can be compressed. This use of PCA is a purely data driven process to find the orthogonal directions and variations (principal component vectors and eigenvalues) to be used in the following MCS-calculations. The orthogonal property makes it possible to treat the principal components separately. As long as the components are not rotated, the additive property is preserved and each component can be handled separately. The number of components (number of modes (m)) to be included has not been the primary focus in this study. The Cumulative Percentage of Total Variation is one of several methods [6] that can be used.

The eigenvalues (λ) define the variances for the modes and the component vectors define the loadings (L), the weight, to be used in each node for each mode. To each node is also linked its mean deviation (see Figure 4b).
Performing an assembly simulation creates new sample shapes of the details, showing their non-rigid behavior in the assembly. By running a new PCA on these results, the observations can be compressed, described by new means and PCA-components. This result can then be used in simulations on the next structural level. PCA is then proposed to be conducted in portions, on each detail separately, which brings down the number of nodes in each PCA calculation.

When using stored PCA-analysis results of separate details simulated on a lower assembly level, these details cannot be varied individually in the MCS calculation on a higher structural level. This is because the joined parts, in the subassembly, do not move independently in the joining nodes. One way to handle this is to let the part with the largest number of PCA components define the variances used in the MCS-calculation, and to find new loadings for the nodes describing the other parts in the assembly. This is possible if there are more joining points than PCA-components, describing each detail, giving rise to a linear equation system with more knowns than unknowns, taking into account that the deviations in the joined nodes shall be equal.

Suppose that the shape deviations in detail A are represented by \( p \) PCA-components and detail B by \( q \) components. Furthermore suppose that the details A & B are joined in \( k \) weld points. If \( q \leq p \) and \( q \leq k \), then the following can be derived.

The variances (PCA-eigenvalues) are used in the MCS-calculation to vary the shape deviation. In each loop new deviations are calculated for each PCA-component:

\[
\alpha = \{ \alpha_1, \alpha_2 \ldots \alpha_p \} \quad \text{for detail A} \tag{1}
\]

\[
\beta = \{ \beta_1, \beta_2 \ldots \beta_q \} \quad \text{for detail B} \tag{2}
\]

The distances from nominal in the joining points:

\[
d = \{ d_1, d_2 \ldots d_k \} \tag{3}
\]

PCA loadings connected to node pairs, \( i = 1 : k \), which defines the joining of details A and B are as follows:

\[
a = \{ a_1, a_2 \ldots a_p \}^T \quad \text{for detail A} \tag{4}
\]

\[
b = \{ b_1, b_2 \ldots b_q \}^T \quad \text{for detail B} \tag{5}
\]

\[
A = \{ a_1, a_2 \ldots a_p \}^T \quad \text{for detail A} \tag{6}
\]

\[
B = \{ b_1, b_2 \ldots b_q \}^T \quad \text{for detail B} \tag{7}
\]

Distances in the joining points shall be equal:

\[
d = \alpha A^T = \beta B^T \tag{8}
\]

Now the following can be derived:

\[
\alpha A B = \beta B^T B \tag{9}
\]

\[
\alpha A (B^T B)^{-1} = \beta (B^T B)^{-1} \tag{10}
\]

New loadings \( L_{AB} \) for detail B in relation to \( \alpha \) can be calculated:

\[
\beta_{LB} = \alpha_{LAB} \tag{11}
\]

\[
L_{AB} = A^T (B^T B)^{-1} L_B \tag{12}
\]

where \( L_B \) represents the initial loading matrix for detail B.

In this manner, it is possible to couple the parts in the MCS so that each detail is not handled separately. If a detail is welded further out in the structure, calculate first new loadings for the detail connected to the master (detail A). Thereafter, use this detail as a new master for the detail further out in the structure.

Finally, caution must be taken when using equation (12) if the number of Wp:s connecting detail A to detail B are close to or equal the number of modes describing the variation of detail B. The deviations of details A and B in the WP:s will be approximately the same when described by a subset of PCA-components. Unwanted extrapolation and enlargement factors can be the result if caution is not taken.

4. Case study

The aim of the case is to demonstrate the proposed use of PCA-components, representing the shape deviation. The simulated assembly consists of 5 parts, joined by spot welding (see Figure 5 and Table 1). The study is performed in 3-steps:

- Creation of reference training sets.
- Applying the proposed method on the training set.
- Evaluating the result by comparing the color maps.

4.1. Creation of reference training sets.

The aim of this step was to create a controlled set of sample shapes (see Figure 6). Historical CMM-measurement data from production environment has been used to distort a nominal assembly creating shape training sets (25-shapes) for each detail in the assembly. The part assembly was modeled in RD&T by its nominal FEM representation and positioned in a N-2-1 fixation, representing the measurement fixture. Nodes representing the nominal welds were connected with rigid beam elements. Support points were added for each key control CMM-measurement point (Key Points), defined in the normal direction to the part surface. To each of these support points, sequential sets of “correlated” deviations, representing the measurements, was coupled.

Using the non-rigid simulation functionality, the mesh was distorted, according to the measurement deviations, creating 25 deviations in normal direction for each node and detail, representing 25 sample shapes for each part in the assembly. The whole part is distorted not only in the key points.

4.2. Applying the proposed method on the training set.

The 25-shape deviations of each whole detail were represented by the deviation in normal direction (from nominal) at each node point. PCA-analysis was performed on each detail separately calculating the eigenvalues (the variances), \( \lambda \), and the
The cumulative percentage of total variation was calculated using the following formula:

$$C(m) = 100 \sum_{i=1}^{m} \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j}$$  

(13)

where \( p \) is the total number of components. The cumulative percentage of total variation for each detail is shown in Figure 7.

The Cumulative Percentage of Total Variation is one of several methods [6] that can be used to choose the number of components to be included in describing the variances of the parts. It was chosen as an easy, straight forward method. The choice of the number of components has not been the primary focus in this study.

As no clear 'elbow' can be detected in the outer and inner graphs, but a change in direction can be seen for ReinfR at approximately 98%, this limit was chosen in this study. With the cumulative limit set ≥ 98 %, the different parts are represented by different numbers of modes (the number of chosen components). These components can then be used in the assembly simulation on the next structural level.

As the details are joined, they cannot be varied independently in the next simulation level. The loadings for Inner, Reinf Left and Reinf Right were therefore transformed to follow the variations of part Outer, using equation (12). The loadings for the Nut plate were transformed to follow the transformed loadings for the Inner part, as it is joined further out in the structure. The original amount of modes (PCA-components) and after the transformation is seen in Table 2.

For each detail, the statistical representation was imported to RD&T, consisting of mean deviation in each node, loadings for each node and mode, plus the variance belonging to each mode. In this case, the variances were given by the Outer part.

### 4.3. Result

In Figure 8, color coded results are shown. The color codes show the part variation in normal direction.

The bottom view shows the variation in the created reference training set. The red arrows indicate the key control points where the nominal assembly was disturbed.

The top view shows the result after applying the proposed method on the sample shapes, the training set, using PCA-technique followed by MCS-calculations.

As expected, we have a good resemblance. The example shows how an N-shape representation can be used to create a statistical representation. In this case, the 25-shapes are compressed to 10 components + mean deviation, a relation (11/25), when the cumulative percentage of total variation limit was set at 98%. As the creation of the training set can be seen as an assembly simulation on a lower level, the case also indicates how results from a lower level assembly simulation can be represented by a statistical model, to be used for further assembly studies on a higher structural level.

### 5. Discussion

A lot of research and tool development is done in the area of forming and molding simulation [20-23]. With the firm belief that these tools are getting better and better at predicting the actual shapes, the presented model focuses on virtual assembly. The assumption is that these virtual tools ("F" and "M") will include variation packages, which enable the export of a number of sample shapes representing the predicted variation. One example of this is the forming simulation package Autoform with its variation package Autoform Sigma, from which it is possible to generate a number of sample shapes in, for example, VRML-format.

With the generalization, seeing the different sources of input as a number of shapes, holes and edges, the different kinds of data sources can be handled similarly. Building the simulation models on nominal base can potentially increase the reusability, minimizing the need for remodeling, which includes coupling of the necessary information, reference schemes, joining points, and contact points to the corresponding nodes. This is possible under the basic assumption of small deviations in relationship to each part's total size and the fact that the changes of stiffness due to the geometry deviations are neglected. Therefore, stiffness and sensitivity matrices are calculated using nominal shapes, using the MIC-technique.

Demands on concurrent engineering, the amount of remodeling, reuse in later phases and the fact that the available detailed information differs in the development phases, are factors that have been taken into account in the proposed method. With a statistical representation, a pure MCS technique can be used to perform variation analysis. It is possible to change statistical model, depending on available data. In early prediction, input could come from forming and molding simulations used in lower structure assemblies or from lower level assembly simulations, depending upon the studied object's structural level in the assembly sequence. Later, in pre-production, these predictions could be verified using scanned and/or CMM measurement data. In running production, key control point-measurements or scanned data can be used.
strengthens the reusability, as different training sets can be used depending upon project phase and available datasets.

PCA was chosen to create the statistical model as the shape deviations are continuous and correlated by nature (and therefore must be coupled, not independently varied in the MCS calculation). Furthermore, PCA is a well-known multivariate technique, and the computation is a straight forward, data driven process. PCA components are orthogonal and independent; they are seldom directly interpretable to physical aspects, as mentioned in the work by [15-16]. Care must be taken in the root cause analysis of the components and their physical interpretation. The rotation of the components cannot be performed in a straight forward manner without losing independency. The PCA components are still useable in variation contribution and sensitivity analysis, as long as their contribution is evaluated, not necessarily the precise underlying physical aspects.

The PCA-components can also be used to compress assembly simulation results, and to make the results usable in studies on higher structural levels. This enables further use of the data in studies on the next structural level.

The PCA calculations are proposed to be performed on each detail separately to minimize the number of nodes in each PCA calculation. It also simplifies the identification and coupling of the statistical information to each node. As the parts cannot be varied independently in the MCS calculation, equation (12) is derived, making it possible to express each parts variances by the part variances with the highest number of components, using the fact that the weld points are connected and the deviations shall be the same in these points.

Finally, deeper studies have to be performed, looking into details concerning the generation of training sets from different kinds of sources. Methods for how to choose the right amount of components to represent the part variation also need to be studied more in detail. Furthermore, efficient methods to judge the need to introduce internal contact modeling within a joined assembly have to be developed. If assembly internal contact modeling is needed, on the higher structural level, then the retroactive contact considering-procedure proposed by Ungemach and Mantwill [17] might be useful.

6. Conclusion

The realistic modeling of the shape of mating surfaces is a vital precondition for contact modeling. That is true, for example, in detailed studies of geometric fault propagation due to weld order. Statistical shape models are proposed to be used to describe the part variation. It is shown that sets of part shapes (here called shape training sets) can be used to create statistical shape models, based upon PCA-technique, to simulate deviations of the part. The part’s shapes (their geometry) are then represented by nominal meshes together with a statistical representation of the part deviation in each node, to model the whole part.

Acknowledgement

The authors acknowledge the financial support provided by Fordonsstrategisk Forskning och Innovation (FFI), VINNOVA and Volvo Car Corporation (VCC). A special thanks to Samuel Lorin PPU at Chalmers and Carl Angervall at VCC for constructive discussions on the subject.

References

A bio-inspired approach for self-correcting compliant assembly systems

L. J. Wells, J. A. Camelio

Grado Department of Industrial and Systems Engineering, Virginia Technological University, Blacksburg, VA, USA

Abstract: Statistical process monitoring and control has been popularized in manufacturing as well as various other industries interested in improving product quality and reducing costs. Advances in this field have focused primarily on more efficient ways for diagnosing faults, reducing variation, developing robust design techniques, and increasing sensor capabilities. However, statistical process monitoring cannot address the need for instant variation reduction during assembly operations. This paper presents a unique dimensional error-compensation approach for compliant sheet metal assembly processes. The resulting autonomous self-correction system integrates rapidly advancing data mining methods, physical models, assembly modelling techniques, sensor capabilities, and actuator networks to implement a part-by-part dimensional error compensation. Inspired by biological systems, the proposed quality control approach utilizes immunological principles as a means of developing the required mathematical framework behind the self-correcting methodology. The resulting assembly system obtained through this bio-mimicking approach will be used for autonomous monitoring, detection, diagnosis, and control of station and system level faults, contrary to traditional systems that largely rely on final product measurements and expert analysis to eliminate process faults.

Keywords: Compliant Assembly, Variation Reduction, Dimensional Error Control

1. Introduction

Traditionally, two approaches have been implemented to reduce variation and increase productivity in assembly systems: 1) robust design and 2) statistical monitoring and control. Robust design methodologies are generally implemented at process and product design stages. As an example, simulation models have been developed to evaluate the impact of component tolerances on the quality of an assembled product [1]. Evaluating variation propagation in an assembly system during design allows for the development of more robust processes and products. However, assembly systems are complex and it is impractical to design a perfect system that is insensitive to the impact of all process/noise variables. Therefore, quality control systems are placed in production environments to reduce variability and ensure that assembly processes result in high product quality.

Common quality control approaches use control charts to detect abnormal quality conditions or faults. Once a fault is detected, an assembly line is halted while engineers and operators work to determine the source of the problem. A faulty process or component must then be located, analyzed, and repaired or replaced. Currently technologies are generally used to monitor final product characteristics, but rely on operator knowledge and expertise for problem compensation. Delays in problem detection and diagnosis result in extended production downtimes for manufacturing operations and defective intermediate products.

With advancements in sensor and measurement technologies, 100% sampling in assembly processes has been achieved in the form of high-density data sets [2,3]. This achievement has led to a third approach towards variation reduction in assembly, active error compensation. These new methods rely on continuing advancements in programmable tooling, in-line dimensional measurement systems, and advanced variation propagation models to adjust the process on a part-by-part basis. Mantripragada and Whitney [4] developed a stage transition model allowing for control theory to be applied to assembly propagation problems. In this paper, part measurements before assembly were used to calculate control actions to minimize the variation of a product’s key characteristics. Jin and Ding [5] developed a design of experiment based Automatic Process Control. In this work, regression models were used with in-line observations of observable noise factors to automatically adjust process control factors to minimize quality loss. Djurjanovic and Zhu [6] proposed the use of a state space model to control variations in multi-station machining applications. This control strategy consisted of feedback and feed-forward control to adjust tools paths and fixture positions. Izquierdo et al. [7] showed that when dealing with rigid sheet metal parts, optimal control action for multi-stage assembly processes can be determined based on optimizing explicit kinematic relationships.

Xie et al. [8] expanded upon the work of Izquierdo et al. [7] to compliant sheet metal assemblies. Their paper used expert knowledge of possible dimensional fault patterns in a compliant sheet metal assembly process to develop a collection of pre-determined control actions to perform active error compensation. These control actions were learned in an off-line step, which utilized highly non-linear Finite Element (FE) modelling (that included friction and contact between parts and tools) coupled with an optimization routine. These pre-determined corrective actions are then implemented as an in-line control strategy based on incoming part measurements. While this method performs well, several draw-backs exist: 1) The dimensional fault patterns have to be known, which may not be realistic; 2) The control strategy cannot learn from the assembly system to identify actual fault patterns; and 3) The control strategy cannot adapt to changes in the system, such as; a change in suppliers, a redesigned material handling system, or system degradation, all of which can result in the appearance of new fault patterns.

Recently, Wells et al. [9] developed a self-healing framework based upon key insights drawn from the human immunological system. In our paper, the immunological principles outlined in Wells et al. [9] are applied to compliant
assembly systems to overcome the aforementioned issues encountered by Xie et al. [8].

The remainder of this paper is organized as follows. Section 2 provides a brief introduction to adaptive immunity and the mechanisms that make it successful. In Section 3 a mathematical adaption of these mechanisms based upon integrating current state-of-the-art data mining methods, physical models, assembly modelling techniques, sensor capabilities, and actuator networks for creating a self-correcting compliant assembly system is developed. Finally recommendations for future research and conclusions on the proposed system are given in Section 4.

2. Adaptive Immunity

There are several levels of defence built into the human immunological system. However, for brevity, only a brief introduction to adaptive immunity will be discussed in this paper as it is the driving force behind the proposed self-correcting assembly system. A more thorough discussion into the relationship between the immune system and assembly systems can be found in Wells et al. [9].

Adaptive immunity utilizes pre-existing knowledge of invading organisms (antigens) to develop responses tailored to maximally eliminate them through the simultaneous implementation of two immunological mechanisms; B and Th Cells. B Cells are the workhorse behind the active immune system as they provide two crucial functions; antigen recognition and eradication. B Cells are continually and pseudo-randomly generated, where each one is capable of dealing with a specific antigen, known as that B Cell’s cognate antigen. Once a B Cell encounters its cognate antigen, it binds to that antigen and awaits activation. Once activated, the B Cell will replicate into a collection of plasma and memory B Cells. Plasma Cells produce the antibodies to eliminate antigens, while memory B Cells are inactive versions of their parent B Cell (pre-activation). B Cells that have never been activated decay and make room for new B Cells. The process of B Cell decay and memory B Cell creation results in an ever-increasing population of B Cells that are able to fight off antigens known to exist within the host.

While B Cells fight the "war" against antigens, Th Cells orchestrate the "war" effort. Th Cells can be divided into two categories; effector and regulatory Th Cells. When an effector Th Cell contacts an antigen bounded to a B Cell, that B Cell is activated. This mechanism acts as a redundancy check to prevent B Cells from attacking good cells. Regulatory Th Cells maintain balance within the immune system by turning off immune responses when the antigen no longer poses a threat. In essence, Regulatory Th Cells are the antithesis of the Effector Th Cells as they limit unnecessary B Cell replication.

To increase the robustness of the immune response, two processes exist to produce a large diversity among antibodies, class-switching and mutation. Different classes of antibodies possess unique tools for antigen eradication. In order to take advantage of this collection of tools, during replication, B Cells are able to "switch" which class of antibody they produce. It should be noted that this class-switching is governed by effector Th Cells. Additionally, when antibodies are produced, mutations occur, resulting in slight variations on their effectiveness towards fighting the antigen.

Adaptive immunity provides a robust response to a wide variety of possible antigens. However, due to a lack of memory or a fierce disease, the immune response may be too slow to react before major damage is caused. In these cases, immunity to a given threat can be obtained artificially from an external source, such as vaccinations. A vaccination involves exposing individuals to altered or weakened antigens. This technique is mostly used as a prevention method, since vaccines aim to induce immunity without having to experience the actual disease. Once the individual has developed its own antibodies, the memory of how to identify and eradicate the infection is permanent, basically becoming an element of active immunity.

2.1. Adaptive Immunity Characteristics

The previous section highlighted the essential mechanisms that govern adaptive immunity. This section identifies the characteristics most vital to the success of the adaptive immune system when these mechanisms are implemented. In addition, the importance of these characteristics towards developing an ideal self-correcting assembly system is explored.

One of the most important characteristics of adaptive immunity is that it is continuously working to become more efficient and robust in detecting and eliminating threats regardless of the host's condition. In fact, for the majority of the time, this process operates without the host being aware that any problem exists. This is made possible by the fact that these two systems (immune and host) operate in parallel and interact only when needed. In an assembly system this behaviour is of the utmost importance as the cost associated with waiting for problems to be solved can be devastating. In addition, modern assembly systems are equipped with sensors and measurement equipment for statistical process control. This provides a wealth of information that could be used by a parallel immune system to further the assembly system's robustness.

The immune system has the ability to fight antigens that fall into two categories: 1) Known - Through the implementation of vaccines the system can defend itself against antigens that are known to exist and 2) Unknown - Through the generation of random B Cells the system can defend itself against antigens that are not known to exist. In addition, due to the nature of cell replication, memory cells, and cell decay, the immune system inherently maintains an optimal ratio of solutions for known and unknown antigens to minimize the use of system resources. It is possible but highly impractical to design/redesign an assembly system to completely account for known faults. Additionally, it is currently not possible/practical to account for unknown faults. Therefore, an effective adaptation of active immune system principles into an assembly system could drastically increase its ability to handle both known and unknown faults.

One of the most unique and effective characteristics of the immune systems relies in its ability to increase antibody diversity through mutations and class-switching. As previously mentioned it is possible to design/redesign an assembly system to defend itself against known faults; however, known faults rarely occur within an actual assembly system. This is due to the fact that faults are continuous. For example, through the use of programmable tooling an assembly system may be designed to perfectly account for a 2’ flange error on a piece of sheet metal in a joining operation. However, since this is a continuous fault, an exact 2’ flange error will never happen. With this in mind, it is of the utmost importance for an assembly system to be able to
handle a wide array of "mutated" faults. Additionally, complex assembly systems have multiple tools (programmable tooling, machine parameters, etc.) at their disposal to correct faults. Optimization tools coupled with physical system models can be used to determine corrective actions to minimize the effect of the fault. However, these models tend to be highly non-linear for complex assembly systems, which will not guarantee a global optimum corrective solution. Therefore, it is in the best interest for the assembly system to be able to perform "class-switching", which would allow multiple local optimal solutions (combination and usage of compensation tools) to be implemented.

3. Bio-Inspired Self-Correcting Compliant Assembly Systems

In this section, the previously discussed active immunity principles will be adapted into a mathematical framework for the proposed self-correcting compliant assembly system.

3.1. Antigens and Antibodies

In order to start building tools for adaptive immunity in compliant assemblies, one crucial question needs to be answered: What are antigens and antibodies in terms of an assembly system? In a general sense, antigens can be thought of as foreign organisms, which cause a possible threat to the host. In assembly systems, this equates to any incoming part with errors \( A_b \) being imparted into the assembly that may result in post-assembly Key Product Characteristics (KPC) deviations \( y \).

The term antigen is short for antibody generator, which means that the presence of an antigen should initiate the production of antibodies. As discussed in Section 2, in order for this antibody production to begin, the antibody must first be matched with its cognate B Cell, which is designed to eliminate that antigen. Therefore, a B Cell (in terms of an assembly system) is a pre-defined part with errors \( B \) that may or may not match an incoming part with errors \( A_b \). In addition, an assembly system antibody must be a unique corrective action \( A_b \) designed to compensate for \( A_b \), which is initiated if \( B = A_b \). Possible correction actions include adjusting; fixture locations, clamping forces, welding current or voltage, etc.

3.2. B Cell Creation

The proposed self-correcting compliant assembly system relies heavily on the ability to randomly produce pre-defined part errors and respective corrective action. In our approach \( A_b \) is determined through the use of virtual assembly models based upon Finite element (FE) analysis. It has been shown that FE models that incorporate friction and contact considerations between parts and tools can produce high fidelity estimates of post-assembly KPC’s deviations \( y’ \) as a function of \( B \) for compliant assemblies [10]. These models can then be used to determine an optimal corrective action \( A_b \) for a given \( B \) by minimizing the weighted sums of squares of \( y’ \) as

\[
J = \min_{A_b} y’^T Q y’
\]

s.t. \( h(y’, A_b) = 0 \) and \( g(y’, A_b) \leq 0 \).

Where \( Q \) is the weight matrix that captures the relative importance among KPCs and must be a positive definite matrix. Functions, \( h(\cdot) \) and \( g(\cdot) \) contain all the product/process constraints effecting the assembly and the capability to perform corrective actions, such as the actuator work space and the degrees of freedom limitations necessary to perform the adjustment. For compliant assemblies, solving Eq. 1 requires numerous evaluations of a highly non-linear virtual assembly model, which is computationally exhaustive. In a production environment, solving Eq. 1 on a part-by-part basis is not a viable option, as the production time would dramatically increase. Fortunately, this problem can be avoided by imparting adaptive immunity into the assembly system.

To rectify this problem, our approach will strictly use non-uniform rational B-Splines (NURBS) surfaces to develop part geometries. NURBS are widely used in computer aided design (CAD) for the representation of free-form curves and surfaces due to their interesting properties such as the ability to handle large surface patches, local controllability (local modification), lower computational requirements, and the ability to represent analytical forms [11,12]. A NURBS surface is defined as

\[
B(u, v) = \frac{\sum_{i=1}^{n_u} \sum_{j=1}^{n_v} B_{ui}(u) \cdot B_{vj}(v) \cdot w_{ij} \cdot \mathbf{p}_{ij}}{\sum_{i=1}^{n_u} \sum_{j=1}^{n_v} B_{ui}(u) \cdot B_{vj}(v) \cdot w_{ij}},
\]

where \( B \) is the surface, \( B(u, v) \) is a 1 x 3 vector defining a discrete point on the surface at location \( (u, v) \) in a Cartesian coordinate system; \( n_u \) and \( n_v \) are the number of control points in the \( u \) and \( v \) directions, respectively; \( B_{ui}(u) \) and \( B_{vj}(v) \) are B-Spline functions in the \( u \) and \( v \) directions, respectively; and \( w_{ij} \) are a 1 x 3 vector defining the coordinates of the control points and their weights, respectively. It should be noted that the B-Spline functions are defined by the order of the defining curves knot sequences. Please refer to Cox [13] and de Boor [14] for further details.

The first step towards generating B Cells requires the creation of NURBS surface representation of the nominal part geometry. First determine the order of the \( u \) and \( v \) directions, \( k_u \) and \( k_v \), respectively for the surface. This paper suggests to use \( k_u = k_v = 4 \), which will result in cubic curves. In addition, \( n_u \) and \( n_v \) need to be determined, based on the following two criteria: 1) the number of control points used should provide an exact representation of the nominal surface and 2) the number of control points used provides enough flexibility to allow a significant amount of randomization amongst the B Cell population, but not too many as that could result in wavy or crinkled surfaces. In order to satisfy these criteria, some visual experimentation may be required. Once \( k \) and \( n \) have been chosen (for both directions) nonperiodic knot vectors and the resulting B-Spline functions can be calculated.

For the next step, define \( B_{ij} = [b_{ij,1}, \ldots, b_{ij,n_v}]^T \) and \( w_{ij} = [w_{ij,1}, \ldots, w_{ij,n_v}] \). Given that the knot vectors and B-Spline functions have been determined and are now constant, the nominal part surface \( B(p) \) can be summarized from Eq. 2 as

\[
B = f(B_0^* + \delta).
\]

Also, define \( \delta \) as a \( 4 \cdot n_u \cdot n_v \times 1 \) random vector, which will be referred to as the B Cell generator. Now a B Cell can be defined as

\[
B = f(B_0^* + \delta).
\]

where a small perturbation of the B Cell generator about zero will result in the creation of a random surface based upon the nominal surface. The B Cell generator will be randomly generated sequentially at times \( t = 1, 2, \ldots, \) i.e. \( B_1, B_2, \ldots \). Also, for notation simplicity let \( B_0 = f(B_0^*) \) and \( B_0^* = B_0^* + \delta_i \).

In order to increase the robustness of the immune system our approach suggests that B Cells be generated from multiple random B Cell types. First define \( n_r \geq 1 \) as the number of random B Cell types to be implemented. Next for each random B Cell type’s B Cell Generator \( R\delta_r, l = 1, \ldots, n_r \) define its random
distribution. Also, set \( r a_1 = \cdots = r a_n = 1/n_r \), let \( r \alpha = [r a_1, \cdots, r a_n] \), and \( R \delta = [R \delta_1, \cdots, R \delta_n] \).

To produce a B Cell at time \( T \), randomly generate an independent multinomial observation defined by the event probabilities \( r \alpha \). Let \( RX_{i, T} \) be an observation classified in category \( i \) at time \( T \) where \( i = 1, \cdots, n_r \). For each \( RX_{i, T} = 1 \) generate the B Cell as

\[
B_T = f(B_T^* + R \delta).
\] (5)

For each \( B_T \) generated, a FE model of the assembly for this geometry is created resulting in the evaluation of Eq. 1 to calculate \( y(T) \) and \( A_S^*(T) \). However, in order to reduce the computational effort in evaluating Eq. 1, for the initial calculation of \( A_S^*(T) \) a subset of the actual constraints on the system should be used. Specifically, replace \( g(\cdot; \cdot) \) with \( g^*(\cdot; \cdot) \), where \( g^*(\cdot; \cdot) < g(\cdot; \cdot) \). The implication of this will be discussed later in the paper. To complete the B Cell creation process set \( T_r = 0 \), to indicate that the \( T^\text{th} \) B Cell has yet to be activated.

It is of the utmost importance to note that the random nature of B Cells could cause impossible surfaces, such as self-intersecting surfaces. However, the occurrence of such a surface is highly unlikely if the distribution of each B Cell generator is realistic. With this in mind \( B_T \) should be analyzed directly after creation to ensure that \( B_T(u, v) \neq B_T(x, z), \forall u \neq x \) and \( v \neq z \).

3.3. Vaccination and B Cell Mutations

During the initial design of an assembly system, it is often possible to identify known error types that may affect the system. Similar to the vaccination of children, it is vital that a new assembly system be vaccinated for these errors. The first step in implementing the vaccination phase in compliant assemblies is to identify \( n_k \geq 0 \) incoming known error types in the form of surfaces \( \{ KB_1, \cdots, KB_{n_k} \} \) for the assembly system. These error types can be determined through expert knowledge or a sensitivity analysis, which aids in determining the most critical error types [15]. Next determine the control points and weights vector \( \{ KB_1^* \} \) for each \( KB_i \) that satisfies Eq. 3, i.e. \( KB_i = f(KB_i^*), i = 1, \cdots, n_k \). In addition rank each known error type \( i \) by its estimated frequency or severity as \( k \alpha \), and let

\[
\begin{aligned}
\text{ka} &= [k\alpha_1, k\alpha_2, \cdots, k\alpha_{n_k}] = \|k\alpha_1, k\alpha_2, \cdots, k\alpha_{n_k}\|.
\end{aligned}
\]

As discussed earlier vaccines subject the host to altered or weakened antigens, which results in a similar immune response to actually having been exposed to the real antigens. In this respect, our vaccination should allow for mutations in the vaccines B Cells. Therefore, the next step in vaccination is to determine an appropriate distribution for the B Cell generator \( \{ KB_i \} \) for each \( KB_i \). This will allow for the B Cells to mutate and cover a wider range of possible antigens.

Define \( B \in [0; 1] \) if \( n_k > 0 \), 0 otherwise, as the proportion of B Cells to be generated from random B Cell types. For each time \( T \), randomly generate an independent Bernoulli observation \( BX_{i, T} \) defined by the event probability \( B \). For each \( BX_{i, T} = 1 \) generate \( B_T \) from a random B Cell type, outlined earlier. For each \( BX_{i, T} = 0 \) perform the following steps. Randomly generate an independent multinomial observation defined by the event probabilities \( k \alpha \). Let \( KX_{i, T} \) be an observation classified in category \( i \) at time \( T \) where \( i = 1, \cdots, n_k \). For each \( KX_{i, T} = 1 \) generate the B Cell as

\[
B_T = f(KB_i^* + K \delta_i).
\] (6)

3.4. Antigen Identification

An antigen must be identified and eradicated before it can cause harm to the system. With respect to an assembly system, this means that each \( A_S^* \) must be determined and matched with an appropriate \( B \) from the pool of B Cells before the assembly process begins so that the respective \( A_S^* \) can be applied. In order to obtain \( A_S^* \) this paper suggests the use of 3D laser scanners implemented in the assembly line. 3D laser scanners are one of the most state-of-the-art dimensional measurement technologies currently being implemented in industry. They can rapidly provide point clouds, consisting of millions of data points, to represent an entire manufactured part surface. Traditionally 3D laser scanners have been used in reverse engineering applications, but are beginning to be applied as inline measurement devices for statistical process control applications [16]. It should be noted that more traditional measurement devices, such as optical coordinate measuring machines (OCMM) can be used. However, this technology may not provide the resolution needed to accurately capture the surface.

In its raw form the data collected from 3D laser scanners is not appropriate for antigen identification for the following two reasons: 1) Finding a match between the current antigen based on the raw point cloud data and a large pool of B Cells would to be computational impractical and 2) The adaptive immune system must create a memory of the antigen’s unique characteristics, which would also prove to be highly computationally impractical as is would require the storage of an entire point cloud. To remedy this problem, this paper suggests that each \( A_S^* \) be defined as a NURBS surface that best represents the raw point cloud, which can be performed through multiple methods found throughout the literature. However, a majority of these methods are based on simultaneously estimating weights and control points through a least-squares strategy. Such methods require a significant amount of computational effort, which is undesirable in our case. Another approach, referred to as lofting or skinning, involves interpolating cross-sectional isoparametric curves to obtain the NURBS weights and control points, which tends to be more computationally efficient. Recently, Koch [17] has shown that both of the previously discussed methods produce the exact same results. Therefore, it is suggested that the lofting method be implemented in our approach. The lofting approach requires that the B-Spline functions are known, which is already the case as they were pre-defined by the nominal geometry in the creation of the B Cells. Therefore, each resulting NURBS surface \( A_S^* \) can be defined with a parameter vector \( \{ A_S^* \} \) consisting of weights and control points form Eq. 3, i.e. \( A_S^* = f(A_S^*) \).

It should be noted that this computational effort will be minimal when applied to compliant assemblies. Compliant assemblies typically involve sheet metal parts, which do not require a significantly large resolution to uniquely capture the surface. Moreover small imperfections in the surface will not affect the final assembly. It is recommended experimentation to obtain a scanning resolution to balance the requirement of computational speed and NURBS surface accuracy.

3.5. B Cell Memory and Mutation

Through the process of replication, memory B Cells retain knowledge of previous antigen encounters. When an antigen is
encountered by a memory B Cell it can replicate itself to become a plasma cell. During this replication, mutations occur, resulting in slight variation in the B Cell. These mutations cause an increased population of B Cells that are very similar to previously encountered antigens. Our approach requires that after each encounter with an antigen results in appending the memory matrix

\[ A = [A_g^1(1), A_g^2(2), \ldots, A_g^l(l)], \]

where \( l \) is the total number of antigens encountered. Once a significant amount of antigens have been encountered \( l > C_p \cdot 4 \cdot n_u \cdot n_v \), a principal component analysis (PCA) can be performed on \( A \). PCA is a dimension reduction technique, where a linear transformation is performed on a set of correlated multivariate variables, resulting in a set of orthogonal variables through eigenvector decomposition. The first step of PCA is to calculate the sample covariance matrix \( S \) from the memory matrix, followed by solving the eigenvalue decomposition:

\[ S = V L V^T, \]

where \( L \) is a diagonal matrix containing eigenvalues \( (\lambda_1 \geq \cdots \geq \lambda_{n_u \cdot n_v}) \geq 0 \) and \( V \) is an orthonormal matrix whose columns are eigenvectors. In addition, the quantity

\[ \lambda_i = \lambda_i (1 / \ lambda_1 + \lambda_2 + \cdots + \lambda_{n_u \cdot n_v}) \]

is the proportion of the variability of the original system that can be attributed to the \( i \)-th principal component. For more details regarding PCA, the reader is referred to Johnson [18].

Once the number of encountered antigens has reached a significant number \( l > C_p \cdot 4 \cdot n_u \cdot n_v \), it becomes possible to generate mutations of memory B Cells obtained from PCA. However, it is still desirable to obtain B Cells from the vaccination and random error types. Let \( W(\tau) = e^{-\gamma \tau} \), where \( \gamma \in [0; 1) \) is the learning rate for producing B Cells from PCA and time \( \tau \), where \( \tau = 0 \) when the value of \( l \) at time \( T_e \) becomes greater than \( C_p \cdot 4 \cdot n_u \cdot n_v \). For each time \( T_e \), randomly generate an independent Bernoulli observation \( (M X_F) \) defined by the event probability \( W(T - T_e) \). For each \( M X \), \( T = 1 \) generate a \( B_T \) from procedure outlined in Section 3.3. For each \( M X_F = 0 \) perform the following steps. Randomly generate an independent multinomial observation defined by the event probabilities \( A \), where \( A = [A_1, \ldots, A_{n_u \cdot n_v}] \). Let \( M X_{i,T} \) be an observation classified in category \( i \) at time \( T \) where \( i = 1, \ldots, 2 \cdot n_u \cdot n_v \). For each \( M X_{i,T} = 1 \) generate the B Cell as

\[ B_T = f(\mu + v_i(\theta_i)), \]

where \( \theta_i \sim N(0, \lambda_i) \), \( v_i \) is the \( i \)-th column of \( V \), and \( \mu \) the \( 6 \cdot n_u \cdot n_v \times 1 \) mean vector of \( A \).

Generating B Cells from memory in this manner is very appealing for several reasons. First, after the system matures the majority of the B Cells that are being generated are reflecting actual error patterns being seen in the system. Second, the system can adapt to changes or drift in the process. Third, the known error types identified during the vaccination stage could have been completely misguided. Therefore less emphasis should be placed on them in the long run. However, if these error types begin to show up they will still be accounted for through B Cells generated from the PCA. Finally, the majority of the principal component vectors will reflect statistical noise, which will allow for truly random B Cells to be created furthering the robustness of the system.

### 3.6. B Cell and Antigen Matching

In our proposed immune system each time an assembly enters a station it is transformed into a NURBS surfaces \( A_g \). For each \( A_g \) the immune system must be matched with an equivalent \( B \). The simplest approach for doing so would be to simply find the \( B \) that is most similar to \( A_g \). However, as the immune system matures the large number of candidate B Cells would make this approach highly undesirable, as this matching needs to be as fast as possible. To rectify this problem, our method will implement a decision tree to find this match. A decision tree is a supervised learning technique that creates a flowchart-like structure for the purpose of data classification. At every internal node within the tree the decision is made on which branch to follow by testing an attribute of the current data set. The final node (leaf or terminal node) contains a class label for the data set. In our case a B Cell’s attributes are the elements of its corresponding \( B^* \), while the class labels are based on whether a group of B Cells are highly similar. After the decision tree is built each \( A_g \) gets classified by the tree. Then its respective \( A_g \) gets compared to each \( B \) in that class instead of the entire B Cell pool.

Typically decision tree induction (learning) is performed with a training set of data; however, since the immune system is continually producing B Cells there is a stream of training data. In order to avoid continually redesigning the decision tree, our paper suggests the incremental method for the induction of the decision tree, namely the ITI algorithm described in Utgoff [19].

The procedure to generate class labels required for the creation of the decision tree is as follows:

Step 1: Define the criterion \( \epsilon_d \), which is the maximum difference allowable between two B Cells for them to be considered as the same class. Also define the criteria \( N_d \), which is the maximum number of B Cells that are allowed to be in the same class.

Step 2: At time \( T \) calculate

\[ \epsilon_{T-j} = \sum_{u} \sum_{v} (B(u,v) - B_{T-j}(u,v))^2, \]

where, \( j = 1, \ldots, T - 1 \).

Step 3: If any \( \epsilon_{T-j} < \epsilon_d \) then \( B_T \) gets labelled as the same class as \( B_{T-j} \). If this criterion is met more than once then \( B_T \) gets labelled as the same class as the \( B_{T-j} \) that corresponds to the smallest \( \epsilon_{T-j} \). If the criterion is never met or \( T = 1 \), \( B_T \) gets labelled as a new class. Also if the addition of \( B_T \) to a class results in the total number of entities for the class exceeding \( N_d \), \( B_T \) is not added and is instead killed (and set \( T = T - 1 \)).

When selecting the criteria \( \epsilon_d \) and \( N_d \) the following considerations should be made: 1) \( N_d \) will limit the size of the B Cell pool; 2) \( N_d \) will also limit the number candidate B Cells for a given \( A_g \) to be compared to; and 3) \( \epsilon_d \) will affect the overall size of the decision tree. It should be noted that since the identification of a class for \( B_T \) may result in its death, the classification should be performed before Eq. 1 is evaluated for \( B_T \).

As previously stated each \( A_g \) will be classified by the decision tree, based upon attributes \( A_g \). The difference between each \( B \) that has the same class label as \( A_g \) is calculated by Eq. 12. \( A_g \) is then matched with the \( B \) that produces the smallest \( \epsilon_m \).

\[ \epsilon_m = \sum_{u} \sum_{v} (A_g(u,v) - B(u,v))^2 \]

129
3.7. B Cell Activation and Decay

One of the main tasks of a T<sub>N</sub> Cell is to determine whether corrective action A<sub>N</sub>(t) should be implemented for the current antigen when B<sub>T</sub> = A<sub>N</sub> (should the t<sup>th</sup> B Cell be activated), or in our case when B<sub>T</sub> ≈ A<sub>N</sub>, where B<sub>T</sub> was the best match identified from the B Cell and antigen matching routine previously discussed. In order to make this decision, define the criterion ε<sub>a</sub> which will be the maximum allowable value of ε<sub>a</sub> that will result in A<sub>N</sub>(t) being implemented. This criterion performs two functions: 1) It prevents any errors from the classification process resulting in incorrect corrective actions and 2) It makes sure that B<sub>T</sub> = A<sub>N</sub> within a prescribed tolerance. If this criterion is met (ε<sub>a</sub> < ε<sub>d</sub>) A<sub>N</sub> is implemented and set I<sub>T</sub> = 1.

Through the process of B Cell activation T<sub>N</sub> Cells govern the process of B Cell decay and ultimate death. This is an essential requirement of the proposed approach, as the system resources and efficiency will significantly degrade if the B Cell population is not kept in check. Define N<sub>P</sub> as the maximum allowable B Cell population. At each time T, if T > N<sub>P</sub> remove B<sub>T</sub>, where ζ = min<sub>t</sub> (|I<sub>T</sub> = 0) and t = 1, ..., T from the decision tree. In addition set T = T − 1 followed by setting all B<sub>T</sub> = B<sub>T</sub>−1, where ζ < t < T.

3.8. Class-Switching

In addition to governing the activation and decay of B Cells, T<sub>N</sub> Cells also control class-switching. Performing class switching in a compliant assembly system requires the generation of new antibodies or corrective actions for an activated B Cell. If the criterion of ε<sub>a</sub> < ε<sub>d</sub> is met by B<sub>T</sub> and the current A<sub>N</sub> Eq. 1 is re-evaluated for B<sub>T</sub>. As previously discussed, the initial solution to Eq. 1 was implemented using a subset of the actual design space, g<sub>∗</sub>(·). A new antibody for B<sub>T</sub> can therefore be designed by choosing a different subset of the actual design space. To perform class-switching, define N<sub>C</sub> as the number of new antibodies to be generated for B<sub>T</sub>. When B<sub>T</sub> is activated, obtain N<sub>C</sub> new subsets {g<sub>1</sub>(·), ..., g<sub>N<sub>C</sub></sub>(·)} of the original design space. For each g<sub>i</sub>(·), i = 1, ..., N<sub>C</sub>, re-evaluate Eq. 1 to obtain candidate values A<sub>N</sub><sub>i</sub>(t) and y<sub>i</sub>(t). Finally set A<sub>N</sub><sub>i</sub>(t) = min<sub>1</sub> (A<sub>N</sub><sub>i</sub>(t)|y<sub>i</sub>(t) < y(t)).

The reasoning behind implementing class-switching in a self-correcting compliant assembly system is based of the fact that the solution to Eq. 1 is not guaranteed to produce a global optimal. It would be possible to use exhaustive search routines to try to obtain a global optimum; however, a relatively complex assembly system would make this extremely challenging. With this in mind, it would be highly inefficient to perform exhaustive searches when it is possible that some A<sub>N</sub><sub>i</sub> may never be applied. In addition exhaustive searches would slow the process B Cell production, resulting in a loss of overall immunity. The effect of class-switching is that the more often a specific antigen is encountered the stronger the antibody that fights it becomes.

4. Conclusions

This paper presents an approach for implementing adaptive immunity into compliant assembly system. The proposed approach emulates two key immunological tools, B and T<sub>N</sub> Cells by incorporating advanced data mining methods, physical models, assembly modelling techniques, sensor capabilities, and actuator networks, which results in a system capable of part-by-part error-compensation. This is owed to the fact that the computation burden of the immune system remains separate from the actual assembly system. This separation allows for immunity to be continually developed and refined over time without affecting the assembly system’s efficiency. While the mathematical framework for the proposed approach has been developed, the actual implementation will rely on two crucial aspects. Firstly, for the immune system itself to perform efficiency requires a sophisticated relational database to keep track of the B Cells and their current status within the system. Secondly, a highly robust finite element model (in terms of mesh generation, constraints, contact interactions, etc.) must be generated for the assembly system to account for the randomness allowed within the B Cell pool.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. CMMI-918055.

References

1. Introduction

To achieve high productivity and flexibility, assembly systems commonly need to use different Levels of Automation, i.e. both robots or machines and human operators are considered as resources. Tasks allocation between these resources is often planned in the design phase of the system due to machine investments, etc. However, if task allocation is done in early design phases, occurrence of unpredicted events leads to an inflexible system that cannot be dynamic and proactive during execution. For instance, if a robot breaks down, this robot cannot continue to assemble products and it must be repaired; this will have two consequences. The first consequence is that the tasks, or operations, allocated to the robot must be reallocated. The second consequence is that a repair task must be performed on this robot by another resource.

For consistency reasons, in the remainder of this paper, the term operation will refer to what the resources of the assembly system can perform or execute, and the term task will refer to high-level phases, e.g. design tasks.

For complex systems, the dynamic allocation of operations cannot be performed manually. Thus, in order to perform optimization on this allocation of operations, both operations and resources need to be precisely modeled. This paper is aiming to define these operation and resource models.

The proposed method for modeling resources will be illustrated in the paper through a simple example. The sequence of operations and the allocation of operations for this example are generated using the Sequence Planner software [3].

2. Sequences of Operations and sequence planning

An important task when designing an automated assembly system is to specify in what order the different operations can or must be executed. This task is called sequence planning and permits to define Sequences of Operations (SOP). In the following subsections a formal SOP language and a software tool that handles this language will be presented. Interested readers are referred to [4], [5] for further information.

2.1. Product example

The method presented in this paper will be illustrated through the example presented in Figure 1. This example is a Product composed of two pieces: Part A and Part B, assembled together with seven Rivets. This product is produced in a cell composed of three resources: a robot, a fixture and a human operator. These resources can perform different operations; their abilities can be redundant or not as will be detailed in the next sections. The following description briefly explains how the product is to be produced:

- First, Part A and Part B must be placed on the fixture. They can be placed either by the robot or the human operator.
- Then, Part A and Part B are fixated on the fixture by clamps controlled by the fixture itself.
- Then, Part A and Part B are assembled together with seven Rivets. This operation can also be performed either by the robot or the human operator.
- Finally, the product is inspected. Depending on which resource performs the operations to place Part A and Part B and the assembly operation, the inspection operation differs as will be detailed in section 4.3.

Figure 1. Product used to illustrate the proposed method
2.2. SOP language

The Sequences of Operations (SOP) language [4] is a graphical language used to specify and visualize relations among operations. This SOP language is based on an operation model. Sequences of operations are defined with the help of pre- and post-conditions related to each operation. Figure 2 presents how an operation can be represented using the SOP language.

Operations represented using the SOP language can be translated into Finite State Machines (FSM) [4], [5]. This FSM formalization can hence be used to apply verification and validation techniques and supervisory controller calculation.

The pre- and post-conditions can be related to several design or implementation phases. The following examples illustrate the use of pre- and post-conditions in different phases.

• Product design phase:
  Precedence relations can be expressed through pre-conditions, e.g. a hole needs to be drilled before riveting. Results from previous quality evaluation and tolerance analysis [6], [7] can also be introduced in order to prioritize a specific assembly sequence.

• Resource booking:
  If operation O1 must be performed by resource R1, then R1 must be booked before O1 is executed and unbooked after.

• Implementation control:
  To generate the control code for the implementation, information from the plant need to be taken into account. For instance, sensor values can be expressed through variables and permits to express conditions that must either be satisfied to start an operation (pre-condition) or to stop this operation (post-condition).

• Other purposes:
  Additional pre- and post-conditions can be generated automatically in order to solve issues related to safety, deadlock avoidance, collision avoidance [8], etc. This point will be explained in the next subsection.

Thus, instead of explicit and static sequence construction that is hard to handle for large scale systems, the SOP language permits to express relations through operations, hence, more flexibility is obtained during the different product and process design, and implementation phases, due to support by an underlying formal computational engine [3], [9] that resolves blocking and deadlocks and guarantees safety.

As a conclusion, this SOP language is based on the fact that a sequence of operations should not be considered as a compulsory input data but as a consequence of relevant pre- and post-conditions on when and how the operations can be executed.

Figure 3 and Figure 4 give an example of operations that are related through pre-conditions. Figure represents the set of six operations that must be performed on the product given in Figure 1 and illustrates pre- and post-conditions of operation Fixate A.

Figure 4 represents explicitly the relations between these six operations. According to the requirements on the product, operations Place A and Fixate A can be executed in parallel with operations Place B and Fixate B. Then, operations Assemble A+B and Inspect A+B must be executed sequentially.

Pre-conditions, and respectively post-conditions, of an operation can be composed of guards and actions. Those guards and actions are defined through variables. A guard is a condition that must be satisfied so that the operation can start (or finish). An action permits to define or change the value of variables when the operation starts (or finishes). For instance, a pre-condition related to the booking of a resource is both a guard and an action: the resource needs to be available (\(R1=\text{available}\)) and the resource is then booked (\(R1=\text{booked}\)).

The pre-condition associated to the resource booking is (\(R1=\text{available} \land \ R1=\text{booked}\)). The fact that both guards and actions can be used in the same condition helps engineers in expressing functional needs.

For the example given in Figure 4, two pre-conditions and one post-condition are defined for the operation Fixate A. The first pre-condition (Place A==finished) means that operation Place A must be finished before operation Fixate A can start. The second pre-condition and the post-condition are related to resource booking. This latter pre-condition means that resource Fixture must be available (Fixture==available) and then booked (Fixture==booked) before operation Fixate A can start. The post-condition means that resource Fixture is unbooked (Fixture==available) when operation Fixate A is finished.

In Sequence Planner, the modeling of pre- and post-conditions is based on the FSM model that handles the SOP language. Each operation and each resource is represented through a unique variable. This representation eases the definition of pre- and post-conditions related to operation states, resource operating modes and counters (see section 4.1).
2.3. Sequence Planner software

Sequence Planner (SP) is a prototype software tool developed to manage the SOP language and to perform sequence planning [3]. SP handles operations and permits to build Sequences of Operations according to pre- and post-conditions associated to each operation. These sequences of operations can be represented from different points of view. For example, SP can represent SOPs from a product point of view (sequences of operations related to one product) or from a resource point of view (sequences of operations performed by a specific resource).

To ease SOPs representation, several concepts have been introduced in order to express parallelism, alternatives, arbitrary order, etc. One important concept is hierarchy. A hierarchical relation can be used to represent in detail how an operation is performed. This hierarchical representation permits to simplify the representation of an SOP and only display information that are important for the end-user: either a high-level view of a whole system or a low-level view of a part of a system.

The input data of this software are the following:

1. A set of product operations. This set contains operations that should be performed on the product.
2. A set of resources with detailed operations. For each resource, each operation that it can realize is detailed through a hierarchical relation.
3. A resource mapping. This mapping permits to define, for each operation, which resources are able to realize it.

As previously mentioned, since the operations and the resources can be formally defined through a FSM model, supervisory control theory can be applied on the global model. SP is linked to Supremica [9] and permits to generate extra guards that must be added to pre- and post-conditions to avoid forbidden states. These extra guards are generated according to a supervisory model that is defined to ensure system liveness, safety, collision avoidance, etc. For instance, the sequence conditions define when the fixture should close clamps to fixate Part A or B, and together with interlocking and resource allocation this defines the pre-condition for the close clamp action. The interlocking checks that the resources do not collide with something or someone, while resource allocation is used to manage zone booking or tool allocation.

As a result, the output data of the software are several SOPs that contain the original operations completed with additional guards. These SOPs describe the minimally restrictive behavior of the system, i.e. the behavior with the largest amount of freedom that satisfy pre- and post-conditions and do not lead to blocking or deadlock. These SOPs can be generated from a product point of view (SOP given in ) or from a resource point of view (SOPs for the human operator, etc.), with different levels of detail, etc.

3. Levels of Automation in an assembly system

There is no simple way to make automation human-oriented and applicable across all domains and types of work. Different processes and domains put different emphasis on flexibility, speed, etc. requiring specific consideration of type appropriate automation [10]. Levels of Automation (LoA) could be defined as “The allocation of physical and cognitive tasks between humans and technology, described as a continuum ranging from totally manual to totally automatic” [2].

In order to provide indicators and parameters that can be used to perform and optimize resource allocation (seen from both a cognitive and physical perspective of LoAs) in assembly systems a 7 by 7 matrix has been developed [11], resulting in 49 different possible solutions of varying LoAs. Results from six case studies show that approximately 90 percent of the tasks performed in assembly systems are still performed by humans and by own experience (LoA=1.1) [12].

There is a need for a dynamic resource allocation that can take advantage of the access to instantaneous evaluation of the situations to choose the best allocation [13]. A case study that uses dynamic resource allocation, involving changeable LoAs [14], shows that it is possible to change from a human operator to a robot-cell and vice versa in order to achieve volume and route flexibility. The issue to be shown in this paper is how to model and simulate this dynamic allocation when alternative resources could be allocated to some operations.

Difference in LoAs implies that different resources need to be modeled as precisely as possible so that these models correspond to these LoAs and not to a generic resource. Furthermore, models of behavior, knowledge and skills for robots and human must be considered in different ways in order to better fit the real resources.

The aim of the proposed resource modeling is to reduce the gap between a resource and its model, and to take into account human roles in early design phases of an automated system to avoid automation abuse [15].

4. Modeling of human operator and robot resources

In previous works [4], [5], sequence planning has been considered only for the automatic mode of a system. In this context, the resource allocation is often the result of an optimization problem; thus only one resource (or several complementary resources) is allocated to each operation on the product. In this paper, we consider an assembly system in which robots and human coexist and can cooperate. If we consider the functioning of such an assembly system over a longer period, several unpredicted events can occur (robot breakdown, misassembled products, etc.). These errors imply that the optimal solution found for the automatic operating mode is no longer the optimal one if the system configuration is changed.

The following sections deal with two improvements that can be added to previous systems modeling. The first one considers operating modes of an assembly system; the second one alternative solutions with different LoAs.

4.1. Operating modes

In the automatic operating mode, a resource performs operations on a product. If several products are considered, the resource executes the same operation several times. This modeling means that products and resources are considered in two different ways: the resource performs operations whereas the product needs operations to be performed.

However, over a longer period, maintenance tasks need to be performed. For instance, after a breakdown, a robot needs to be maintained, reprogrammed, set-up, etc. In that case, the robot, which was previously considered as a resource, is now considered as a product, and a human operator performs operations on it.

The SOP language can be used to represent these operating modes. An operation is associated to each mode. Then, using a hierarchy relation, detailed operations are added to each mode according to the resource abilities. For the studied example, five modes can be considered for the robot resource: Production, Unavailable, Maintenance, Set-Up and Ramp-Up. These
operating modes are presented in Figure 5, and a detailed view of the Production operating mode is given.

![Image](image_url)

**Figure 5. Operating modes of the robot resource**

Since the SOP language is used to represent operating modes, relations between the different modes can also be defined through pre- and post-conditions.

For example, if we assume that the robot can assemble 100 products before maintenance is needed, the relations between the operating modes can be defined through the pre- and post-conditions defined as follows:

- post-condition of operating mode Production is defined by count_assemble==100.
- a part of the pre-condition of operating mode Unavailable is defined to be true when operation Production is finished. This pre-condition is defined by Production==finished.
- a counter count_assemble is incremented by 1 each time the operation R_Assemble is finished. This implies that a post-condition of operation R_Assemble is defined by count_assemble+=1.

Besides, the robot can be switched from Unavailable mode to Maintenance mode if a human operator who has abilities to perform maintenance operations contained in this mode is available. Then, the robot can be switched to Set-Up, Ramp-Up and Production, sequentially.

### 4.2. Use of resource operating modes

Resources used in flexible assembly systems typically have many abilities. Modeling a resource through several operating modes permits to organize these different abilities according to the operating mode they are related to.

As mentioned previously, a detailed optimal planning is hard to obtain for complex systems and is no longer relevant when an error occurs (resource unavailability, etc.). The hierarchical resource modeling through operating modes can be used to consider sequence planning from two hierarchical levels:

- A detailed optimal planning for each production mode: For example, an optimal sequence planning for the robot resource when it is in its Production mode. This planning takes into account detailed operations (R_PickUpA, R_PositionA, etc.)
- An organizational planning considering only the different operating modes. For example, the robot will produce 100 products being in its Production mode, and then a human operator will be needed to perform maintenance in the Maintenance mode. Once Maintenance is finished, the robot can switch to Set-Up, Ramp-Up, and then Production.

This hierarchical representation can also be used to define policies that must be applied to each mode. For instance, in the Set-Up or Ramp-Up modes, human operators and robots may use the same area; this means that the safety policy must be on its highest level [16]. On the opposite, in the Production mode, the time policy can be the leading policy. Moreover, according to the current operating mode the control code implemented in the robot can be slightly different. For example, in the Production mode, since the robot acts automatically extra guards can be added by the supervisory controller in order to take into account collision avoidance with other robots. On the other hand, in the Set-Up mode, the robot should execute step-by-step; extra guards can be added so that human validation (through a push button) is needed between two steps. Other policies such as deadlock avoidance, energy efficiency, etc. can also be applied to the different modes.

The same approach can be applied to a human operator resource. The operating modes can be different for each human operator since they all have different abilities. In that case, the objective of operating modes is to define what type of activities the human operator is performing: mental activities, manual activities, production tasks, maintenance tasks, pause, etc. These modes can then be used, for example, to conduct an optimization according to a human operator’s workload.

### 4.3. Levels of Automation in Sequences of Operations

Robots and human operators can sometimes have the same abilities, but in many cases their abilities and capabilities are different (in this paper capabilities are defined as detailed abilities, quality parameters, maximum load, etc.). If they do not have the same abilities, then the matching between resources and operations is simplified. If several resources can perform the same operation, the best one (according to an optimization criterion) is allocated to each operation. However, the result of an optimization problem depends on hypotheses. If these hypotheses are no longer satisfied the whole optimization problem must be reconsidered, and this takes time. In many cases, if a robot breaks down an easy solution is to use human operators as a replacement resource. Unfortunately, if the different alternatives have not been considered during the process design the reallocation of operations is not easy to perform. To tackle this issue, we suggest generating SOPs that keep track of alternatives according to different LoAs.

In what follows, only the Production operating mode will be considered. For the example illustrated in Figure, the operations are performed using three resources: a robot R, a fixture F and a human operator H.

The fixture F can perform operations Fixate A and Fixate B. The robot R can perform operation Place A, Place B, Assemble A+B and Inspect A+B (see Figure). The human operator H can perform operation Place A, Place B, Assemble A+B and Inspect A+B (see Figure).

![Image](image_url)

**Figure 8. Detailed abilities of the human operator resource**

As presented in Figure and Figure 8, even though the robot and the human operator have exactly the same global abilities their detailed abilities differ. Even though both can Place A, Place B and Assemble A+B, they do not perform these operations with the same precision. Thus, they have the same abilities but different
capabilities. The following statements illustrate these differences and their consequences:

1. If the human operator performs Place A or Place B, the geometry of the resulting assembly must be inspected. The robot is supposed to be accurate and repeatable enough so that geometry inspection is not required when both Place A and Place B are performed by the robot.

2. If the robot performs Assemble A+B, the rivets quality must be inspected. If Assemble A+B is performed by the human operator, we assume that rivets quality inspection has already been done by this operator during assembly.

3. Geometry inspection is performed by the robot. This inspection is done automatically using a contact sensor.

4. Rivets quality inspection is based on visual inspection. This inspection is performed by the human operator resource.

Statements 3 and 4 are used to define the robot ability \( R_{\text{InspectGeometry}} \) and the human operator ability \( H_{\text{InspectRivet}} \). Statements 1 and 2 are instantiated in the pre-conditions of operations \( R_{\text{InspectGeometry}} \) and \( H_{\text{InspectRivet}} \).

These four statements illustrate an example of an assembly system that requires two LoAs. Neither the human operator nor the robot can perform the whole assembly alone. However, several alternatives are possible, as illustrated in the next section.

5. Simulation and results

Sequence Planner has been used to generate and represent the different alternatives that permit to produce the product shown in Figure.

The operations that must be performed have been modeled as shown in Figure. Only the necessary conditions have been used to define pre- and post-conditions. The three resources previously presented have been modeled in the software with their detailed abilities, as shown in Figure and Figure 8, using hierarchical relations.

Figure 9 presents an SOP from a product point of view. Since both the robot and the human operator can perform operations Place A, Place B and Assemble A+B, alternatives are proposed for these operations. Alternative sequences are represented by a single horizontal line. Since Place A and Place B can be performed by different resources (e.g. Place A performed by the robot and Place B performed by the human operator), they can be executed in parallel. Parallel execution is represented by double horizontal lines. Dashed horizontal lines on top and bottom of operations \( H_{\text{InspectRivet}} \) and \( R_{\text{InspectGeometry}} \) mean that these operations are order independent, i.e. \( H_{\text{InspectRivet}} \) can be executed before or after \( R_{\text{InspectGeometry}} \).

However, this graphical representation is not sufficient to describe precisely and unambiguously the assembly system behavior. Pre- and post-conditions associated to each operation permit to define more precisely this behavior. For instance, according to Figure 9, \( \text{Fixate A} \) and \( \text{Fixate B} \) could be executed in parallel; however since they are performed by the same resource (fixture \( F \)) they cannot be executed in parallel. Pre-conditions of operations \( \text{Fixate A} \) and \( \text{Fixate B} \) permit to define this through resource booking conditions.

The SOP given in Figure 9 does not express an optimal sequence planning but the minimally restrictive behavior of the considered assembly system. This SOP permits to represent alternatives between a high LoA (operations performed by the robot) and a low LoA (operations performed by the human). If possible, alternatives are represented locally; these local alternatives permit to define local triggers to define if the human or the robot should perform operations that follow. The definition of these triggers is not unique. One solution is to consider that the first resource available performs the operations of the alternatives. Another solution is to define additional guards according to the solution of an optimization problem. Since many optimization problems can be defined (global time minimization, resource workload minimization, human workload adjustment, stock in process minimization, route flexibility maximization, etc.), generation of an optimal planning is not considered in this paper. However, the obtained SOP can be used as an input to an optimization problem, since costs can be attached to each operation.

Figure 9. Sequence of Operations with different possible alternatives
6. Conclusion and prospects

This paper has presented a way to consider both human operators and robots as specific resources with their own abilities and capabilities. The modeling language used to define and represent these resource models is flexible and permits to express many relations between operations and resources. The SOP automatically generated permits to represent all alternatives to assemble a product without deadlock.

On-going work considers three topics:
1. Automatic generation of control code for robots and instructions for human operators. The automatic generation of instructions for human operators would aim at reducing the learning time and improving product quality.
2. Sequence planning optimization with regard to flexibility and proactivity.
3. Definition of timed and stochastic models. These models would permit to implement criteria such as Mean Time To Failure, Mean Time Between Failure, etc. and could be used to define relations between operating modes.

References

The influence of assembly design methods on work exposures – an analytical examination

J. Egbers\textsuperscript{a}, G. Reinhart\textsuperscript{a}, A. Hees\textsuperscript{a}, W.P. Neumann\textsuperscript{b}

\textsuperscript{a} Institute for Machine Tools and Industrial Management (IwB), Application Center Augsburg, Technische Universität München, Augsburg, Germany

\textsuperscript{b} Human Factors Engineering Lab, Department of Mechanical and Industrial Engineering, Ryerson University, Toronto, Canada

Abstract: The increasing average worker age and high numbers of health disorders among the working population cause ability limitations. Mismatches of worker abilities and work demands inhibit successful worker employment and pose need for action towards the assembly designer. This paper presents an analytical examination to locate workload-influencing parameters at the assembly design process. The results described in this contribution give the opportunity to prospectively influence the integration of concerned workers when designing an assembly system. Furthermore, the importance of a design methods’ influence on workloads becomes quantifiable.

Keywords: Human Factors; Production Planning; Assembly Design

1. Introduction

1.1 Changing Worker Abilities

Due to the demographic change, worker abilities and therefore worker assignments to assembly workstations are changing. Ageing not only influences human performance factors on an individual level but also leads to an increasing interindividual variability of capabilities and competences [1,2]. The increasing average age of the working population causes a rising number of workers with ability limitations (AL). When physical capabilities decrease, for example the movability of the cervical spine or grasping forces, even workloads at a reasonable level may cause critical exposures to employed workers. As a result, the work ability can be negatively affected. Personnel placement problems in assembly operations emerge when workstations mismatch worker abilities [3] due to decreasing physical capabilities, higher workloads, or shortened cycle times when introducing a new assembly system generation. Corrective adaptations of existing workstations, worker transfers to other production divisions, and “convalescent” workplaces outside regular assembly lines will not facilitate the needed number of suitable workstations for future workforces [4]. Built to fit individual needs, only a small number of workers can be employed using these measures. High job performance requires successful matching of individual capabilities and work demands [5], but designing all workstations to fit a relatively small proportion of the workforce is inappropriate for economic reasons. Considering workers with ALs, this matching even is a prerequisite for sustained employability.

1.2 Integration of Workers with ALs into Assembly Design

The necessity to maintain older employees at work and to support continued health and productivity poses need for action [6]. Hence, future workstations need to be designed according to specific abilities of existing workers [3,7], especially to workers with ALs. Existing approaches mostly focus on healthy staff members. Especially, production engineering methodologies disregard physical differences in worker abilities and implications on worker employability although specific data on current worker abilities is available. Existing ALs and compensative measures must therefore be considered during the design process of assembly systems to provide a sufficient number and type of suitable workstations. Transparency regarding influences of assembly design methods on work exposures becomes crucial [8] to include these measures even before the assembly system is being implemented. This requires a holistic approach that comprehends both early conceptual stages that focus on the assembly line and the layout design as well as detailed design stages that include the design of workstations and assembly tasks. By identifying assembly design methods at which sufficient workload data on future workstations are known, an identification of mismatches between work exposures and worker abilities allows early revelation of adaptation needs [9] and a prospective integration of compensative measures for concerned workers becomes possible.

1.3 Aim of this Contribution

The following examination therefore integrates input parameters for workload equations into the assembly design process. The overall goal is to determine time points during the design process at which mismatches between worker abilities and work exposures can be assessed to purposefully integrate compensative measures.

2. Workload Assessment and Profile Comparison

To establish quality gates for ergonomic assessments during assembly system design, [10] identify workload-relevant design parameters and chronologically locate these parameters at the design process. Identified parameters include input information for the design project (such as product variants, productions quantities or lot sizes), are dependent on the product design (geometric shapes, product weights, related assembly processes), determined by the assembly design itself (workflow and workstation design) or can be related to the workforce (number of workers, worker assignment, job rotation schemes).

Workstation evaluation either bases upon commonly used methodologies for work requirement analysis, such as OWAS [11], NIOSH [12], RULA [13], EAWS [14], or on qualitative
estimations of workloads [15]. For application in industrial environments or specific companies, criteria selection bases upon present workloads. Worker abilities are analyzed using functional capacity evaluations (FCE) [16] and can afterwards be classified according to equated workloads. During assembly operation, profile comparison systems (PCS) combine both workload and ability information, identify suitable workstations for workers with ALs and initiate needs for adaptation. These systems assess and compare ability levels of individuals with physical work demands using different profile comparison criteria (PCC) to classify both worker and workstation using the same scales. Each PCC is defined by a classification standard for worker and workstation evaluation as well as algorithms to compare both databases. This may also include criteria weightings to balance the relative importance of PCC at a certain work environment. When applying work requirement analyses for workstation evaluation, each PCC consists of values for different input parameters (IP), and an algorithm or qualitatively formulated guidelines that determine levels of work exposure. For use in assembly, consistent and standardized PCS with formalized criteria definitions, consistent guidelines for ability as well as workload classification, and a criteria selection that addresses typical workload situations are necessary [17].

3. Methodology

Prior to the conceptual design, the first step when designing an assembly system should clarify possibilities for ergonomic interventions within the design process. The methodology determines the influence of assembly design methods on workloads by connecting PCC definitions to assembly design methods and comparing the outcomes of each design method with IP that are needed to examine PCC values at designed workstations. Contrary to existing approaches for integrating human factors into the assembly design process, this methodology does not require percentile data or qualitatively formulated guidelines that determine levels of work exposure. For use in assembly, consistent and standardized PCS with formalized criteria definitions, consistent guidelines for ability as well as workload classification, and a criteria selection that addresses typical workload situations are necessary [15].

3.1 Analysis of PCC Definitions and Design Processes

In the first step, each PCC is analyzed with regard to needed IPs for workload equations. Moreover, the first step includes the analysis of IPs regarding to units of measurement and a documentation of gained results. Common IP definitions include distance values, time durations, frequencies of process executions, angles, forces, postures, and physical quantities. Afterwards, the chosen assembly design procedure is analyzed and documented. This includes the chronological order of design processes, a compilation and documentation of applied design methods, design tools, corresponding outcomes, and types of data.

3.2 Identification of Connections

As a second step, the influence of each assembly design method on each IP is examined. First, a quantitative ascertainment of connections is conducted by comparing output parameters of each assembly design method with each IP. To ascertain indirect relationships between IP and assembly design methods, a second analysis considers qualitative relationships.

Fig. 1 shows the results of step 2 using the example of layout design and the three PCC ‘Working height in standing positions’, ‘Neck strain’ and ‘Load carrying’, each including needed IPs to equate workloads. Assembly layout design results in positioning planned workstations at the chosen assembly system location. Given the relations between product, process and resource (PPR-Model) as input data, the layout design also defines distances between and material flows among workstations. Consequently, the layout design directly influences IP 3.2 when loads need to be carried between workstations or from material supermarkets to workstations. As a result, covered distances can be calculated after designing the assembly layout. The assembly layout also qualitatively relates to IP 1.2, IP 2.3 and IP 3.3: Besides further influence factors, postures depend on the distance to be covered when grasping material, handling material or executing assembly tasks, all of which factors that are qualitatively (but not quantifiable) influenced by the assembly layout. Working height (IP 1.1) and deflection angles (IP 2.2) are an outcome of the detailed workstation planning and therefore not influenced by the assembly layout. The same also applies to IP 2.1: Process durations and durations of neck strains depend on task times and the workstation design. Load weights (IP 3.1) are determined by material weights, transport containers and the number of goods to be carried, all three design decisions that are taken before or after deciding the assembly layout. Hence, no connection between the layout design and these IPs are documented.

3.2 Determination of Relevant Design Points in Time

The third consecutive step analyzes connections between design phases and PCC. The main goal is to determine points in time of assembly design for each PCC at which IP values are fixed. These points in time can be used for worker assignments during assembly system design and for assessing the design
status towards its suitability for the current workforce. Since worker assignment based on unstable workload data may lead to inaccurate and volatile matching results, the goal is to determine the last design phase after which each IP remains unchanged. Identified relations are next transferred to an impact matrix that contains each direct and indirect connection between design methods and IPs. Evaluation of this matrix is not restricted to an identification of design points for worker assignment, but also allows a determination of the absolute and relative importance of design phases, a deduction of time windows during assembly system design that influence IP and PCC as well as time-dependent analyses regarding different IP data types.

4. Results

The following paragraph describes results, analyses and utilization possibilities when applying the methodology to a PCS used in automotive industry and to each of two reference design processes. The PCS includes 19 PCC that analyze workloads such as workings heights, physical strains, movability of different parts of the body, but also takt times, information provision, illumination of the work area or climatic conditions.

Beginning with a product analysis and ending with the implementation of the assembly system, reference design process 1 starts with a determination of design targets and consists of 41 design methods [18] (Fig. 2). The second reference design process was developed by the ADiFa consortium [19]: It consists of ten consecutive assembly system design phases and 21 design methods, starting with a product analysis and ending with the system implementation. In both reference design processes, a chronological order for design methods is defined. For each design method and each PCC, possible connections were analyzed and - if a design method influences an IP - connections were documented.

4.1 Points in Time for Profile Comparison

Fig. 2 shows the time spans during which assembly design methods influence PCC. Furthermore, derived design points in time for preliminary worker assignments by conducting profile comparisons (PC) are marked.

Reference design process 1 influences PCC in a nonuniform way, leading to four different points in time for profile comparison after completion of the design methods 21 (assembly layout design), 31 (workstation layout design), 34 (compilation of work instructions) and 35 (means of assembly design) [18]. Reference design process 2 influences PCC in a more uniform way, taking existing means of assembly into account at design step 7 and therefore influencing 18 out of 19 PCC during early design phases. At design step 18, the design quality of means of assembly is being assessed; this also includes the estimation of future workloads. Since nearly all design processes influence workloads either quantitatively or qualitatively, the only points in time for assessing the integratability of workers with ALs by comparing profiles are after design phase 18 and design phase 19 [19].

4.2 Influences of Design Processes on Workloads

Fig. 3 shows the cumulated number of identified connections between assembly design methods and PCC in a chronological order for both reference processes. Furthermore, the relative importance of each design method, based on the proportion of connections, is depicted.

Using reference process 1, the first design method with a relevant influence on workloads is the calculation of target processing times (10), a method that bases upon product analyses. The same applies to reference process 2 (5). The next influencing conceptual design methods split the required
assembly capacity into subsystems, determine operation sequences, determine basic organizational forms, and assign assembly processes to work stations. At reference process 1, design methods 18 to 23 focus on technical aspects of assembly design, like means of transport and suitable conveyors, workpiece carriers and material provision. Using reference process 2, the design method with the highest influence on workloads plans the material provision and logistic concepts (no. 14).

5. Discussion

Typically, early design methods analyze the product to be assembled and derive target process times. As a result, only time-dependent PCC can be assessed before deciding upon the assembly concept. The type of available PCC information changes from early to late design methods: in the beginning, information has a qualitative character since only single IP data is quantifiable and reasonable workload calculations are not yet possible. Correspondingly, PCC that examine workstation exposures are applied at later design stages.

Both reference processes do not include the product design itself - a design parameter which serves as a main source of workloads. Expectedly, PCC were influenced to a great extend by late design phases at which workstation layouts, operations sequences on a workstation level and technical means of assembly are designed. However, this examination does not support the often stated influence of early design decisions on workloads [20,21].

Using PCSS with a large number of PCC, an almost continuous assessment of the current design status becomes possible. In practice, each assessment leads to a collection of IP data, an equation and classification of PCC values and a worker assignment which implies additional effort for assembly designers. Focusing on the most important design stages is necessary to give a reasonable cost-benefit ratio when assessing the design quality. At least one worker assignment should be conducted before the assembly concept is approved.

6. Conclusions

The described approach allows the identification of design methods that influence work exposures. Several quantifiable assessments can be applied, including the influence of a design method on certain units of measurement of IPs, or the relative importance of a design method or design phase for certain workloads. When only limited design capacities are available, assembly designers can focus on the most important design stages according to the abilities of the present workforce.

Due to its analytical view on workload data during the design of assembly, the described approach also causes difficulties: even though input data for workload-related attributes may be calculable at determined points in assembly design, it remains questionable, if assembly designers put additional effort into a repeated workload equation. Reduced workloads in future assembly operations can be achieved using this methodology, but encouraging assembly designers becomes a management task.

References

Training by augmented reality in industrial environments: a case study

S. Iliano, V. Chimienti, G. Dini
Department of Mechanical, Nuclear and Production Engineering University of Pisa, Pisa, Italy

Abstract: Training activities play an important role in any industrial environment. Classical training methods, like on-the-job training, show many limits, considering also the wasting of productive time of skilled operators. Focusing on manual assembly lines, Augmented Reality looks very promising in representing a tool to assist operators: virtual instructions are provided real time via a display (e.g.: Head Mounted Display) and overlaid on the real environment. From a software point of view, however, AR implementation is quite difficult, time-consuming, and does not give any feedback concerning the action performed by the operator. Therefore, the aims of the paper are: i) to create a software routine able to guide programmers during AR implementation; ii) to combine AR and feedback sensors. The first aim has been accomplished by a “variant” approach, creating an interface for entering data characterizing each operation and connecting them in order to obtain a logical sequence. As a case study, the assembly of a simple electrical switch is used to show an application of the previous concepts.

Keywords: Assembly, Augmented Reality, Training.

1. Introduction
Teaching and learning have an important role in industrial environments. The technological evolution of production systems, the presence of sophisticated CNC systems, require the operator has basic computer knowledge to understand how to control the process. As shown in a study of the Ontario Skills Development Office (Canada), 63% of US manufacturing industries have implemented, over the past years, new technologies requiring, for their use, training of personnel. In addition, a third of respondents wants to improve their job-performance and rewards best employees [1].

Looking at the production department of any manufacturing company, typical training methods, like on-the-job training, show many limits, considering also the waste of productive time of skilled operators involved in these activities.

In particular, focusing on manual assembly lines, training difficulties are related to: high number of parts to be assembled, wide variety of operations to accomplish, dexterous assembly tasks, assembly errors, risk of not completing assembly operations and therefore introducing incomplete products in the assembly line.

The difficulty is also in capturing trainee attention, keeping him focused on the whole phase and making him aware of right assembly priority, even during repetitive procedures. His activity should be performed in adequate time, ensuring product quality and minimizing errors (e.g.: using poka-yoke devices).

Table 1 sums up the main advantages/disadvantages of classical training methods, in comparison with a new approach using Augmented Reality (AR) as an instrument for giving to the operator the needed information.

2. Training by Augmented Reality

AR looks very promising as a tool to assist and support operators during training in any industrial environment [2, 3]. Virtual instructions, contextualized and provided real time via a display (such as Head Mounted Display - HMD), can facilitate this process, replacing paper instructions.

This innovative technique, in fact, allows the user immersion in a world made of two video streams: 1) the real environment in front of him, shot by a webcam set on his forehead; 2) virtual contents, overlaid on the reality and aligned with it.

Table 1. Training methods: advantages and disadvantages

<table>
<thead>
<tr>
<th>Training Method</th>
<th>PRO</th>
<th>CONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-the-job</td>
<td>- Performed directly on the workplace</td>
<td>- It could take a long training time</td>
</tr>
<tr>
<td></td>
<td>- Both general and specific skills transferred</td>
<td>- Waste time of skilled operators</td>
</tr>
<tr>
<td></td>
<td>- Most effective method of training in any area</td>
<td>- High costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Unstructured approach</td>
</tr>
<tr>
<td>Computer-based</td>
<td>- Distance learning</td>
<td>- Computers set outside work environment</td>
</tr>
<tr>
<td></td>
<td>- Specific computer programs</td>
<td>- Limited effectiveness</td>
</tr>
<tr>
<td></td>
<td>- Modularity</td>
<td>- Unsuitable for training on manual operation</td>
</tr>
<tr>
<td>Augmented Reality</td>
<td>- Suitable for programs and software training</td>
<td>- Lack of feedback</td>
</tr>
<tr>
<td></td>
<td>- Performed directly on the workplace</td>
<td>- Difficult implementation</td>
</tr>
<tr>
<td></td>
<td>- Operator guided step-by-step in the operations to perform</td>
<td>- High cost of hardware and software</td>
</tr>
<tr>
<td></td>
<td>- Standardization of teaching</td>
<td>- Research in training applications still in early stage</td>
</tr>
<tr>
<td></td>
<td>- Conveyance of less human resources</td>
<td></td>
</tr>
</tbody>
</table>

Virtual contents, usually represented by interactive training material (text instructions, animations, pictures, 3D models), are referred to the real environment by means of a tracking system. One of the most common tracking methods is the marker-based, which uses binary references set on the workbench to align the virtual contents to the real objects.

Some interesting examples of AR applied to support manual assembly can be found in literature [4, 5, 6, 7, 8, 9]. Tom Caudell, the former who coined the word “Augmented Reality”, as first applications of this technology, showed an example of wire bundle assembly and connector assembly [10]. The reference [11] reports an interesting example of an AR system for helping pump assembly process at one of the leading pump producers.

From the literature analysis it results that the chance to use AR for training operators in assembly tasks has already been
examined, and showed good prospects of application. Nevertheless, from a programming point of view, the implementation of this procedure is quite difficult and time-consuming. Moreover, the mere use of AR tools and methods does not allow a closed-loop control of the assembly task performed by the operator, since no feedback signal is usually present in these kinds of applications.

Taking into account the previous considerations, the authors believe that a new approach in this field could be obtained by: i) creating a software routine able to guide the programmer in implementing AR-based procedures for training purposes in manual assembly; ii) developing AR-based facilities in manual assembly lines, integrated with feedback sensors, able to give appropriate signal to the AR system.

3. Variant approach to AR implementation

A variant approach is proposed to facilitate and speed up the implementation of the AR procedure.

The principle of this approach consists in creating a "standard AR procedure" for each "standard elementary assembly operation" to be performed during assembly (e.g.: take a component from..., use tool to..., etc.). The specific AR procedure is obtained by retrieving from a database the standard AR procedure and by varying it according to the specific product.

The method is depicted in Figure 1. As a first step, the assembly sequence is divided in tasks and, each of them, in standard elementary operations characterized by:

a) name of the objects associated to the elementary operation;

b) virtual elements used in AR procedure for guiding the operator (text, arrows, 3D animations, pictures, etc.);

c) positions of virtual elements (coordinates and orientation) with respect to the markers placed in the working environment.

Considering the i-th standard elementary operation, the standard AR procedure is extracted from the database and varied depending on the previous parameters. The standard AR procedure is then added to the previous ones in order to create a logical sequence and compose the AR training for the whole assembly sequence.

4. Integration of Augmented Reality and feedback sensors

Figure 2 shows the structure of proposed method for mixing AR devices and feedback sensors.

Figure 2. Proposed method for mixing AR and feedback sensors.

Considering a generic i-th task, the AR system produces virtual aids which are sent to the AR device worn by the operator (e.g.: HMD). The operator is therefore able to see the working environment together with the instructions on what to do and, consequently, acts guided by these virtual elements.

Infra-red sensors are distributed in the workplace, in order to give feedback signals on the correctness of the actions done by the worker. The aim of these sensors is to detect the operator’s movements, verifying that he acts according to the assembly sequence. Two different feedback paths are considered: the former directly to the working environment, in order to alert the worker with proper devices placed on the workbench (alarms, LEDs, etc.); the latter to the AR software running in the computer, in order to create other virtual elements for supporting the operator in performing recovery actions: a virtual instruction on what to do appears on screen if the sensor detects the worker is acting improperly.

5. Case study

As a case study, the assembly of an electrical switch has been considered. This product consists of 15 components and it is illustrated in Figure 3.

5.1. Assembly sequence

The assembly sequence is structured as follows: positioning of switch case (#1) in a proper slot on the assembly fixture; assembly of the terminal blocks (#2) and spacers (#3) inside the switch case; insertion of the spring (#4) on the shaft inside the case, followed by the assembly of the switch (#5) inside the spring; insertion of the actuator (#6) inside switch hole, aligned with the spring; positioning of the switch cover (#7) on the switch case.
At this point, the assembled parts are rotated and placed upside down in another slot of the fixture. The wire connectors (#8) are therefore tightened to terminal blocks and the correct functioning of the switch is tested. At the end, the cover plate (#9) is placed on a third slot of the fixture, the previous subassembly is forced inside the plate, and the top cover (#10) is used to hold all the components.

5.2. Implementation of AR procedure

In order to apply the variant approach previously described, the assembly sequence is divided in task and in elementary operations. As an example of application, the task “connector wires (#8) tightening to terminal blocks (#2)” is considered. This task is formed by 14 elementary operations. The first one is “take connector wire (#8) from container” and the corresponding standard AR procedure is named “take an object from...” and it is extracted from the database. The variation is obtained by setting the following parameters:

a) name of the object: “connector wire”;
b) virtual element: “arrow indicating the container”;
c) position of virtual element: values taken from 3D full-scale reconstruction of the whole assembly station, using CAD software.

5.3. AR facilities

As shown in Figure 4, an HMD using a Video See-Through (VST) device has been worn by the worker. This device includes a pair of goggles with one mini-display for eye, and two webcams, aligned with the eyes in order to have a better perception of the real environment.

A Graphical User Interface (Figure 5) is seen by the worker through the HMD. This interface includes: assembly operation number, assembly instructions, picture of the part to be manipulated. Using a wireless remote controller, the worker is able to navigate through the training session.

5.4. Sensorized AR station

The assembly station (Figure 5 and 6) is formed by an assembly fixture, 10 containers and 6 tool-holders. As far as the tracking system is considered, it was decided to place 4 markers at the corners of the fixture in order to make the system more robust, by implementing an algorithm that displays virtual contents even though only one marker is visible from the camera. Additional markers were placed on each group of containers.

Each container is equipped with IR sensors, consisting of a 38 kHz infrared emitting diode and a receiver module, placed in the middle top of each container in order to detect hands taking components from container. The output of each sensor is transmitted both to the working environment and to the AR system.

In the working environment, the sensors activate two LEDs positioned on each container: a red LED for error signals to the worker; a green LED for indicating the correct action to be performed. Sensors and LEDs are connected to a hardware platform based on the microcontroller ATmega1280, characterized by 54 digital input/output pins and a USB connection. The platform analyzes signals coming from IR sensors and sends digital outputs towards LEDs and interfaces with AR software running on the computer.

An AR tool for managing virtual contents was created using C# programming language. With this tool, the user is able to receive the instructions about the assembly task. Moreover, this program is in charge of communicating to the platform the step
number performed, so that the right LED can light. At the same time, an iterative check is carried out every 10 ms on sensor outputs. Sensors reveal the presence of the worker hand in a certain container, and the output is sent to the AR system, which re-configures itself depending on the right/wrong sensor triggered.

When an elementary operation is performed (e.g.: “take connector wire (#8) from container”), the computer, besides showing virtual contents on the HMD, communicates with the microcontroller that sets on the green LED on the correct container; the microcontroller waits for the signal coming from IR sensors and the procedure is paused. This means that the worker keeps seeing virtual elements (text, arrows etc.), but he cannot go forward until the sensor is activated. At this point three cases may occur:

- if the worker triggers the right sensor (i.e. he takes the right component), the microcontroller sends a confirmation signal to the computer and the worker can move to the next step;
- if the worker triggers the wrong sensor (i.e. takes components from the wrong container), the red LED is set to on and the AR procedure is stopped until the right sensor is activated;
- if the worker does not trigger any sensor (i.e. he does not take any component) but uses controller and tries to change the step, an alert message appears on the HMD.

Three cases may occur:

5.5. Training session

Several tests have been carried out, to check the correctness of training and to evaluate its feasibility. In Figure 7, some steps of a training session have been represented. The task is “connector wires (#8) tightening to terminal blocks (#2)”. As already mentioned, this task begins with “take connector wire (#8) from container”.

In Fig 7a, it has been supposed that the operator takes a component from the wrong container: a warning message appears on the screen, the red LED is on, and the virtual arrow indicates the right container. In Fig 7b, the operator takes the connector wire (the green LED is on); in Fig 7c, 3D models of the switch and the wire in the middle of the GUI, overlaid on the real environment, showing how to assemble them. Fig. 7d gives instructions about the tool to take for tightening the wire.

6. Conclusions

The main innovative aspects introduced by this work concern:

- the development of a general tool to be used by workers to learn how to execute an assembly sequence;
- the creation of a software routine, based on a variant approach, able to guide the programmer in implementing AR-based training session;
- the implementation of an AR based system for supporting operators in performing assembly tasks, integrating AR devices and feedback sensors.

Acknowledgments

This work has been funded by the European FP7 Research Project FLEXA - Advanced Flexible Automation Cell (Grant agreement no 213734).

References

Interaction between complexity, quality and cognitive automation

T. Fässberg, Å. Fastha, F. Hellmanb, A. Davidssonc, J. Stahrea

aDepartment of Product and Production Development, Chalmers University of Technology, Gothenburg, Sweden
bVolvo Trucks, Gothenburg, Sweden
cVolvo Cars Corporation, Gothenburg, Sweden

Abstract: Today’s assembly systems are adapting to the increased mass customisation. This means shorter cycle times, more variants and a more complex environment for the operators. An industrial case study has been performed in order to describe relations between complexity, quality and cognitive automation. This article use quantitative methods to describe the complex environment. This is done in order to create a better understanding for the importance of using cognitive automation as an enabler in order to create a more competitive assembly system for the future.

Keywords: Assembly, Complexity, Quality, Cognitive automation, LoA

1. Introduction

The future holds a more customized market. Complexity related to the increased number of product variants induced by mass customization has huge effects on the final mixed model assembly lines in modern factories. This leads to more unique products and a more complex work environment for the operator who will assemble the products, case studies show that 90 % of final assembly tasks are still performed by humans [1]. One definition of complexity is by Weaver [2] whom defines complexity as the degree of difficulty to predict the system properties, given the properties of the systems parts. Schleich means that a driver for assembly system complexity is the high variety of products and parts [3]. Similar ideas can be found by Urbanic et al. which, presents a model of complexity were quantity, diversity and content of information is direct associated with complexity [4]. The focus in this paper is the complexity related to mass customization i.e. caused by an increase of number of products and parts to assemble (increased amount of information). To meet requirements from mass customization, many assembly systems are using a mixed-model assembly approach as an enabler for the high variety of products. Although mixed model assembly is an enabler for high variety, such systems tend to get very complex as variety increase [5]. An important aspect of complexity is the “perceived complexity”. From an operator point of view this is a subjective factor such as competence and information [6]. Cognitive help tools are seen to reduce the perceived complexity by supporting competence and information.

The increased task complexity in assembly needs to be handled otherwise the quality of the product and productivity in the system could be affected. In order to maintain high quality and reduce the complexity, one solution could be to consider cognitive automation for the operator e.g. technical support to know how and what to assemble and to be in situation control. An industrial case study has been executed in order to investigate the effects cognitive automation have on quality, in terms of assembly errors, in a complex final assembly context.

The aim of this paper is to: Investigate if cognitive automation can be used to increase quality in a complex final assembly context.

An industrial case study has been executed to test if there is a relation between cognitive automation, quality and quantitative (objective) station complexity.

2. Case company

Volvo Car Corporation manufacturers around 400 000 cars per year. The two main assembly plants are located in Gent, Belgium and in Torslanda, Sweden. In the Torslanda plant five models; V70, XC70, S80, XC90 and V60 are produced with a total volume of 136 323 cars for the year 2010. The five models are based on three different platforms.

One serial flow mixed-model assembly line is used for all five models in the final assembly plant. The assembly line is divided into different line segments, which can have buffers in between. A driven conveyor line continuously paces a majority of the assembly line segments. The assembly is characterized by short cycles at each station and a high division of work. The current tact time (September, 2011) is about 66 seconds but can vary between the different line segments. To some extent subassembly lines are also used at different parts of the line. At the sub-assemblies other tact times may be used.

2.1 Selected area

In order test the aim of the paper, an area of interest was selected. The area is one of the most complex in the final assembly with a very high product variety and a large number of parts. The chosen area consists of a total number of sixteen stations were seven have been studied within this project (the grey operators in figure 1 represents the chosen stations). The chosen stations are a part of the pre-assembly area for the preparation line of engines. In the line the engines are customised with correct driveshaft, cables etc. The engines assembled are used in all models and variants on the main assembly line. There are three areas for the pre-assembly of the engines and this is the second area, Power Pack 2 (PP2).

Figure 1. Selected area, total number of stations and selected stations
The layout of the pre-assembly area is organized as a serial flow assembly line without buffers between the stations. A driven assembly line conveyor paces the line. The assembly line is characterized by short tact times, currently 63.2 seconds, and a high number of different product variants and a large number of different parts. There is one operator working at each station. Both sides of the line are used resulting in that two stations can use the same line range but on different side of the assembly object. Some stations utilize both side of the line, which for instance can be due to large size components.

The work organization is designed so that one team is responsible for 6-8 stations. There is one team leader within each team. The operators rotate between the stations in the team. For the stations chosen in this study a total number of two teams are involved as seen in figure 1. All of the investigated stations are considered to be complex due to the large number of variants and parts.

3. Quantitative methods used

Three different measurements have been used to verify the hypothesis of this paper namely; operator choice complexity, assembly errors (quality) and cognitive automation. How these measurements have been gathered is explained in the following sections. Data have been gathered in the final assembly plant at Volvo Cars in Torslanda, Sweden during the summer and autumn of 2011.

3.1 Operator Choice Complexity

The complexity in mixed-model assembly caused by the high levels of variety is by Hu et al. called “Operator Choice Complexity” (OCC), which concerns all choices that the assembly operator can make and the risk for error associated with these choices [5]. The measurement of complexity at each station that is used for comparison in this paper is the operator choice complexity proposed by Hu et al. [5] and Zhu et al. [7]. The model can be used to calculate a complexity measure for mixed-model assembly lines.

The complexity model is based on entropy function. A definition of the operator choice complexity that is induced by product variety is given as follows: Complexity is the average uncertainty in a random process i of handling product variety, which can be described by entropy function $H_i$ in the following form:

$$H_i(P_{i1}, P_{i2}, ..., P_{iM_i}) = -\sum_{j=1}^{M_i} P_{ij} \log P_{ij}$$

(Eq 1)

where $P_{ij}$ is the occurrence probability of a state $j$ in the random process $i$, $j \in [1 M_i]$, $C$ is a constant depending on the base of the logarithm function chosen. If $\log_2$ is selected, $C = 1$ and the unit of complexity is bit [5].

Equation 1 is used to calculate the operator choice complexity for each of the seven selected stations. Input to the equation is the number of variants that occurs at each station and the demand for each variant based on 3835 cars produced during one week. The probability $P$ is calculated for each variant $j$ and for each station $i$ and the total operator choice complexity is calculated with the entropy equation 1. The result for each station $i$ is present in figure 2. The unit scale in the figure is bit.

Other complexity parameters such as number of tools, parts and tasks to perform have been gathered, seen in table 1. These parameters are used as a complement to the OCC measure. The parameters have either been collected through direct observations at the assembly line or the balancing and sequencing system used at Volvo Car Corporation.

Table 2. Complexity parameters related to each station

<table>
<thead>
<tr>
<th>Station number</th>
<th>30</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>13</th>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Parts*</td>
<td>15</td>
<td>23</td>
<td>20</td>
<td>14</td>
<td>18</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>Number of Tools*</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Number of Tasks**</td>
<td>14</td>
<td>22</td>
<td>26</td>
<td>17</td>
<td>15</td>
<td>25</td>
<td>17</td>
</tr>
</tbody>
</table>

* Sequenced parts seen as one
** Mean value of two products

3.2 Assembly errors

Assembly errors are discovered at control stations or directly by the operators at the assembly stations. All errors are reported by team leaders or by responsible quality personnel. The errors are connected to the product architecture. This means that even if a problem is discovered downstream from where it actually occurred it can be traced back to station, responsible team and individual operator causing the error.

Errors reported to the internal quality system have been extracted for a time period of 16 weeks from the system and sorted by station. The results are presented in table 2. The errors have been cleared from errors caused by material and parts defects i.e. only assembly errors are included. The errors found had the following characteristics:

Table 3. Type and number of errors

<table>
<thead>
<tr>
<th>Error type</th>
<th>Number of errors</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Connected</td>
<td>106</td>
<td>30%</td>
</tr>
<tr>
<td>Incorrectly fitted</td>
<td>83</td>
<td>24%</td>
</tr>
<tr>
<td>Missing</td>
<td>51</td>
<td>14%</td>
</tr>
<tr>
<td>Not tightened</td>
<td>38</td>
<td>11%</td>
</tr>
<tr>
<td>Total</td>
<td>278</td>
<td>79%</td>
</tr>
</tbody>
</table>

Assembly errors are categorized in eleven categories, were the top four categories accounts for (278) 79% of the total number (353) of errors. These categories are associated with errors were parts not have been connected properly, incorrectly
assembled or that the parts are missing or not tightened correctly. Other quite common errors are that parts are loose or that e.g. plastic covers have not been dismantled.

3.3 Cognitive Automation

In order to measure the cognitive level of the station a components of a method called DYNAMO++ was used. The DYNAMO++ method [8] and a concept model [9] for task allocation were developed during 2007-2009. The main aim is to evaluate and analyse changes in an assembly system due to triggers for change i.e. the company’s internal or external demands and Levels of Automation (LoA). The LoA analysis is done at working place level [10] i.e. on task, in stations [11, 12] and from an operator’s perspective.

The measurement parameters used for task allocation is a seven by seven matrix [8], seen in figure 3, further developed from Frohm’s taxonomy [13].

The LoA measure was made from direct observations and from standardised assembly instructions. An advantage of the use of two sources of information is that the standardised assembly instruction does not always correspond with the reality, which we wish to capture. Two models were assessed for each station, the most common model (C) regarding demand and the heaviest model (H) to produce regarding time. The distributions of the tasks for the two models are presented in the matrix illustrated in figure 4.

Results show that 62 percent (H) and 64 percent (C) were made with LoA level= (1;1) i.e. by hand and with own experience. The fact that so many tasks are done without cognitive support could have an impact on quality. Further, 25 percent (H) and 24 percent (C) is done with LoA cog =5 (often Pick-By-Light or indicators of what bit to use for the pneumatic screwdrivers). These are examples of tools, which are used to guide the operator to make a correct action and avoid errors. Examples of the different cognitive support tools (levels of cognitive automation) used at the stations

The parts are presented to the operators in the material facade in bulk packages or by sequence racks (which could be seen as cognitive automation because they are sorted i.e. LoA cog=2=working order). Many different sizes of the bulk packages are used in the facade. Poka-yoke solutions such as Pick-By-Light systems, illustrated in figure 5, are used for some parts but not all, see table 3.

Operators are supported in their work by screens, which show current model and variant and status of required tightening operations. The operators are also provided with feedback from the Pick-By-Lights and haptic feedback from some of the tools used. Operator instruction sheets are available at every team area gathered in binders. Both manual tools and automated tools are used in the assembly work.

<table>
<thead>
<tr>
<th>Station</th>
<th>30</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>13</th>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Tools</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Number of PBL</td>
<td>7</td>
<td>15</td>
<td>13</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Sequenced articles</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
4. Relations between the three areas

In order to answer the hypotheses an investigation between four relations (illustrated in figure 6) has been done and is discussed in following sections.

![Figure 6. Overview of the three investigated areas](image)

4.1 Relation 1; between operator choice complexity and assembly errors

The first relation between the operator choice complexity and the assembly errors is illustrated in figure 7. As seen there is a relation between the lowest complexity and the lowest number of assembly errors (Station 13) and vice versa (Station 23). Station 11 differs the most from the pattern that the assembly errors follows the measure of OCC. Therefore these three stations have been further compared in the other relations.

![Figure 7. Relation between operator choice complexity and assembly errors](image)

4.2 Relation 2; between assembly errors and cognitive (and physical) automation

The station that sticks out is station 11, if assembly errors had a correlation with OCC the anticipated number errors found would have been approximately the half, why is it so high? Due to the fact that over 60 percent of the tasks are done with own experience and that “incorrectly fitted” and “not connected” has the highest assembly errors in the further investigated stations (11, 13 and 23), a summary of the assembly errors is shown in table 4, could be an indicator that there is a need for more cognitive support within these stations.

Station 11

A total of 241 assembly errors were found during the investigated time period. 181 assembly errors were excluded due to that these errors were associated with errors from a supplier and not the assembly operation. Leaving the total number of errors for the investigated time period to 60 errors. 63% (38) of the errors were classified as “not connected” and one single part and task accounted for 38 % (23) of the total errors. The LoA of this specific task was (1, 1)\(^1\). Meaning that the operation was performed without any support.

Station 23

A total of 91 assembly errors were found during the investigated time period. 60% (54) of the errors were classified as “incorrectly fitted”. One single part and task accounted for 51 % (47) of the total errors. The part was either placed in wrong position or missed. The LoA of this specific task was (1,1). Meaning that the operation was performed without any support.

Station 13

A total of 10 assembly errors were found during the investigated time period. They were all classified as “not connected”. The low error rate at this station could be explained by that most operations at the station were associated with a high LoA. Part assurance was made with a hand scanner and tightening operations were counted by the system to match the number of tasks supposed to be performed.

<table>
<thead>
<tr>
<th>Error Code</th>
<th>Nr. of errors (station 11)</th>
<th>Nr. of errors (station 13)</th>
<th>Nr. of errors (station 23)</th>
<th>Total nr of errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrectly fitted</td>
<td>4</td>
<td>-</td>
<td>54</td>
<td>58</td>
</tr>
<tr>
<td>Not Connected</td>
<td>38</td>
<td>10</td>
<td>5</td>
<td>53</td>
</tr>
<tr>
<td>Not Tightened</td>
<td>2</td>
<td>-</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>Missing</td>
<td>14</td>
<td>-</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>10</td>
<td>91</td>
<td>161</td>
</tr>
</tbody>
</table>

4.3 Relation 3; between operator choice complexity and cognitive automation

The choice complexity is directly influenced by the choices and variance of solutions. An increased number of models, parts, tools etc. will result in an increase of choice complexity. Table 5 shows the number of variants and the demand of the most common variants. The choice complexity measure cannot directly be reduced by cognitive automation. However, introducing cognitive automation can reduce the perceived complexity caused by the increased choice complexity.

<table>
<thead>
<tr>
<th>Objective Complexity Elements</th>
<th>Station 11</th>
<th>Station 13</th>
<th>Station 23</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCC</td>
<td>3.9</td>
<td>3.8</td>
<td>4.5</td>
</tr>
<tr>
<td>Number of tools</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Number of variants</td>
<td>31</td>
<td>27</td>
<td>51</td>
</tr>
<tr>
<td>Demand for each variant</td>
<td>8 variants accounted for 77 percent</td>
<td>9 variants accounted for 78 percent</td>
<td>6 variants accounted for 51 percent</td>
</tr>
</tbody>
</table>

At the investigated stations decision support is given by Pick-By-Light and process support is given by monitors and tools associated with tightening operations. Tightening tasks are easy

---

\(^1\) Not observed during the LoA assessment assessed afterwards

---
to control and restrict while assembly operation, which is done without any use of a tool, are very hard to monitor and control. Many manual tasks on the stations were to connect electrical connections. However neither decision nor process support was given when performing contact operations. The information regarding these operations was to be found in binders at the stations. If the contact operations are missed or badly performed the error is not acknowledged until on later control stations while tightening operations are controlled within the station boundaries by a control system connected to the tools.

4.4 Relation 4; is there a relation between cognitive automation, quality and quantitative (objective) station complexity?

Earlier empirical results [15] show that in general, system complexity, does affect performance negatively and that training and that man/machine interface plays important roles in minimizing the negative effect of system complexity on performance.

Results from previous sections show that relations could be made between quality, complexity and cognitive automation. Believes are that cognitive automation can be used as a mean to reduce the negative effects of choice complexity in terms of quality.

5. Conclusion

This paper shows that it is possible to use quantitative measures in order to show relation between station complexity, quality and cognitive automation. These methods could be further used in order to improve both the resource efficiency and resource allocation in order to get an effective assembly system. Then, the operators’ competence and experience should also be taken into consideration, which is not fully covered by using the three methods.

The main conclusion is that there is evidence that cognitive support is needed in final assembly to minimize the negative effects of complexity.

Acknowledgment

The authors like to express their gratitude to all involved personal at the case study and the collages within the COMPLEX project. This work has been funded by Vinnova within the program Production Strategies and Models for Product Realization.

This work has been carried out within the Sustainable Production Initiative and the Production Area of Advance at Chalmers. The support is gratefully acknowledged.

References

A hybrid human-robot assistance system for welding operations - methods to ensure process quality and forecast ergonomic conditions

F. Busch², C. Thomas¹, J. Deuse³, B. Kuhlenkoetterb
¹Chair of Industrial Engineering, TU Dortmund University, Dortmund, Germany
²Chair of Industrial Robotics and Production Automation, TU Dortmund University, Dortmund, Germany

Abstract: This paper discusses the development of a multi-robot system to reduce labour-intensive manual handling and positioning of heavy parts in welding processes by using two vertical joint arm robots for work assistance. It focuses on technical means to ensure a high product and process quality in hybrid robot systems. The presented results are part of the project “rorarob” and are supported by a prototype at the TU Dortmund University, Germany. First concepts were presented at 3rd CATS 2010, Trondheim [1], Safety Aspects and Human Factors of Hybrid Robot Systems have been discussed in [2],[3],[4].

Keywords: Assembly, Robot, Welding

1. Introduction

The paper presents aspects of process and human factors engineering for prospectively designing a hybrid robot assistance system for welding applications. This includes the forecast of dimensional variations due to influences of heat and inaccuracies of the kinematic chain (process engineering) and the evaluation of the ergonomic conditions (human factors engineering). Both key aspects are taken into account early in the planning by designing the hybrid robot system using an offline simulation system.

Welding is a labour-intensive production process, because batch-sizes are small and products are mostly custom-made. Especially in small and medium sized businesses, welding operations are performed manually, with a low level of automation. As a result, the worker performs several other tasks beside the welding process, like handling and positioning of parts or preparing the weld seam.

Because sizes and weights of the assemblies exceed in many cases the load carrying capacity of a human, it is not possible to reposition the part without high effort. Due to these restrictions, the worker has to change his position several times to reach all weld seams and use cranes or other handling tools to reposition the parts during the process. This leads to a high physical load on the employee during the process [5] and long lead times. In consequence several breaks are necessary to compensate the physical stresses and keep the high level of concentration needed for a high quality output.

The idea of the research project “rorarob” [1],[4] is a robot-based assistant system for supporting the handling of heavy parts in manual welding operations. Because flexibility requirements of custom-made constructions do not allow a full automation with closed robot cells, known from the automotive industry, the project intends a combination of manual and automated operations in direct human-robot cooperation. In a hybrid system for robot assisted welding, the worker has the possibility to influence the pre-planned welding task during the manually done process while the robot carries the weight of the parts and does the part-handling between the production steps. The objective is an improvement of ergonomic conditions and a significant reduction of time required for handling. Test runs performed with product samples showed a decrease of manual handling up to 50 %.

For a direct human-robot interaction a flexible safety concept has to be integrated. A concept of four different operation modes for collaboration between humans and robots are described in [2] and [3].

Figure 1. Demonstrator at TU Dortmund University

The described assistance system is realised as a demonstrator at TU Dortmund University (Figure 1). All components of the demonstrator have a high maturity level regarding industrial usability. The multi-robot system contains two handling robots. One handles the assembly and has a maximum payload of 400 kg and a working range of 2.55 meters. The other robot transports the additional parts during the welding process. It has a payload of 150 kg and a working range of 2.2 meters.

The layout and the path-planning of the hybrid robot system are done by an offline programming system. For this purpose a digital human model (DHM) has been integrated into an existing offline simulation tool for multi-robot systems. The simulation of human movements within the offline programming allows a prospective planning of the whole process regarding ergonomics and safety (e.g. collision avoidance).

2. Challenges in Robot Assisted Welding

Main challenges from the process side are internal and external influences on the multi-robot system. Internal influences are based on errors during the gripping process or inaccuracies of the robots. External influences like position errors are following by welding distortion, based on the welding process. Subsequently the welding technology will be described briefly.
In industrial manufacturing several processes appear along the production process. According to the German DIN-standard 8580 [6] manufacturing processes are classified by function in casting and molding, shaping, machining, joining, surfaced and changing material property. In most welding procedures the basic metal and additional material is melted up to combine parts. By solidification of the material the parts are united. [7] Welding procedures are classified by acting principle in pressure welding and fusion welding. In fusion welding technology heat is applied by different sources. In arc welding, a form of the fusion welding, an electric arc is generated between base material and the cathode of the welding tool. [8]

The geometrical dimensions are influenced by the size of heat impact. The level of energy and the position of heat influence are important factors regarding welding distortion. These can be countervailed by fixations. The disadvantages of fixations are strains in the welded material and a high time effort to realise fixed positions for all parts and positions of an assembly (Figure 2). Therefore aim of the on-going research is to use robots for substituting standard fixations.

![Figure 2. Worker during the fixation of parts](image)

Other sources of inaccuracy in robot-based handling systems are incorrect gripped workpieces, grounded on errors in the gripping process. Also high tolerances of semi-finished products and deviations in the kinematic chains are sources of inaccuracy.

3. Welding Strategies for Robot Assisted Welding

For ensuring a high accuracy of the welded assemblies a suitable strategy is necessary for welding in an robotic system. By joining parts a rigid connection between both robots is established. The challenge is to avoid influences of welding distortion on the mechanical connected multi-robot system. In cases of too high position errors the robots stop their movements. The distortion during the process is measured by an optical coordinate measurement system K610 of Nikon Metrology NV with help of infrared LEDs. The system is build up with three linear CCD cameras for measuring the position of each LED by triangulation. Configurations with three or more LEDs are able to measure position and orientation of coordinate systems with up to 1000 Hz.

Experiments showed relative position errors at the tool center points (TCP) of both handling robots, increasing in different situations: Even by synchronous, but unconnected movements position errors are followed by the inaccuracy of the kinematic chain. This fact increases especially during reorientation of parts. Beside the system-related inaccuracy of robots, relative movements occur during the welding process as an impact of welding distortion. A comparison of two different welding strategies is described in the paragraph below.

In a stop position of the handling robots relative movement of the robot grippers is measured during a manual welding process of gripped workpieces, exhibited in figure 3.

Two different welding strategies with equal parameters (position and orientation in working range of the robots, identical workpieces, welding parameters, 200 mm weld seam) are analysed. These are tested in the robot system to analyse if there are differences in comparison to manual welding with standard fixations. In the case A (Figure 4) two workpieces are directly combined without a spot welding before. In the second strategy workpieces are stapled by spot welding before (case B).

The measured data represent only the weld seam; relative movement during the welding process is referred to the beginning of the weld seam (Figure 5). Both measurements start with the welding process. Because of the sensitivity of the optical measurement system to ultraviolet light, measurement during the welding process can fail. After welding (case A: 1:05 minutes, case B: 1:30 minutes) the diagrams show cooling phases up to more than 10 minutes.

Both diagrams present a similar characteristic. Especially the relative movement in x direction (as direction of gravitation) is nearly identical. A conspicuous characteristic of both measured data is the run along the welding line (y axis). During welding process the relative movement is maximal in comparison to other axes. But during cooling phase the gap becomes smaller. An equal reaction of the system can be observed (case A: ~1.35 mm, case B: ~1.60 mm). At first no significant influence to the distortion seems to arise from different welding strategies. But for the usage of a robot based assistance system a cooling time of more than 10 minutes is time intensive and uneconomical. For an uninterruptible production flow it is necessary to open the gripper directly after welding and bring the next part to the next welding position. By comparing the two strategies the distortion after welding (case A: 1:05 minutes, case B: 1:30 minutes) is in each direction half as size when the parts are stapled before.

Followed by these measurements it is clearly that even in a robot based assistance system the stapling process is necessary. This is essential for a high process stability and product quality, basic requirements for an industrial applicable solution.

Finally the different measurements of robot movements and welding distortion clarify an undesired relative movement of the two handling robots to each other. Especially during...
are able to prospectively design and simulate the entire process, single and multi-robot operations, effectors and material supply. Including cell layout, path planning and motion simulation of manufacturer to manufacturer. These software tools in general are an integral part of turnkey robot systems and differ from programming systems are available on the market. Often the user is able to simulate the entire system, including path and robot movements can be simulated in one environment, the robots, requires a digital human model (DHM). Only if human stress on the human body is minimized and collisions are excluded. Therefore a digital human model (DHM) is integrated into a hybrid robot system virtually in advance.

In comparison to position preservation during welding, an reorientation or as a consequence of welding distortion, the position errors achieve a critical value. During welding operations, the stiffness of fixation has to be as large as possible. In comparison to position preservation during welding, an unlocked connection of the robots during the movements affords unrestricted robot movements to allow ergonomic welding positions. A mechanical compensation unit in form of a clutch will be designed, constructed and implemented into the demonstrator, for compensation of inaccuracies.

4. Path Planning for Robot Assisted Welding

The choice of a suitable welding strategy it is necessary to plan the robot trajectories in a way that physical stress on the human body is minimized and collisions are excluded. Therefore a digital human model (DHM) is integrated in an offline simulation environment to design and prove the hybrid robot system virtually in advance.

To plan and simulate robot systems, several offline programming systems are available on the market. Often they are an integral part of turnkey robot systems and differ from manufacturer to manufacturer. These software tools in general are able to prospectively design and simulate the entire process, including cell layout, path planning and motion simulation of single and multi-robot operations, effectors and material supply.

The computer-aided design and simulation of hybrid robot systems, including a direct cooperation between humans and robots, requires a digital human model (DHM). Only if human and robot movements can be simulated in one environment, the user is able to simulate the entire system, including path planning, collision analysis and evaluation of process parameters and human factors (e.g. ergonomics).

DHMs represent the shape and movements of the human body in a virtual work environment and are part of the “Digital Factory” [9], the generic term for computer-aided planning tools along the product life cycle, which have significantly increased in recent years [10].

Empirical studies carried out by the Chemnitz University of Technology showed, that three different DHMs have the most industrial relevance in Germany [11]. The most important DHM is RAMSIS by Human Solutions (42%), developed for the automotive industry. The two others are Human Builder by Dassault Systems (26%) and Jack by Siemens PLM (16%) [12]. In most cases, DHMs help to design and analyse products or workplaces regarding ergonomic aspects [12]. A good example is the RAMSIS model, initially developed for designing passenger compartments and cockpits by performing visibility and reachability analysis.

All these models have in common that they have reached a very detailed mapping of the human musculoskeletal system. But the necessary effort for programming is often disproportional to the expected benefits. The available software solutions are complex, expensive and need expertise and training. For this reason, the usage of DHMs is limited to larger companies with specialised planning divisions or to specialised service providers.

Part of the research project “rorarob” is a new approach in implementing a human model and its movements into an offline simulation environment. The aim is to significantly reduce the programming effort for simulating human movements. The basis for this approach is the software FAMOS (www.carat-robotic.de), an offline programming system developed especially for small and medium-sized businesses. FAMOS is able to simulate single- and multi-robot applications, like for example grinding or polishing. To simulate a hybrid robot system, with manual and automated operations in direct human-robot cooperation, a DHM is implemented into the software. The common existing human models, like RAMSIS, Jack or Human Builder, are big and complex tool boxes, where even a single license is mostly too expensive for an economical usage in small and medium sized businesses.

Therefore a new approach is followed in this project, to develop a solution, which concentrates on the basic function to simulate hybrid robot systems. The aim is to significantly reduce the time needed for building up a simulation model for manual human operations. For this purpose, a character animation system coming from the entertainment industry is used as basis for implementing a DHM into FAMOS. The implementation of the basic functions is done by carat robotic innovation GmbH, the Chair of Industrial Engineering at TU Dortmund University is responsible for the ergonomic design of the human model. Part of the implementation is a database of different human models for representing a variation of body sizes (e.g. from 5-percentile female to 95-percentile male operators), based on a skeleton model covered by a wire frame overlay. The software is able to simulate the movements of the human and the robots and perform a collision and ergonomic analysis during the process.

Beside the detection and prevention of collisions, an ergonomic assessment procedure for evaluating the motion sequence of the human model regarding ergonomics is
performed during simulation. In this case the OWAS method (Owako Work Analysis System) [14] is utilised. This evaluation methodology is based upon observation of body postures, which are described by the combination of all joint positions of the human model. The OWAS algorithm identifies each posture and its duration and percentage in the analysed motion sequence. The joint coordinates and motion data are provided by the human simulation and are used for evaluating the resulting load on the musculoskeletal system for back, arms and legs. Final result is a classification for the level of risk for injuries on the musculoskeletal system, which indicates the need and urgency for corrective actions regarding the movements of the human body in the analysed configuration.

Figure 6. Path Planning of a Digital Human Model created with the EmotionFX Plug-In of FAMOS [13]

For building up a simulation model of a specific welding task, the movements of the human and the robot have to be mapped. At first, similar to offline programming of robot trajectories, the walking paths of the human model is mapped by drawing trajectories on the virtual shop floor (Figure 6). This path can be associated with different human actions, like specific work tasks. These actions can be for example standard motions, like get up, sit down, kneel, walk or run, stored as motion capture clips in a database. This motion library can be expanded indefinitely, especially for assembly tasks. Beside these standard movements, parts of the human model, for example the hand, can be associated with other geometries and moved along a trajectory. Typical example is a welding process, in which a welding torch is moved along the weld seam. Motions like this depend on the individuality of the work task and cannot be covered by a standardised motion library, it is planned to integrate standard time building blocks, described in predetermined motion time [8].

To further extend the motion library, it is planned to integrate standard time building blocks, e.g. reach, grasp, move, position, release and higher aggregated blocks, described in predetermined motion time systems like MTM (Methods Time Measurement).

5. Conclusion

This research shows possibilities and constraints for designing a hybrid robot assistance system for welding applications. The results point out what dimensional variations arise during the process due to influences of heat and inaccuracies of the kinematic chain and how they can be taken into account in the system and process design. The next step is the development of a technical solution to compensate these inaccuracies in real time during the process. Beside process parameters, a robot system designed for a direct human-robot interaction necessitates adapted offline simulation systems. The implementation of a digital human model in existing software solutions allows to plan the design of hybrid robot systems, including ergonomic aspects, in advance. For a broad use of DHMs in the planning of manual or hybrid work systems, the effort for modelling human movements needs to be reduced significantly. One approach is the described combination of standard motions stored in a database and individually calculated hand-arm movements.

Acknowledgements

This Research & Development project is funded by the German Federal Ministry of Economics and Technology as part of the technology program “AUTONOMIK-Autonomous and simulation-based systems for SMEs”. The project management agency “Convergent ICT/Multimedia” of German Aerospace Center (DLR), Cologne is in charge of the project. Moreover we want to thank our associated industry partners carat robotic innovation GmbH, MAN Diesel Turbo SE and Boecker Maschinenwerke GmbH.

References


Design for collaboration: 
a development of human-robot collaboration in assembly

J. T. C. Tan\textsuperscript{a}, T. Inamura\textsuperscript{a}, T. Arai\textsuperscript{b}

\textsuperscript{a} Principles of Informatics Research Division, National Institute of Informatics, Japan
\textsuperscript{b} Department of Precision Engineering, University of Tokyo, Japan

Abstract: The changing production requirement and the emerging of intelligent robots have introduced the concept of collaborative robots into manufacturing. However, most of the recent assistive robots developments are focused on the robot design based on existing assembly processes. In order to achieve effective collaboration, this work initiates a proposal to redesign assembly system for the feasibility of collaboration. Based on several prior developments, this paper integrates all the developments to illustrate the overall concept of design for collaboration, from (a) the proposal of task-based modeling, collaboration analysis and simulation to study the assembly process design for collaboration, (b) safety management that involves safety strategy development and safety evaluation by risk assessment and human mental safety study, to (c) an actual prototype development and (d) performance evaluations. The positive performance outcomes in handling, assembly and overall collaboration system have greatly built up confidence towards the proposal of the design for human-robot collaboration in assembly.

Keywords: Man-machine system, Assembly, Simulation

1. Introduction
The changing production requirement, from fewer designs with large quantity to various designs with small quantity, is pushing the evolution of manufacturing systems, from mass production lines to flexible production cells. The short product life cycle and costly automated systems have made fully automated robotics systems impractical to address the flexibility requirement in modern assembly processes. Skillful human operators are still much preferred in flexible manufacturing systems. However, this solution trades off much productivity for the flexibility. To address this shortcoming, the emerging of intelligent robots has introduced the concept of assistive robots in many recent developments.

1.1. Assistive robots
Many developments in the area of human-robot collaboration in terms of kinematic assistance were established in recent years, e.g. Robot Assistant rob@work [1], COBOT [2], KAMRO (Karlsruhe Autonomous Mobile Robot) [3], CORA (Cooperative Robotic Assistant) [4], Humanoid Service Robot HERMES [5] and The Manufacturing Assistant [6]. In order to realize fully flexible and optimum collaboration, the assisting robots have to be competence for workplace and time sharing systems [7]. These robots are able to perform handling and assembly tasks independently and collaboratively, to improve working efficiency by automating portions of the work and enabling the human operator to focus on the portions of the work requiring human skill and flexibility.

Human-robot collaboration in assembly seems to be an optimistic and promising proposal in the future development of manufacturing systems. However, most of the current works are focused on the robot design based on existing assembly processes. In order to achieve effective collaboration, this work initiates a proposal to redesign assembly system for the feasibility of collaboration. Based on several previous works, this paper integrates all the developments to illustrate the overall concept of design for collaboration, from the study of assembly process design for collaboration (Section 2), safety management (Section 3), to an actual prototype implementation (Section 4) and performance evaluations (Section 5).

2. Assembly process design for collaboration
2.1. Task-based modeling
The study starts with the concept of human-robot collaboration in assembly and relations among assembly, human operation and robot control. A task-based modeling framework is developed to integrate the abstract model of assembly, the task model of human operation and the low level robot control into a single representation [8].

The entire assembly process is regarded as a long chain of many sub task units. Each sub task unit can be recursively decomposed into smaller sub task units and eventually the whole process can establish a hierarchical structure of tasks (Fig. 1). The top levels of tasks contain of assembly relationship (assembly task) and branch down into human operations (control task). Certain human operations that can be automated by robots are further branch down into robot motions. The definition of assembly task and control task are as follows [9].

Assembly task describes the assembly relationship between the product and the parts (or sub-assemblies) it contained.

\[
\text{Assembly Task, } T_a = \{X, Y\} \tag{1}
\]

where,

\[
\text{Part 1, } X = x_1, x_2, x_3, \ldots
\]

\[
\text{Part 2, } Y = y_1, y_2, y_3, \ldots
\]

Control task is the description of human operation to perform the task, in terms of assembly action, the parts (or sub-assemblies) involved, and the locations of assembly.

\[
\text{Control Task, } T_c = \{V, X, Y, L_x, L_y\} \tag{2}
\]

where,

\[
\text{Action, } V = v_1, v_2, v_3, \ldots
\]

\[
\text{Part 1, } X = x_1, x_2, x_3, \ldots
\]

\[
\text{Part 2, } Y = y_1, y_2, y_3, \ldots
\]

\[
\text{Assembly Location of Part 1, } L_x = l_{x_1}, l_{x_2}, l_{x_3}, \ldots
\]

\[
\text{Assembly Location of Part 2, } L_y = l_{y_1}, l_{y_2}, l_{y_3}, \ldots
\]

The above task description definitions of assembly task and control task are proposed as the decomposition criteria to decompose the composing process in hierarchical task analysis development. Starting from the assembly task level, the product
can be decomposed (disassembled) into the parts or sub-assemblies it contained. The sub-assemblies are then further decomposed into their respective parts or sub-assemblies. The decomposition continues until the assembly task contains only two single parts, i.e. Part 1 and Part 2.

The task decomposition process is then continued in the control task level. Similarly, using the approach used in assembly task, the decomposition of control task is carried out until only a single Action, Assembly Location of Part 1 and Assembly Location of Part 2 are present.

Among the control tasks, certain human operations that can be automated or collaborated by robots (determined in collaboration analysis) are further branch down into robot motions based on the robot programming.

2.2. Collaboration analysis

Along with the task analysis, collaboration analysis is developed [10] to describe human-robot collaborative operations in assembly. The collaboration is analyzed from the given assembly operations and the tasks are reassigned to the appropriate working agent based on the capabilities of the human operator and the robot system.

The analysis can be done in two stages, qualitative and quantitative, based on the complexity to determine the optimum collaboration solution for a given task. In qualitative analysis, the performance requirements of the task are compared qualitatively with the capabilities of human and robot to identify possible collaboration solution. If the optimum solution is not apparent, quantitative analysis by analytic hierarchy process (AHP) [10] can be conducted to score the possible solutions based on the performance requirements. For example, four main criteria (productivity, quality, human fatigue, and safety) are used in previous study [10] to evaluate the cable-connection insertion task between human only and human-robot system.

2.3. Simulation

A human-robot collaboration simulation environment is under new development (Fig. 2) [11] to simulate the collaboration task model for assembly process design and assessment. The tasks are represented in simplified agent motions for the collaboration operation. Human operator can control his/her own avatar to perform task motions in collaboration with the robot agent. The robot agent will response to the human avatar’s actions based on the collaborative operation sequences. The collaboration processes are assessed based on the feasibility of the assembly and the tact time comparison between different models.

3. Safety management

Besides conceptual investigation on the collaboration in assembly, this work also discusses on the safety management, which weighs paramount importance in real world implementation.

3.1. Safety strategy and development

As the current revision of international safety regulations towards allowing more human-robot interactions in a robot cell, this project has proposed several key safety techniques to safeguard human-robot collaborative assembly operations [12]. The objectives of the safety developments are: (1) Work cell and robot safety design, (2) Operation control safety, and (3) Collaboration safety assessment.

---

**Figure 1.** The hierarchical task-based modeling concept illustrated in a cable-connector insertion assembly.
3.1. Work cell and robot safety design

The entire work cell is divided into human and robot areas with a low safety fence and is monitored by photoelectric sensors. The safety fence provides a physical barrier, while the photoelectric sensors connected to the operation control system act as a secondary defense. Across the workbench, two sets of double light curtains are setup to establish the safe collaborative working zones (Fig. 3). In Zone C (Collaboration Mode), the robot system is restricted as part of the safety control strategy in a highly reliable dual check safety (DCS) system. The system monitors both robot speed and area restrictions according to ISO regulations.

Apart from the force sensor attached on the wrist of the gripper to monitor collision, the entire robot is built with safe mechanical design as the fundamental safety strategy to improve robot reliability. The overall mobile robot manipulator system is built with a low center of gravity to prevent toppling. Moreover, for precise movement and to avoid collisions, the system utilizes a vision system to detect marks on the floor with an accuracy of 5 mm and 0.1 degrees. The base is equipped with a bumper switch to detect collision with objects as well as wheel guards to prevent foreign object from becoming entangled with the wheels.

In this development, a triple stereo visions system is developed for safety monitoring in human-robot collaboration [13]. Network cameras are used to capture images for tracking of color areas on the human operator. The image coordinates by particle filter and human body model are combined to estimate the 3D positions for the human motion monitoring. Based on the human motions, high risk or danger conditions of the human operator can be detected to safeguard the operator during production.

3.1.2. Operation control safety

Besides optimizing production rate, operation planning and control play important roles in ensuring safe collaborations between a human operator and robot manipulator. From the early task modeling of the collaborative operations (Section 2), operation content and detailed planning including safety status can be defined for each step of the operation. During the operation, the operation control system coordinates the collaboration process, while monitors the inputs from control switches, safety sensors (i.e. photoelectric sensors on safety fence and vision system), and light curtains to ensure safety status of both human and robot are verified throughout the whole operation. For instance, the robot system’s position in the designated zone with respect to operation progress is determined using input from the front double light curtain system, and any unintended intrusion will trigger an emergency stop by the operation control system. If an emergency occurs, the operator can identify the error source from the status of each device, and recover the system from the user interface to continue from the previous step.

3.1.3. Collaboration safety assessment

Two safety assessment approaches are proposed in this work to verify collaboration safety. The first approach is the development of risk assessment [12] and the second approach is the investigation on human mental safety, which will be discussed in the next section. Based on industrial standards ANSI/RIA R15.06 with reference to ISO 14121 (JIS B9702), ISO 13849-1 (JIS B9705-1), and BS EN 954-1, the risk assessment is conducted to gauge the safety performance of the developed system in terms of safety design. The initial risk identification (before implementing safety design) is obtained from the task modeling of the assembly operation. The identified risk is then measured by the severity of the hazard, exposure and avoidance categories to determine the risk reduction category. Afterwards the safeguard selection matrix is used to determine the minimum safeguards, which are then incorporated into safety design. To analyze the safety design based on risk reduction, safeguard selection validation is conducted according to the matrix. For comparison, the assessment includes the equivalent performance level, PL, and Category based on ISO 13849-1.

3.2. Human mental safety

Another important development concerning safety in this work is the initial effort to quantitatively assess human mental burden during collaboration operation. J. W. Crandall and M. A. Goodrich [14] have proposed two aspects of safety studies: physical safety and mental safety. The aforementioned risk assessment study indicated that the system safeguards minimized physical injury of the human operator by the robot system. On the other hand, the close proximity to a robot might induce a mental workload on the human operator. Thus, mental safety strives to prevent such a mental workload, which can influence the operator’s physiology and affect productivity.

The main purpose of this development is to propose several practical methods to study on human mental workload in the collaboration design. There are two different approaches in...
assessing operators’ mental workload: physiological parameters assessment and subjective assessment. In physiological parameters assessment, the biological reactions induced by mental strain (Fig. 4) are measured during collaboration experiments. Skin potential reflex (SPR) data, which is an effective indicator of nervousness, was recorded as a physiological measurement (amplitude ratio and occurrence rate). For a parallel comparison, the rating scale method (0 – 6 scores) was used as a subjective assessment to investigate ‘fear’ and ‘surprise’ levels. Several experiment studies [15] had been conducted to investigate the effect of the front protruded robot motion speed on the collaboration, the human-robot working (bin-picking) distance, the robot motion trajectory, and the effect of preknowledge on the robot motion. Finally, a low mental workload production framework and evaluation development were proposed [15].

Figure 4. Change in physiological parameters caused by cognitive and emotional stress [15].

4. Prototype implementation
All the proposals in this work have been implemented in a prototype cellular manufacturing system (Fig. 5) [12]. In this development, a twin arms manipulator mobile robot system and an information support interface system are designed and developed to facilitate human operator in a flexible cable harness assembly.

Figure 5. Human-robot collaboration prototype system.

The earlier task modeling development is not limited to a modeling representation of the collaboration system that facilitates assembly process design, but it is also closely integrated with the human-robot support interface system [9] in the actual implementation to coordinate all the collaboration flow in the operation.

5. Performance Evaluations
Several system performance evaluations, for instance, bin-picking of assembly components and cable insertion assembly are conducted for comparison studies between conventional manual operation and human-robot collaboration operation.

5.1. Handling performance
The bin-picking task in the handling experiment is to collect 9 assembly parts from the containers on the floor and arrange them into the parts tray as shown in Fig. 6(a). The objective of this experiment is to compare the productivity of robot and human operator per day. Human operator is assisted by digital picking system, where light indicator is attached on the containers to prevent mistake.

The average time required by the robot system to prepare one part into the parts tray is 10.5s while human operator required 6.0s. To calculate based on the fact that robot system can work 8 hours x 3 shifts per day, the productivity of robot system is 72% more than human operator (Fig. 6(b)). This result has proven the rational to automate the bin-picking task, while having the robot to work parallel while waiting for human operator to complete other assembly task will definitely improve the overall system productivity.

5.2. Assembly performance
In the cable insertion assembly experiment, the task is to insert 64 cable pins into the connector based on a randomly instructed order (Fig. 7(a)). The objective of this experiment is to compare the performance of human-robot collaborative assembly operation over manual operation by the human operator. In the human-robot collaborative assembly, the robot will indicate the insertion holes by a laser pointer attached on its gripper. In manual operation, the human operator needs to refer to a graphical diagram showing the insertion holes.

From the experiment results (Fig. 7(b)), after 5 times of operations, the human-robot collaborative assembly required only 50% of the manual operation time to complete the assembly task. It is proven the highly effectiveness of the
collaborative operation over manual operation in the repetitive assembly task.

Figure 7. Cable insertion assembly experiment.

5.3. Overall collaboration system performance

A comparison study was conducted between human manual operation (Exp I) and the human-robot collaboration system (Exp II) in a cable harness assembly [12]. A comparison between novice and expert operator is also included in the experiments.

The experiment results (Fig. 8) show that better performance is achieved in the collaboration setup (Exp II). It is proven that the collaboration between human and robot was carried out well with the support interface system. Also, the identical performance of both novice and expert subjects in Exp II has shown that the skill difference is less significant as the cable harness assembly is relatively simple. The minimum assembly duration was recorded even from the first assembly trials, proving that the system enabled high adaptability of the operators to work in the collaborative system. This result is a good evidence of the effectiveness of the collaboration system.

Figure 8. Overall collaboration system performance.

6. Conclusion and future work

The aim of this work is to initiates a proposal to redesign assembly system for the feasibility of collaboration. Based on several prior works, this paper integrates all the developments to illustrate the overall concept of design for collaboration, from (a) the proposal of task-based modeling, collaboration analysis and simulation to study of assembly process design for collaboration, (b) safety management that involves safety strategy development and safety evaluation by risk assessment and human mental safety study, to (c) an actual prototype development and (d) performance evaluations.

The positive performance outcomes in handling, assembly and overall collaboration system have greatly built up confidence towards the proposal of design for human-robot collaboration in assembly. This work has served as the basic development framework and further studies are required for optimization.

References


Adaption of processing times to individual work capacities in synchronized assembly lines

G. Reinhart*, M. Glonegger*, M. Festner*, J. Egbers*, J. Schilp*
* Institute for Machine Tools and Industrial Management (iwb), Application Center Augsburg, Technical University Munich, Germany

Abstract: Increasing demands on the flexibility of assembly workers lead to numerous times of high exposure rates within one shift. Inflexible processing times in synchronized assembly lines mismatch individual, circadian variable work capacities of operators. Hence, frequent performance peaks due to limited possibilities for individualization elevate physiological and psychological workloads. Main objective is a temporal individualization of assembly system performance requirements. The conceptual framework presented in this paper gives the opportunity to investigate how processing times can be adapted to individual work capacities. An adaption of cycle time leads to an adaption of working speeds to reactively level workloads within one shift. The authors focus on the identification of influence factors on cycle time and their qualitative impacts.

Keywords: Assembly Design and Control, Human Factors in Assembly, Individual Performance

1. Introduction

Increasing competitive pressure, know-how diffusion and market globalization are modifying the competitive patterns of manufacturing enterprises [1]. Accordingly, the high number of product variants presents enormous difficulties in the design and operations of assembly systems [2]. Since the introduction of assembly lines for mass production by Henry Ford, several developments led to changes from strictly paced and straight single-model lines to more flexible systems [3].

In the last decades, manufacturing system design methods were developed to investigate the dynamic behavior of manufacturing systems, to estimate their performance and to support their efficient design, improvement and reconfiguration [4]. Numerous scientists and practitioners focus on assembly line design, assembly line balancing and other subjects in the context of planning, optimizing and controlling assembly production systems. Configuration planning generally comprises all tasks and decisions which are related to equipping and aligning the productive units for a given production process before the actual assembly starts [5]. This includes setting the system capacity (cycle time, number of stations, station equipment) as well as assigning the work content to productive units (task assignment, sequence of operations) [5].

The above listing shows the commonly used technological approach to balance assembly lines. In this context, individual work capacities are planned in a constant manner, which leads to inflexible work demands and performance requirements in synchronized assembly lines. By contrast, humans do not perform work tasks in a constant way within one workday. A multitude of factors influences the current work capacity of an assembly worker. Accordingly, the human circadian rhythm, the chronotype as well as organizational aspects as shift times play a major role. Contemporary production systems cannot adapt to individually fluctuating human performance levels. Consequently, integrating elderly operators within assembly systems with fixed cycle times becomes a challenge for production designers. Physiological and psychological deterioration inhibits long-term ability-equitable employment.

An optimized adaption of processing times to individual work capacities countervails impairments due to high work exposures.

2. Cyclic assembly operations

2.1. Current approaches

An assembly line (AL) is a flow-oriented production system where the productive units (PU) performing the operations are aligned in a serial manner [5]. PU can be defined as a series of manual or automated assembly work stations through which one or multiple product(s) are sequentially assembled [6]. An AL consists of stations arranged along a conveyor belt or a similar mechanical material handling device. Work pieces are consecutively launched down the line and are hence forwarded from station to station until they reach the end of line. A certain set of operations is performed repeatedly on any work piece which enters a work station, whereby the time span between two entries is referred to as cycle time [5]. Cycle time applies to the time between two consecutive finished products leaving the AL (output) and represents the maximum or average time available for each work cycle [3; 6].

For dividing assembly processes among work stations, the total work content to assemble a work piece is sectioned into subsets of elementary operations called tasks. Tasks are indivisible work units and thus, each task is associated with a processing time referred to as task time. Due to technical and organizational requirements, tasks cannot be carried out in an arbitrary sequence, but are subject to precedence constraints. General input parameters can be conveniently summarized and visualized by a precedence graph [5]. The decision problem of optimally partitioning (balancing) tasks among work stations with respect to some objective is known as the assembly line balancing problem (ALBP) [3].

Work stations of unpaced ALs operate with slightly different task times, leading to mismatches between performance requirements and work contents to be executed within one cycle. In this context, despite operating on equal cycle times, different work stations have variable processing times. Buffers between work stations can partially reduce these difficulties. In
case of a buffered AL, the ALBP needs to include positioning and dimensioning of buffers [3]. In order to avoid excessive capacity utilizations, the cycle time is planned as an average time representing all models. Consequently, processing times of certain models are higher than the cycle time. This can finally result in work overloads whenever the task cannot be finished within the station’s cycle time borders [3; 7].

2.2. Problem statement

Buffers connecting work stations serve as balancing marginal fluctuations of processing times [8]. Working speed regulations on an individual level depend upon buffer sizes up- and downstream the work station. Nowadays work schedulers do not differentiate cycle-times according to different performance levels of employed operators. Current cycle-time models assume a constant work speed [8]. Consequently, assembly workers have to perform at an equal level of operating speed during the whole workday (see Fig. 1).

The straight line visualizes constant performance requirements of an assembly system during the morning shift. Due to cycle time restrictions, ALs have fixed production rates [3]. The cycle time has a uniform level at the beginning, the middle, and the end of a shift. Though, restrictions within production systems with a high division of labor between workplaces are so numerous that performance requirements and working speed can be seen as relatively inflexible between shift times. In contrast, due to a multitude of influence factors, human work capacity is individually fluctuating. This causes numerous times of stress as well as under-challenged demands within one shift (colored grey).

Considering the integration of elderly or medically constrained employees into multi-variant assembly systems, three development targets can be set up [9]. One of those is the transition to adaptable human work levels within one shift. To reduce physiological and psychological stresses and strains, design of working speed should take the human circadian rhythm into account. Existing ergonomic approaches as well as approaches from applied psychologists can be summarized as follows [9]:

- A decoupling of work stations within the AL and according tasks at the work station enables more flexible timing and work performance.
- A more flexible superordinate structuring of work with at least partly redundant or oversized systems increases the range for an individualized line-balancing.

- By providing selective information needed at a work station, a consistent consideration on systems ergonomics and industrial psychologies acts alleviating on stress.

The current situation in synchronized ALs reveals inflexible processing times in contrast to individual human work capacities. Subsequently, individually fluctuating working speeds lead to numerous times of high exposure within one shift.

3. Conceptual framework

The main objective is a temporal individualization of assembly system performance requirements. This gives the opportunity to investigate on how to adapt operators’ processing times (control variable) to individual work capacities. In this context, the approach allows to reactively level workloads. The conceptual framework presented in this paper focuses on identifying influence factors on cycle times to expand cycle time ranges (control lever). Fig. 2 visualizes the superordinate methodology as well as the conceptual framework to be described in the following.

![Figure 1. Constant cycle times cause an equal level of working speed during shifts and lead to varying work exposures](image)

The straight line visualizes constant performance requirements of an assembly system during the morning shift. Due to cycle time restrictions, ALs have fixed production rates [3]. The cycle time has a uniform level at the beginning, the middle, and the end of a shift. Though, restrictions within production systems with a high division of labor between workplaces are so numerous that performance requirements and working speed can be seen as relatively inflexible between shift times. In contrast, due to a multitude of influence factors, human work capacity is individually fluctuating. This causes numerous times of stress as well as under-challenged demands within one shift (colored grey).

![Figure 2. Conceptual framework of the approach described in the following](image)

3.1. Identification of influence factors on cycle time and their qualitative impacts

As a first step, influence factors on cycle times that restrict planning and adapting cycle times must be detected. Restrictions characterize factors which limit modification boundaries when adapting cycle times. Restrictions can be subdivided into technical, personal and organizational influences.

<table>
<thead>
<tr>
<th>Parameters (rising value)</th>
<th>Impact on cycle time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conveying velocity</td>
<td>↓</td>
</tr>
<tr>
<td>Availability</td>
<td>↑</td>
</tr>
<tr>
<td>Overall Equipment Effectiveness</td>
<td>↑</td>
</tr>
<tr>
<td>Qualification level</td>
<td>↓</td>
</tr>
<tr>
<td>Motivation</td>
<td>↓</td>
</tr>
<tr>
<td>Experience</td>
<td>↓</td>
</tr>
<tr>
<td>Buffer capacity</td>
<td>↓</td>
</tr>
<tr>
<td>Product variety</td>
<td>↑</td>
</tr>
<tr>
<td>Output</td>
<td>↓</td>
</tr>
</tbody>
</table>

As a first step, influence factors on cycle times that restrict planning and adapting cycle times must be detected. Restrictions characterize factors which limit modification boundaries when adapting cycle times. Restrictions can be subdivided into technical, personal and organizational influences.
In the following, several parameters are described and qualitative impacts on cycle times are derived (see table 1). In this context, rising cycle time means slowing down the system while decreasing cycle time means fastening the AL.

### 3.1.1. Technical restrictions

Technical restrictions on cycle times mainly consist of mechanical equipment components of the AL. Subsequently, the authors focus on three restrictions: conveying velocity, availability and overall equipment effectiveness (OEE).

The conveying velocity represents one of the most decisive factors on cycle times within an AL. Assuming a constant technical length of a work station, conveying velocity directly influences cycle times. Task times within continuous material flow systems are limited by the velocity of the conveyor belt [11]. The upper and lower border of cycle time can directly be deduced from the maximum and minimum speed of the conveyor belt. The maximum conveying velocity is predominantly defined by technical or security-relevant design factors. If a conveyor belt features a minimum speed due to mechanical parameters, the maximum cycle time can be calculated. As a result, if it is possible to change conveying velocity, it will define a technically limited range of possible cycle times. E.g. when assuming a constant station length and possible line balancing, cycle times can be bisected with doubled conveying velocity.

Availability represents a major quality parameter of AL performance and directly affects cycle time. It is defined as the probability to come across a system at a given point of time in an operational status [12]. It is calculated as follows [12]:

$$A = \frac{MTBF}{MTBF + MTTR}$$

Hence, to raise an AL’s availability (A), mean time between failures (MTBF) must be increased as well as mean time to repair (MTTR) should be decreased. With a rising availability, an assembly system longer run productively. To realize comparable productivity, cycle times can be raised. Because of fixed minimal task times, rising availability can only be used for increasing cycle time in a limited way [13].

The definition of cycle time is narrowly connected to OEE which allows assessment of a systems’ productivity. Waste like idle time or setting time negatively influences OEE. It is calculated by multiplying the performance ratio, capacity effectiveness and quality rate [13]. Cycle time can be increased when OEE is enhanced at a stable systems’ output.

\[
\text{OEE} = \frac{\text{Performance ratio}}{\text{Capacity effectiveness}} \times \text{Quality rate}
\]

### 3.1.2. Personal restrictions

The individual constitution of assembly workers influences the way, an operator copes with work load. Hence, while quantifying task times, parameters like qualification level, motivation and experience play a major role. In the following, these restrictions and their qualitative impact on cycle time are depicted.

The qualification level is defined as a specific capacity to act and can be fragmented into professional competence, personality competence, methodical competence and social competence [14]. Considering AL operators, the focus lies on the first two of this listing. On the one hand, professional competence is decisive for “daily business” within the AL. Know-how (e.g. work sequence of single tasks) has to be learned previously and applied while performing at the work station. On the other hand, personality competence is defined by operational readiness, self-discipline and frustration tolerance. Superior qualification leads to a higher performance level and, subsequently, to a higher quantity of produced goods. Producing a comparable amount, cycle time can be reduced.

Motivation can be described as the personal attitude, which directly affects one’s willingness to perform [14]. Especially, intensity and endurance of human behavior are of interest for defining cycle time. These factors border the range of interference. Hence, more motivated workers can increase their performance in the long run. This leads to lessened times needed to execute the same task. Consequently, cycle time can be decreased.

Doing something for the first time means investing more effort compared to routine activities. Repeted task executions lead to decreasing task times down to an individual minimum. This circumstance can be explained by experience. A distinctive division of work allows the utilization of the so called effect of experience [15]. With escalating the cumulative production volume, single piece expenses constantly decline. Repeating tasks cause learning effects based on practice and routine. A shorter task time leads to a decreasing cycle time without suffering quality losses.

### 3.1.3. Organizational restrictions

Limitations due to processes within the AL, supporting events or previously configured parameters can be summarized with organizational restrictions. In this context, predetermined amounts, e.g. planned output or number of operators, allegorize quantifiable restrictions. In the following, buffer capacity, product variety and output are focused.

Buffer capacity pictures a technical arrangement between adjoining work stations and is responsible for their decoupling [11]. Fluctuations, malfunctions and breakdowns can lead to downtimes of the whole system. To enable the previously planned output, more capacious buffers ensure a running system. Hoppers in between the line can be seen as operator buffers. They help to fix bottleneck-capacities and repair smaller interruptions. As a result, major buffer capacity enables decreasing cycle time.

A rising number of different product types leads to dispersing packages of task time. Product variety causes losses (e.g. idle time) due to model-mix within a multi variant AL. If a high number of capacious task times converge, planned execution times are naturally high. Such an extreme case has to be considered while planning cycle time. Hence, a high product variety leads to a rising cycle time.

The output of a system can be defined as the quantity of produced goods that meet quality requirements. It helps defining basic limitation for cycle time. Assuming equal task times, cycle time has to decrease with a rising output. This reciprocal proportionality early finds its borders. The minimum cycle time is limited by a huge number of restrictions while the maximum output has defined upper limits.
The analyst must know upper and lower cycle time limits prior to solving the ALBP [6]. The authors face this problem by analyzing influence factors on cycle time. By evaluating those restrictions in a qualitative way, every factor represents a trend (rising or decreasing cycle time). These single trend-lines lead to the definition of cycle time ranges at which an AL can be operated (qualitative expansion of cycle time range see fig. 3). The dashed lines represent upper and lower boundaries to the cycle time range.

Figure 3. Expansion of cycle time range

After expanding cycle time ranges, AL’s performance requirements can be adapted to an individual human work capacity. The black line shows different cycle times within one shift. The dark grey colored fields represent the modification from the previously planned average cycle time. This adaption approximates human pre-conditions. This can lead to decreasing stress and strain resulting from high pressure deadlines within cycle dependent ALs.

4. Summary and outlook

Current production systems run inflexible processing times due to temporally restricted synchronized ALs. On the other hand, human work capacities are very variable. This leads to an individually fluctuating working speed and subsequently to numerous times of high exposure within one shift. The main objective is a temporal individualization of assembly system performance requirements. This gives the opportunity to adapt processing times to individual work capacities. In this context, the approach allows to reactive leveling of workloads. The conceptual framework presented in this paper focuses on an identification of influence factors on cycle time for expanding cycle time range.

The next research steps therefore include an approach to quantify identified qualitative restrictions. Subsequently, with evaluating influence factors on cycle time, its range can be exactly specified.

Acknowledgements

The project “BioTakt – Adaption of Production Planning and Control to Human Circadian Rhythms” is funded by the Bavarian research trust. Furthermore, support from our industrial partners serves as the basement for research activities in this field.

References

Automatic assembly path planning for wiring harness installations

T. Hermansson¹, R. Bohlin¹, J. S. Carlson¹, R. Söderberg²
¹Fraunhofer-Chalmers Centre, Chalmers Science Park, SE-412 88 Göteborg, Sweden
²Product and Production Development, Chalmers University of Technology, SE-412 96 Göteborg, Sweden

Abstract: This paper presents a novel method for automatically planning and finding a smooth and collision-free mounting of connectors in a wiring harness installation. The assembly of electric cables and wiring harnesses is hard due to its concealed routing, multiple branching points, weights and the flexibility in the material. To overcome this challenge, we propose a path planning algorithm that operates in the following steps: constraint relaxation, handle path planning, unfolding and path smoothing. The method has been implemented and successfully applied to an industrial test case.

Keywords: Assembly, Cosserat rods, Path planning for deformable objects, Wiring harness

1. Introduction

The automotive industry of today is focusing on electrified and hybrid solutions, where both conventional combustion engines and battery supplied electrical engines need to fit in an already densely packed vehicle. The placement of each component must be evaluated with respect to geometrical interference with other disciplines and their components. Also, the assembly aspect must be considered early during conceptual design with respect to feasibility and ergonomics. In particular, the assembly of electric cables and wiring harnesses is hard due to its concealed routing, multiple branching points, weights and the flexibility in the material.

Automatic path planning for deformable objects is widely acknowledged as a very difficult problem. It involves both efficient simulation of elastic materials, fast collision-checking of non-rigid objects and path planning that handles and utilizes the flexibility in the material. The task is to find a collision-free transformation of a set of manipulation constraints. Not only does the infamous curse of dimensionality make an early appearance in the number of constraints – a transformation history is embedded in each state resulting in a many-to-one identification between manipulation constraints and the topology of the scene. For example, a rope held in the rope ends could have a different number of knots or be tangled up around a static geometry, all states corresponding to the same rope end locations.

Figure 1: A wiring harness.

1.1. Related work

For a comprehensive introduction to the field of path planning, the reader is encouraged to read [1], [2] and [3]. Due to the complexity of the problem (in fact, it is proven to be PSPACE-hard for polyhedral objects with polyhedral obstacles, see [4]), completeness has been traded for speed and simplicity with various sampling based strategies, see [5]. Examples of such methods are the Probabilistic Roadmap Planner (PRM), see [6] and [7] and Rapidly-Exploring Random Trees (RRT), see [8]. High-dimensional path planning has been addressed in [9]. Simulation of deformable one-dimensional objects has been carefully studied over the years; see for example [10], [11] and [12]. There has been limited success in developing planners for deformable objects. Issues that arise when planning for flexible parts were highlighted in [13] and in [14] path planning techniques were applied to untangle mathematical knots. [15] proposed a planner combining a sampling-based roadmap method with minimal energy curves and [16] suggested methods for deformable robots. [17] introduced a motion planner for robots to obtain different topological states of deformable linear objects and [18] found motions with contact response to completely deformable surroundings.

1.2. Scope of the paper

This paper presents a novel method for automatically planning and finding a smooth and collision-free installation of a wiring harness. The method is not only restricted to wiring harness but can be applied to any tree-like structure of slender one-dimensional objects. In this paper, such a system is modelled with Cosserat rods to ensure a geometrically exact and at the same time computationally efficient simulation. We propose a path planning algorithm that operates in the following steps: constraint relaxation, handle path planning, unfolding and path smoothing.

The paper is structured as follows. In section 2 we give a formal statement of the path planning problem we aim to solve. Section 3 provides a brief description of the rod model that we use in our simulation. The proposed method is outlined in Section 4 and in Section 5 the method is applied to an industrial scenario. Finally, results are concluded in Section 6.

2. Problem statement

Throughout this paper a harness refers to a system of flexible cables connected without loops (see Figure 1). Furthermore, we only treat quasi-static cable deformations, neglecting inertial effects in the cable when subject to manipulations. A set of manipulation handles constrains the harness in some way and are physically manifested as grip
points, rigid connectors or cable clips clamped to the harness. A harness configuration \( q \) is a harness in mechanical equilibrium, satisfying the manipulation handle constraints position and orientations \( \{ c_i \} \). A manipulation of the harness is then a synchronization of prescribed rigid motions parameterized by \( t \) for each manipulation handle \( t \rightarrow (c_i(t)) \).

Given an initial stress-free harness configuration \( q_{\text{init}} \) originating from a 2D build board layout and at a safe distance from the surrounding \( \mathcal{S} \), the task is to find a manipulation that transforms the harness into a given target configuration \( q_{\text{target}} \) known from the design without colliding with any obstacles, i.e. intersecting with \( \mathcal{S} \). Note that what makes this hard is that there could be a many-to-one correspondence between cable configurations and manipulation constraints – a transformation history is embedded in each state. This means that even if we find a collision-free harness manipulation that satisfies the target constraints, the harness configuration we arrive at might not be topologically consistent with the designed configuration \( q_{\text{target}} \). Our method ensures that the correct state is entered with additional manipulation handles (see Section 4.4).

3. The simulation model

To allow for a proper coupling with a path planning algorithm, our cable simulation model must both account for large deformations and at the same time be computationally efficient. Geometrically exact Cosserat rods (see [12]) are gaining widespread popularity for exactly those reasons.

3.1. Kinematics

A cable can express large overall deformations although locally the stresses and strains remain small. Under the assumption that the cable cross section always remains planar and rigid the kinematics of an anisotropic cable are captured in a framed curve. An extensible Cosserat rod of undeformed length \( L \) is a framed curve parameterized by arc length \( s \),

\[
[0, L] \ni s \mapsto (\varphi(s), R(s)) \in SE(3) = \mathbb{R}^3 \times SO(3). \tag{1}
\]

\( \varphi \) is the center curve of the cable piercing the centroids of the cross section along the rod and \( R(s) = (d_1, d_2, d_3) \) describes the cross section orientation; \( d_1 \) and \( d_2 \) are directors in the cross section plane and \( d_3 = d_1 \times d_2 \) is the director orthogonal to the cable cross section (see Figure 2).

Boundary conditions are posed by the manipulation handles together with the branch break-outs and are realized by elimination of degrees of freedom in the rod models. At an arc length position \( s \), a rod can be constrained either to a point \( (\varphi(s) = c_p) \), to a point and a material direction \( (d_3(s) = c_3) \) or to a point and a material frame \( (R(s) = c_R) \).

3.2. Static mechanical equilibrium

To find the static mechanical equilibrium one seeks a stationary point to the potential energy functional \( W \), of which the elastic part is composed of quadratic forms in terms of frame invariant strain measures,

\[
w^T(s) = \frac{1}{2} \Omega(s)^T K^e \Omega(s), \tag{2a}
\]

\[
w^\Omega(s) = \frac{1}{2} \Omega(s)^T K^\Omega \Omega(s), \tag{2b}
\]

for stiffness tensors \( K^e \) and \( K^\Omega \). The stretching/shearing strain vector \( \Gamma \) and curvature/torsion strain vector \( \Omega \) in material coordinates read \( \Gamma(s) = R(s)^T \partial_s \varphi(s) - e_3 \) and \( \Omega(s) = R(s)^T \partial_s R(s) \). The (static) mechanical equilibrium for a single cable in the presence of a gravitational force field \( g \) is then found as a minimizer to the total potential energy;

\[
\min_{w} W, \tag{3a}
\]

\[
W = \int_{s=0}^{L} \{ w^T(s) + w^\Omega(s) - K_p(g, \varphi(s)) \} ds. \tag{3b}
\]

An adaptive finite difference discretization of a Cosserat rod coupled with an efficient Quasi-Newton method to minimize the energy (3b) together with the complete set of boundary conditions forms the foundation of our cable simulation.

3.3. Contact forces

To handle non-frictional contact forces, a repelling potential energy density term \( w_{\mathcal{C}} \) as a function of the penetration depth \( d \) is added to \( W \). Hertz contact theory [19] suggests a potential on the form \( \bar{w}_d(s) = K^c d(s)^2 \). The energy minimization problem is then

\[
\min_{W} + \int_{s=0}^{L} \bar{w}_d(s) ds. \tag{4}
\]

4. The proposed method

To avoid the difficulties associated with high-dimensional path planning we propose a simple and more playful approach: Given a harness mounted at its target configuration \( q_{\text{target}} \) (see Figure 3a), imagine that we first release all manipulation handles and then pull out the harness from the narrow areas with a certain minimal clearance and unfold it where the free space is unrestricted. By tracing all (free) handle movements during this process we construct a harness manipulation where the individual handle paths are guaranteed to be collision-free. The manipulation is then reversed and locally smoothed. As a final step, contacts still present during the manipulation (if any) are resolved by attaching a set of supplementary handles.

4.1. Constraint relaxation

First, we identify a centrally located point \(^2\) in the harness at \( q_{\text{target}} \) and attach a master handle \( p_H \) to it by introducing an additional point clip. We then relax all other manipulation handle constraints. The harness will immediately attain a new configuration held in balance by an applied force at \( p_H \) and external contact forces from \( \mathcal{S} \) (see Figure 3b). During this energy minimization, which for a Quasi-Newton method is composed of a series of minimizing steps in descent search directions, we carefully compose a manipulation by storing the trajectories traced out by the free-hanging handles. These are certainly collision-free due to contact forces but they do not necessarily correspond to energy minima of the harness (save for the last one); when combined into a quasi-static manipulation of the harness a slightly different deformation can be attained. What matters is though the fact that when

\(^2\) A central point can for example be a main branching point or the centre node of the graph corresponding to the complete harness system.
reversing the manipulation we still trace the same energy minimum and arrive at the configuration \( q_{\text{target}} \).

To allocate clearance for the subsequent smoothing steps, we keep the handle trajectories further away from obstacles by turning up the contact potential field to \( \overline{w}_{\delta+d} \) for some parameter \( \delta > 0 \). Additionally, the gravitational field is removed, since it highly influences the elastic response in a harness held at \( p_H \) only.

4.2. Handle path planning

At this point we have a harness configuration held with the master handle \( p_H \). The next step is to find a collision-free path for a ball \( B_H(r) \) around \( p_H \) out from zones of small clearance (see Figure 3c). For a large enough ball radius \( r \), the tree structure of the harness will ensure that the entire harness passively will follow \( p_H \) due to internal material forces and moments. We apply a two-step path planning procedure in \( \mathbb{R}^3 \).

First, we conduct a local search in order to maximize clearance until \( B_H \) can fit without intersecting with \( S \). We then employ a standard A* search finding the shortest collision-free path for the ball on a deterministically sampled grid in \( \mathbb{R}^3 \). As in 4.1, the trajectories of the free handles are stored and added to the solution.

4.3. Unfolding

Close to the configuration \( q_{\text{init}} \) the harness is ultimately unfolded. Even though the untying of knots is a delicate research problem in itself (see [14]), we assume that our harness can be untangled with a simple stiffening procedure where we gradually increase the stiffness matrices \( K^u \) and \( K^I \) of the cable segments from the handle centre point and out. The motivation to settle with this simple step is that we do not model friction, \( q_{\text{init}} \) is far away from \( p_H \) and the harness has a tree structure. As before, the trajectories of the free handles are stored and added to the solution.

4.4. Path smoothing

In the final step, we reverse the solution so that we now have a manipulation starting in \( q_{\text{init}} \) and ending at \( q_{\text{target}} \). Also, the gravity field is turned on and the master handle \( p_H \) is removed. The quasi-static deformations and contact forces in the previous operations have most likely generated handle paths of discontinuous curvature and that are close to \( S \). We therefore employ a simple smoothing procedure where we locally adjust each handle path by reducing it in a way so that the clearance is increased and the curvature is smoothed. The paths are kept synchronized by velocity tuning.

Finally, when applying the smoothed manipulation to the harness with gravitational influence, contacts with the surrounding \( S \) could occur. These are resolved by detecting the set of contact points on the harness and attach supplementary handles. Assume for example that the first contact occurs at arc length \( s \) on a cable. To obtain a suitable collision-free trajectory, the unprocessed manipulations from the steps in 4.1 - 4.3 can be rerun (or precomputed for a number of distinct cable points evenly distributed along the cable) and provide a collision-free trajectory for the cable point of interest. After having added a sufficient number of supplementary handles, we arrive at a collision-free manipulation of a harness. Similarly, the smoothing operation and the addition of a gravity field could cause the attained target configuration after manipulation to be not topologically consistent with the intended configuration \( q_{\text{target}} \). This is also resolved by adding supplementary handles.

5. Industrial test case

The method has been implemented in Industrial Path Solutions [20], a software for virtual assembly verification and optimization. The method was applied to the industrial test case illustrated in Figure 4.

A wiring harness of length 400 mm and radius 12 mm with two break-outs is to be connected at the inside of the engine compartment of a car. We use our planning method in order to verify that there is a collision-free path connecting the harness at the desired design configuration. The detailed surrounding contains components that could cause interference when manipulating the harness in the narrow areas. As shown in the picture, our planner finds a feasible manipulation for a contact potential parameter \( \delta \) set to 10 mm and a master handle ball radius \( r \) set to 20 mm without using any supplementary handles in the final smoothing step. The total computation time of the planning algorithm was 10 minutes.
6. Conclusions and remarks
This paper has introduced a promising method for finding collision-free manipulations of wiring harness from an initial to a target configuration. The method avoids the problems associated with high-dimensional path planning and utilizes internal forces and moments in the harness as the driving attractor for generating collision-free manipulator handle trajectories. One should also note that the method is not dependent on the choice of underlying simulation model. The method was successfully applied to an industrial test case of high resolution with good computation speed.

7. Acknowledgements
This work was supported by VINNOVA’s FFI Sustainable Production Technology program and is part of the Sustainable Production Initiative and the Production Area of Advance at Chalmers University of Technology.

8. References
Abstract: The battery system design is a cross-sectional issue that relates to natural sciences as well as engineering. This collaboration between various fields of research and the number of components defining the battery system can lead to difficulties during the development process. State of the art methodological approaches for electric vehicle battery systems can be used during the conceptual phase of the development process to shorten development time. This automation could be for example CAD-based in order to process the geometrical or mass specific characteristics of the components of a battery system. A major drawback of this approach is to leave out requirements of the following manufacturing and assembly processes.

The approach described within this paper is associated with the methodology of simultaneous engineering as well as design for assembly. The objective is to shrink the overall development and manufacturing time of a battery system through parallelization of engineering design (i.e. the conceptual development of the battery system) and manufacturing (i.e. the assembly process design).

To solve the problem of parallelization, this work focuses upon the analyses of assembly process characteristics. In addition to these boundary conditions referring to the battery system design are taken into account. These two sets of boundary conditions are processed in a software-based tool in order to give recommendations for the battery system design and/or the assembly line design.

A starting point for this approach is done by an analysis of the battery system. This is done by splitting it into three (sub-) systems which are at the cell, module and system levels. The next step after the completion of the requirement analysis for battery systems will be the analyses of existing assembly strategies, handling and gripping technologies and the connection of derived parameters via mathematical functions.

Keywords: Assembly, Conceptual Design, Battery System

1. Introduction to DFA for battery systems

This paper describes a CAD based design for assembly approach fitted for battery systems of electrically propelled vehicles (EV). The scope of this work is based on the methodological approach of design for assembly (DFA) as well as the integrated process development being a part of simultaneous engineering. This paper can be seen as a review of existing assembly strategy solutions and builds a foundation for the incipient stages of research.

The analysis of an assembly process and the optimization regarding cost, time and quality is a cross-sectional procedure. In order to perform a sufficient optimization, knowledge of the assembly process, the product to be assembled and the assembly line have to be taken into account [1].

The following chapter is subdivided into two different sections. The first deals with the analysis of the design of state of the art battery systems for electric vehicles to identify critical components and derive specific characteristics. The second half compromises a review of existing approaches for DFA. Thereafter a methodology is explained how to extend and combine state of the art tools with new approaches to create a suitable software solution for the design of battery system assembly.

1.1. Battery system design for state of the art electric vehicles

The commercially available electric vehicle (EV) is no longer a vision of original equipment manufacturers (OEM), but is slowly becoming a reality. Over the last 100 years, several waves of innovation regarding the development of electric vehicles have been evident. In [2] a phase model for the evolution of the electric car in comparison with the gasoline powered vehicle is depicted. This shows that electrically propelled vehicles have been the focus of research for engineers and scientists several times.

One of the bottlenecks hindering EVs from becoming a widespread means of transportation has been the lack of an acceptable energy source. A major breakthrough was achieved by Sony with their secondary lithium ion cells which through mass production came at a modest price [1][3]. Their extensive use in laptops and other mobile devices has paved the way for the application as an energy source in electric vehicles. This was successfully demonstrated by Tesla Motors in 2007 with the introduction of the Tesla Roadster as described by Gene Berdichevsky et al in [4] and H. Wallentowitz in [5]. In contrast to Nickel-metal hydride based cells which are common in hybrid electric vehicles like the Toyota Prius, the number and topology of sensors/components for lithium ion batteries is much higher [6]. In addition to the amount of cells needed for power supply, this leads to a challenge for the packaging process and inherently the assembly of the complete system.

The following chapter gives a short review of parts and components of the battery system of an EV and is based upon reviews like Li, Wang et al in [7]. From this a sufficient requirement analysis for the DFA approach is done in order to identify components assembled and topologies to be modified.

1.1.1 Battery cells and their shapes

According to the German Standardization Roadmap, the standardization of battery cell dimensions would restrict the overall vehicle design and thus should not be standardized [8]. The absence of standards and guidelines opens the doors to a wide range of cells for automotive applications. These include a wide spectrum of shapes, sizes and characteristics. Three cell shapes have immerged and taken on market predominance, which are pouch cells, cylindrical cells and prismatic hard-case cells. The
pouch type is a stack or winding of electrodes fitted inside a flexible housing, typically made from an aluminum laminated film deep drawn or thermoformed to a case [9]. The prismatic hard case cell uses a solid housing typically made from polymers, steel or aluminum [10]. Aluminum or steel cases are the typical housing type of high energy batteries for EV applications for prismatic cells as well as for cylindrical cells. Examples of these cell types can be seen in figure 1.

**Figure 1. Battery cell types [10][11][13]**

In commercial applications such as laptops and mobile phones small cylindrical cells are used. These can also be found within the battery system of the Tesla Roadster [4]. Unlike Tesla large OEMs like Mitsubishi or Nissan focus upon high energy cells. In this case Mitsubishi uses prismatic hard case cells within the battery system of the iMiEV, Nissan uses pouch cells within the battery system of the Leaf [14][15]. Making use of high energy cells leads to a reduction in the total amount of cells and a simplification of the connection configuration within the battery system [16]. In addition to the cells, various peripheral components need to be fitted into the module case.

### 1.1.2 Modules and cell periphery

Each EV that is currently available is compromised of a specific number of cells to provide the required amount of energy for an adequate driving range. In order to handle these designs during the production and assembly process, it is common to connect several cells in a serial/ parallel configuration and build up modules. This includes the corresponding peripheral components that are necessary for the operation of the system [16]. According to the International Electrotechnical Commission (IEC) voltages up to 60 V DC are not lethal and no specific protection is needed to handle these modules. This leads to a simplification of safety related production processes and thus has made 60 V DC a common battery module voltage [17]. Those modules are then used to complete the battery system.

The battery management system (BMS) slave unit senses the current flow, the voltage and the temperature of the cells within the modules in order to prevent the cells from damage, prolong the battery life and keep the cells in the best possible operating condition [18]. Different configurations for the placement of hardware to control the cells are existent, and according to [19] there are 4 common topologies. The number of BMS-slaves depends upon the management capabilities of the hardware. Apart from the BMS-slave, fuses and switches are common to ensure electronic safety protection within the module.

In order to keep cells at an optimal operating temperature, measured by sensors and processed by the BMS units, the cells have to be cooled or heated [20]. For the application in electric vehicles several methods for the thermal management of cells have been introduced and can be found inside state of the art EV’s. Simple ways of keeping cells at operating temperature is to use air cooling/ heating. An example for this type of thermal conditioning is applied in the Mitsubishi iMiEV [21]. If a liquid cooling medium is used pumps and cooling plates are required. A third way of cooling is possible by using solid state coolers. This could be a phase change material (PCM) that is adjusted to the thermal operating conditions of the cell [20].

To ensure that the battery cells are fixed in a requested configuration as well as keeping the components for thermal management, control and safety in place, various structural fixtures exist. These are typically made from metals such as aluminium or steel but could also be produced from polymer cases or fiber reinforced polymer housings. The latter are typically used for the pouch cell battery type as seen in [22][23][24]. Figure 2 shows different structures for the mechanical safety of the battery cells as well as the fixture during production and use of the battery system.

**Figure 2. Module structures [25][26]**

### 1.1.3 Battery system assembly

The complete set up of the battery system is mostly done via positioning and fixtureing the modules in a specifically designed housing. The adoption of this housing to the free space within the car body can lead to complex structures with extensive surface areas. An example is Mitsubishi’s iMiEV/where its battery system housing is made of a glass fiber reinforced plastic which covers approximately one third of the floor of the car [27]. In addition to the fixture of the modules within the system, cooling elements, connectors and electronics need to be installed to complete the system.

### 1.3. DFA tools

Various types of design for assembly tools have been developed and introduced. Reviews and evaluations have been done for example in [28] and in [29]. The following DFA methodologies were used as a starting point for the approach presented within this paper.

The IDAES system introduced by Sturges and Kilani in [30] incorporates a software/CAD based knowledge system to identify assembly critical components of a specified product during the conceptual phase of the development. This is done by separating between assembly process-level, system-level and component-level, and the sharing of knowledge between these levels. The evaluation of parts and products leads to recommended modifications of the product structure in order to reduce assembly time. Another interesting approach is presented by Sha Li, Hui Wang, S. Jack Hu, Yhu-Tin Lin and Jeffrey Abell in [31] that incorporates a methodology integrated into a software based solution. This solution aims at an automated process to find system configurations and equipment selections for the assembly of lithium ion battery cells.

Santochi and Dini introduced a method of assembly sequence planning that is based upon the geometrical structure of the product represented through a CAD software package. This could be used as a flexible manufacturing system [32].
Hsu, Lee and Su introduced an approach that uses information from the assembly planning, relinks this information into a software based system and gains information about potential redesigns of the product regarding assembly friendly design. Furthermore intelligent CAD-DFA solutions have been introduced by Pradeep K. Khosla, Raju Mattikali in [34] and [35].

Apart from various approaches introduced by researchers, commercial tools are available and field tested as well. These are for example the workstation based systems in Pro/Engineer and Unigraffics as well as the DFMA tools by Boothroyd and Dewhurst [36]. In [37] a method of integrating skeleton entities into a CAD based system is presented in order to process engineering knowledge for the designers in an early, conceptual stage of the product design.

The approach presented in this paper compromises a partially CAD based approach that uses geometrical as well as electrical characteristics in different product levels to create a model that can be visualized, modified and optimized for assembly.

2. Approach for a conceptual DFA method for electric vehicle battery systems

The variety in cell shapes, cell types and cell configurations in addition to the complex and diverse structures of the free space within an EV leads to time consuming processes for the system design. This coincidence shows that it is useful to incorporate an automatic/semi-automatic process that links the assembly process with the product design and vice versa. This simultaneous engineering approach leads to a reduction of assembly time and helps engineers to design the battery system. This chapter deals with the process of identifying parameter sets of the battery system. These sets of parameters are then merged to mathematical expressions and integrated into a software based system in order to create a modular and suitable solution that is adoptable to different boundary conditions (e.g. battery cell, cooling solution, housing space). Thereafter a way for the integration into a CAD tool to visualize the result is presented. The assembly optimization is done via a rule/guideline data comparison process and the usage of available DFA systems.

2.1. Parametric analyses of the battery system

The main task for the presented approach is the identification and mathematical connection of necessary parameters. These are based upon the product, i.e. the battery system, as well as the assembly process and used to calculate dependencies ultimately allowing the solution to be visualized. The following chapter deals with the description of a requirement analysis in order to collect these parameters.

At first it is necessary to identify parameters that are related to the battery system. In order to keep the extent of this paper at an appropriate level, the parameter analysis is shown solely for the cell-level which is the fundament for the constitutive analysis.

As mentioned above cell shapes and therefore mechanical as well as electrical characteristics are diverse. The three cell types differ in shape and size and are therefore described through their main dimensions (length, width, height, diameter/radius). Additional parameters are needed if a description and visualization of cell specific features like the sealed frame of a pouch cell is necessary. Apart from the basic volume defining parameters of the cell housing, a set for the description of terminals is introduced. In compliance to the diversity of cell shapes the shape and dimensions of terminals differ as well but are in focus for standardization [8]. Typical representatives are plain terminals at both ends of the cell for cylindrical cells, flat foil-type terminals for the pouch cell and screwed terminals for high energy hard case cells. The following figure shows some examples of different terminal types [38][39].

Alongside the geometrical characteristics mentioned above, the gravimetric characteristics of the cells have to be identified and considered during development. These are predominantly defined by the housing-type and the active materials within the cell. The housing of the hard case prismatic cells is either a polymer or metal based case. Metals used are steel and aluminium. The latter is used to decrease the mass of the cell in order to obtain a higher gravimetric energy density. To increase the stiffness of a steel case, the aluminium wall thickness is increased which leads to slightly bigger cells [10]. Polymer or fiber reinforced polymers are also available [11]. The cell mass as well as the volume that is defined by the main dimensions define the volumetric and gravimetric characteristics of the lithium ion battery and are important for the packaging process.

Eventually the parameters that describe the electric properties of the cell need to be taken into account. The most important are the nominal cell voltage, the charging-end-voltage, the discharge-voltage, the capacity, the C-rates and the internal resistance [40].

The requirement analysis is done for the module level as well as the battery system level. The peripheral components analysed are the cooling systems, the electronic components like the BMS/fuses/switches, the interconnectors as well as possible case/housing structures on module and system level. These have been characterized by its dimensions, material properties and position related to a single cell or pack respectively.

2.2. Software build up and case-study

The implementation of parameters of the cell-, module- and system level into a software based solution and the interconnection of these parameters are presented in the following chapter.

Up to now the created methodology features an input section for the user of the system, a calculation section which is done via visual basic applications (VBA) and a visualisation section which is a CAD system. Considering that the assembly optimization is done in the conceptual phase of the battery system development, the design of cells and peripheral components is kept as simple as possible. The structure of the system can be seen in figure 4.
This structure will be described through a case-study, compromising the integration of a number of cylindrical cells into a specified space and its assembly optimization. The first step is to implement or choose a set of assembly product and process specific data. In this case information about the geometric and electric properties of the cell, the peripheral components, the battery system and the available space within the vehicle is needed. If data is newly created, it can be stored for reuse in spreadsheet format. Apart from precise cell characteristics like dimensions or electrical properties, remarks can be added regarding uncommon characteristics like the terminal shape/orientation. In addition to cell characteristics, the data for the properties of the module housing, the cooling system, the BMS type and the terminal connectors has been implemented. In compliance to the cell characteristics, these sets of parameters can be created directly or chosen out of existing data sets. The former is likely to be common, as these components are engineered for the specific battery system. An example for this could be the cooling plates. These are designed to be suitable for a specific module. The approach uses the main dimensions to optimize the place required by the plates, its position related to the cells and the derived mass. The input of the cell type, cell configuration, cooling type, BMS type, structural components and peripheral components are used to calculate volume, mass, topography and centre of gravity for modules in order to gain first boundary conditions for the handling technology.

The next step is the definition of global boundary conditions of the electric vehicle. These are for example the specified space, the type of peripheral components and the characteristics of the electric motor and power electronics. The available space within the vehicle structure is abstracted to cubic elements for terms of simplicity. These are defined by their three dimensions and six degrees of freedom within the coordinate system of the associated electric vehicle.

Following the input of the described parameter sets a visualized solution for an energy and package optimized battery system is calculated. This is done through an algorithm that uses the cell and peripheral component characteristics to create and compares virtual module configurations within the space of the electric vehicle. In this case study a cylindrical cell is chosen. The algorithm calculates all possible cell topologies based upon parameters like the space between one cell to the other and the angle between the grid lines. The following figure shows an example of the top view of a configuration of cylindrical cells wherein the angle between the grid lines is changed.

![Cylindrical cell configuration](image)

**Figure 5. Cylindrical cell configuration**

The characteristics for peripheral components are used to complete the virtual module. In the case of water cooling, the position of cooling plates is calculated for the best possible overall energy density of the battery system. The algorithm used in this approach systematically varies cell topologies, BMS configurations, cooling peripherals locations, structural attributes and the electrical interconnection. Each virtual module is mirrored within the space for the battery system and stored for reuse in a database. The chosen solution for an energy maximized system is visualised through the connection to the CAD system. The user is now able to use the CAD data for the detailed design. The design for assembly approach reroutes and compares the input of the CAD system to a database that is based upon existing and tested DFA rules and guidelines throughout the design of the battery system. The topology, the number of cells and components and further assembly critical parameters are therefore compared with restrictions defined by the user in order to gain an assembly modified version of the battery system.

### 3. Conclusion and outlook

The paper describes a DFA approach that uses CAD systems, spreadsheets and VBA in order to help the battery system developer to optimize the package and the process planning during the conceptual phase of the design process. The approach begins by naming necessary and possible components that define a battery system for an electric vehicle. In addition to these a short review of existing approaches for DFA especially CAD based DFA is presented. The core of the paper incorporates a description of an analysis of possible parameter sets that can be used to describe, modify and optimize battery systems in an early design phase without limiting the creativity of the engineer. Finally, the design for an assembly approach is shown via a case study using cylindrical cells.

The first step for the completion of the tool is being finalized so that the battery system package can be calculated and visualized. The optimization regarding the orientation of all cells as well as peripheral components is partially working and will take some additional time for completion. A major step will be the implementation of assembly process and technology specific parameter sets to create a knowledge base for the user. Furthermore an interface to commercial DFMA tools is likely to be created in order to gather information about potential assembly times and costs.

Finally the implementation of mass specific and cycle time critical characteristics are going to be linked to the CAD/VBA tool in order to receive information if an existing production technology is able to process a specific number of battery systems.

---

*Figure 4. Structure of the battery system DFA approach*

<table>
<thead>
<tr>
<th>Input Data</th>
<th>Battery System</th>
<th>Electric Vehicle</th>
<th>Assembly Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g. Cell data, thermal configuring, connecting configuration...</td>
<td>e.g. available space, electric driven interfaces...</td>
<td>e.g. alignment, configuration, available technologies (gripping, handling)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Calculation</th>
<th>Design</th>
<th>Assembly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Battery System</td>
<td>e.g. module design, topology of cells, connecting configuration...</td>
<td>DFA e.g. Battery System specific, having standardization and commercial tools</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Output Visualization</th>
<th>Package-optimized Battery System</th>
<th>Assembly-optimized Battery System</th>
</tr>
</thead>
</table>

---

172
References


[15] Kennedy, B., Patterson, D., Camilleri, S., 2000, Use of lithium-ion batteries in electric vehicles Journal of Power Sources, Volume 90, Issue 2, 1 October 2000, Pages 156-162


[33] Kennedy, B., Patterson, D., Camilleri, S., 2000, Use of lithium-ion batteries in electric vehicles Journal of Power Sources, Volume 90, Issue 2, 1 October 2000, Pages 156-162

[34] Pradeep, K., Khosla, R.M., 1989, Determining the assembly sequence from a 3-D model, Volume 20, September, Pages 153-162


Beyond human tetris: simulation-based optimization of personnel assignment planning in sequenced commercial vehicle assembly

L. März\textsuperscript{a,c}, W. Mayrhofer\textsuperscript{a,b}, W. Sihn\textsuperscript{a,b}

\textsuperscript{a}Vienna University of Technology, Inst. of Mgt. Science, Theresianumgasse 27, 1040 Vienna, Austria
\textsuperscript{b}Fraunhofer Austria Research GmbH, Theresianumgasse 7, 1040 Vienna, Austria
\textsuperscript{c}LOM Innovation GmbH & Co. KG, Kemptener Straße 99, 88131 Lindau am Bodensee, Germany

Abstract: This paper presents a simulation-based planning software, that is developed for high complexity production settings and currently undergoes first real-data industry application tests. The software is tested in commercial vehicle assembly, but will be also employable in other industries with a sufficiently high number of different products and variants. The solution enables the forecasting of required personnel and the requirements with regard to temporary additional workers (“floaters”) for every cycle at the production line. To this end, the software periodically collects process times, sequence and feature data for the vehicles. The cumulated workload of required personnel and the requirements with regard to temporary additional workers (“floaters”) for every cycle at the production line is calculated under consideration of flexibilities. A web-based user interface enables to set and change reference data, define process simulations and allows for the compilation of diagrams and charts for the analysis of the results of the different simulation runs. Various filters as to time periods, vehicle types or feature variants allow the planner a target-oriented analysis and provide feedback for improving the tact of the production line. Finally, some qualitative results of a first industry test in commercial vehicle assembly are presented. Those preliminary results give valuable feedback about the usability of existing analytical features and desirable additional features.

Keywords: Production Planning, Productivity, Sequencing, Simulation

1. Variant-rich series production in sequenced assembly lines

The principle of variant-rich series production in sequenced assembly lines is current practice in the automotive industry [1]. Due to high wage and non-wage labour costs, available personnel has to be carefully planned with respect to changing requirements at the assembly line in order to evenly utilize capacity [2]. Hence, mixed-model assembly line production requires careful program and sequence planning. The key requirement is to distribute orders with different work content over the production program in a way, that overload and underutilization are avoided [3, 4, 5].

Present sequence planning systems are able to accommodate such planning objectives, by means of sequencing rules that stem from a multitude of vehicle criteria [6]. Yet, sequencing does not always prevent dynamic load peaks at certain stations, since sequencing rules are not created from a forecast, but are based on experiences from the line. Once an excessive peak load as a result of the sequence of vehicle types and/or configuration reaches a team, this specific overload can be avoided in the future by defining a rule, that will prevent this specific scenario causing the overload [7]. A variety of possible vehicle configurations can result in an overload and it is impossible to exclude all possible scenarios that lead to overload, for unacceptable computation runtime or lack of a solution at all [8]. In addition, increased processing time requirements which become apparent in the course of sequences built in the future are not detected. Reasons for this are rare vehicle sequences or shifts in the proportion of certain product types or configuration variants.

Hence, a preview of processing time requirements per cycle and team is of high value. A forecast of the impact on employee workload requires variables representing additional degrees of freedom stemming from the organization of labor within each team (i.e. negative or positive drift). Organizational measures such as “roll-over teams” with different characteristics regarding the execution of assembly tasks in comparison to station-bound teams, should be part of the analysis. Further, the extent of the individual processes that form a task affects the tact of the line.

It is daily practice that process time requirements exceed the capacity of the teams. The simulation is aimed at accurately forecasting process time requirements and resulting workloads of the assembly. This is the foundation of exact personal assignment planning and especially alerts the planner if provisions with respect to wo/manpower (i.e. floaters) are necessary. Hence, demand for assembly personnel is known even before assembly starts. Necessary measures can be taken to fulfill those capacity requirements. Its application is geared to personnel assignment planning in daily operations (shift) as well as in medium term (weekly preview) and strategic planning.

With the planning assistant presented in this paper, transparency in the analysis and evaluation of planning data increases (Figure 1). Subsequently, with the application of mathematical optimization methods a comprehensive decision support system will be developed.

However, the application discussed in this paper so far does not optimize the planning results in a mathematical sense, but supports the finding of an “optimized” solution to the planning problem. This is achieved by making complex system behavior visible and further allowing a step by step optimization of to date mathematically unsolvable planning problems.
2. Objectives and input data for simulation

2.1. Objective

The objective of the development and implementation of the web-based simulation software was to allow continuous use in planning of operative personnel of assembly lines and to give planners a tool for line balancing [1, 9, 10, 11]. The following requirements were central to the design of the support system:

- It is developed for the analysis, evaluation and design of the personnel assignment of sequenced production lines.
- The simulation of staff workloads is based on operational process data and the sequence.
- The area of application shall range from daily operations (use of floaters) via medium-term planning (personnel assignment) to strategic aspects (line balancing).
- Each planner within the network shall, by means of a simple browser, be able to edit the data, simulate the sequence and to analyze the results.
- The functionalities for the evaluation of results will need to provide extensive capabilities and views for analysis, preparation, selection and export of diagrams and tables.

In a nutshell, the system will assist the planner in smoothing the personnel capacity demand curve and identify possible starting points. For this reason, the control variables with the greatest leverage effect on smoothing work content (capacity demand for work content) at the stations have to be identified.

2.2. Procedural aspects of a simulation application

A simulation application can be defined into the phases: model, scenario, calculation and evaluation. The model consists of master data and operational data (sequence, process and attribute data). The master and sequence data can be edited in the application and stored as data records. By selecting master and operational data as well as the depicted time period, a simulation scenario is defined. The calculation of personnel assignment will be performed by the simulation. Subsequently, different analysis functions are available for the interpretation of the results.

2.3. Model data

The master data set contains all elements needed to describe a production line. It represents the production system and all physical and organizational elements of a production line.

The operational data is split up in sequence, process time, and attribute data. The sequence represents the system load and depicts the order (vehicle) sequence. The sequence list contains product name, cycle (operating number within the sequence) and date of the laydown at the last station of the production line. The sequence is updated daily and stored in a separate sequence data set. Sequences can be duplicated, edited and created anew.

The process time data includes all process steps required for each vehicle. By creating a link between work plans and assembly positions, the interrelationship between to the affected teams can be established. Process time data is automatically updated daily and cannot be edited in the application.

The attribute data provide information about equipment options of the vehicles. The attribute data for each vehicle are automatically updated daily and with help of the filter function in the analysis of results allow the analysis of the effects of equipment related process-time strains.

2.4. Definition of simulation runs

The definition of the simulation configures a scenario for the simulation of the flow properties of the production line. The application is designed for:

- manually triggered simulation runs and
- automatic simulation runs.

For a manually started simulation run a master data set, a sequence data set and the start and end cycle needs to be defined. For an automatic simulation, the master data set and the cycles to be simulated are to be specified. The sequence data are updated once a day together with the process time and attribute data and are available for simulation afterwards, allowing daily analysis of the current production program, particularly for floating staff deployment. The scenarios are simulated on a standard computer with a valid runtime license of the software simulation SLX.

3. Simulation of the assembly process

3.1. Balancing process time requirements with available capacity

Each task is assigned to a team that is qualified to perform certain tasks. The different teams have a spectrum of tasks often corresponding to a specific technical area (suspension, engine, cabin, etc.) the team is able to perform. Some manufacturers periodically move around their teams along the assembly line for job-enrichment and in order to have a more flexible workforce.

The predefined sequence of orders determines the position of the vehicle in the station at any given time. With the assignment of teams to the stations, the tasks per cycle and team can be identified. The process time requirements are matched with the available capacity of the teams. Should one station be short of staff, the tasks can start in the previous cycle (negative drift) or can be finished in the following cycle (positive drift) [12]. The review of the process time requirements for each vehicle and team will be made before assigning employees. Basically, the following two scenarios can be distinguished (Fig. 2):

- Scenario 1: processing time requirements are less than the capacity of the team.
- Scenario 2: processing time requirements are greater than the available capacity of the team.

![Figure 2. Comparison between processing time and capacity requirements](image)

As long as process time requirements are higher than available capacity (Scenario 2), tasks might be pushed up. Whether this is possible depends on several factors:
• The team is allowed to move up-to what amount this is possible is defined in the table team.
• The workers still have idle capacity, which is evaluated at the simulation run and results from dynamic calculation.

In scenario 1 there is generally no negative drift. However, due to the size of the process modules and the necessity to deploy several workers to the process, it is possible that processing time exceeds the cycle limit. Typically, those time components are comparatively small. After balancing capacity demand with available capacity the simulation runs through a planning and implementation process. The optimization of the resources is supported through shifting of tasks and processes through the planner. The changed configuration subsequently can be evaluated and compared to the results of previous runs.

3.2. Assigning of workers
Employees are scheduled by booking of capacity “pots” per employee and cycle by means of assigning individual tasks [13]. Upcoming tasks are allocated by the following order of priority:
• tasks requiring more than one employee to be performed, sorted by the number of required employees per task
• tasks sorted by their process time in descending order.
Assigning of workers is done by debiting the capacity pots of the workers of the involved teams. To do so, the first worker’s capacity pot is checked for sufficient capacity. If this is the case, the task is allocated to the worker’s capacity pot. This results in (almost) filling up the capacity pot of the first worker before engaging a second worker. Whether the capacity of an employee is completely utilized depends on the process time variables. The smaller the process steps are, the higher is the chance to utilize a workers capacity to a high extent. The selected distribution logic is similar to the approach used for line balancing in planning.

The procedure covers checking capacity over the current cycle and previous or subsequent cycles in case of negative drift or positive drift. If the next worker does not have sufficient capacity, the employee after that will be checked. If there are no workers with free capacity in a team, an employee of the group of floaters, marked in the master data as such, will be summoned. The handling of the distribution of tasks in the floater group works similar as capacity assignment of the worker teams. The floaters have the same limits with regard to negative or positive drift as the teams to which the floaters are assigned. If the capacity of the floater group is exhausted, the superordinate floater group will be engaged. If there is no such floater group registered any more, a virtual floater will be created, who has no restrictions on positive drift. This prevents that operations with too high of a process time due to pre-, cycle and rework capacity cannot be modeled with the existing workers.

3.3. Carrying out of the tasks
The selected employees are assigned to tasks. At the time of the simulation run, the registered teams and floater groups are scanned for idle workers. Once all employees are available for one task, the employees are blocked for the process time. Contrary to task-allocation amid workers in planning, tasks are now distributed to each employee that turns idle, i.e. the first two tasks of a team can’t be allotted to the first worker, since the worker is already busy carrying out the first task. The second task will be executed by the next idle worker. Compared to the original planning this altered allocation of tasks to workers might require a small positive drift. The selected assignment logic is similar to the real allocation of tasks at the line.

In case of increased or decreased processing time requirements cycle limits can be underrun (negative drift) or overrun (positive drift) [10]. The following basic configurations are possible in the case of process times that are less than twice the cycle time (Fig. 3). In the case of tasks exceeding cycle time limit, the time fraction per cycle is calculated and assigned to the respective cycle. The model does not use punishing factors, but attempts a representation of the processes at the line with high validity and all cases below can appear in reality.

4. Evaluation

4.1. Bar and matrix diagrams
Evaluation functionalities include a variety of diagrams and tables providing a variety of choice for analysis of results. Structure and arrangement of evaluation diagrams is oriented in such a way, that workloads (caused by the vehicles) are shown with respect to cycles and organizational units. By determining the organizational unit, capacities are defined and set in relation to processing time requirements per cycle and the division into temporal sections (cycles) is arbitrary and can be reduced or extended by changing line cycle time, thus affecting productivity. Spatial structuring in stations is not evaluated. However, a link is established by the cyclically releasing of the next vehicle into the line at station 1 and the gradual passing along during the next cycles as well as the assignment of the teams to stations. Dependent upon the view of the analysis of workloads, different types of diagrams (Fig. 4) ensue.
The diagrams contain at least one structural component concerning the organization of personnel or cycle time-related.

Table 1 shows the graph types dependent on the selected category (organizational unit and selection of type of diagram).

Table 1 Result chart types

<table>
<thead>
<tr>
<th>organizational unit</th>
<th>selection chart</th>
<th>chart</th>
<th>chart</th>
<th>chart</th>
<th>chart</th>
<th>chart</th>
<th>chart</th>
<th>chart</th>
<th>chart</th>
<th>chart</th>
</tr>
</thead>
<tbody>
<tr>
<td>team</td>
<td>utilization</td>
<td>teams</td>
<td>utilization</td>
<td>team</td>
<td>utilization</td>
<td>teams</td>
<td>utilization</td>
<td>teams</td>
<td>utilization</td>
<td>teams</td>
</tr>
<tr>
<td>all</td>
<td>utilization</td>
<td>teams</td>
<td>aggregation</td>
<td>matrix</td>
<td>over</td>
<td>cycles</td>
<td>product</td>
<td>over</td>
<td>view</td>
<td>matrix</td>
</tr>
</tbody>
</table>

The histograms showing process time quantity distribution and variant spread mainly analyze process time data in relation to the cycle time or the capacity of the team.

4.2. Example: utilization diagram of a team

The chart ‘workload group team’ shows process time requirements and the utilization of a team per cycle (Fig. 6).

![Figure 5. Example graph utilization staff group](image.png)

Each cycle is assigned to a vehicle and the required process time represents the assigned work content, which results for the current vehicle in a cycle. One set of bars visualizes the work content in relation to the capacity of the team. Another set of bars shows the utilization of the team in the cycle for the vehicle. Work content can exceed the 100% level of staff capacity. In this case, the additional work can be visualized as follows:

- negative drift of the team
- positive drift of the floater group

By clicking on a bar, a table below the chart will appear showing each process time data. In addition, filter functions allow the selective choice of results for products, product types, product attributes as well as cycles.

5. Conclusion

The described application is used in a commercial vehicle manufacturing company since mid-May 2011. With a run-time behavior of one second per cycle, simulation runs on the maximum preview horizon of up to 1,500 vehicles are possible in less than half an hour (is equivalent to a 10-day forecast). By means of the web-based application, the planners themselves can define scenarios and analyze the effects by means of analysis charts and filter functions. The integration of the application in the daily operations of the planners and segment leaders primarily assists in personal assignment and planning of floaters. By simulating the daily sequence, down time of floaters can be minimized by exact (to the cycle) scheduling of work assignments. Moreover, a better planning and use of non-productive time of the floaters will be enabled.

In medium-term planning workload matrices allow the assessment of the ‘seriousness’ of the expected program and an accurate prediction of the necessary weekly personnel resources. The goal of the further development of the application is to improve operational and evaluation functionalities. In future, the planning assistant is intended to be tied to an optimization algorithm to enable the management of headcount as a function of the applied sequence.

References


Assembly path planning by distance field based shrinking

S. Björkenstam\textsuperscript{a}, J. Segeborn\textsuperscript{b}, J. S. Carlson\textsuperscript{a}, R. Bohlin\textsuperscript{a}
\textsuperscript{a} Fraunhofer-Chalmers Centre, Chalmers Science Park, SE-412 88 Göteborg, Sweden
\textsuperscript{b} Volvo Car Corporation, Virtual Methods and IT, 405 31 Göteborg, Sweden

Abstract: This paper presents a method to support dynamic packing in cases when no collision-free path can be found. The method, which is primarily based on path planning and shrinking of geometries, suggests a minimal geometry design change that results in a collision-free assembly path. A new method to shrink geometry based on adaptively sampled distance fields is suggested. Compared to previous work, the new method is less sensitive to input mesh quality. The method works directly on the triangle mesh and requires neither tetrahedralization nor calculation of a medial axis. The algorithm paired with a state of the art path planner has been applied to several industrial cases demonstrating that the proposed method can be used effectively for assembly verification during geometry design.

Keywords: Dynamic packing, path planning, shrinking, distance fields.

1. Introduction
Automotive analysis and verification gradually moves from a physical environment to a virtual one. As virtual models of products and processes are developed long before physical prototypes are built, analysis and verification are initiated at earlier stages of the car development projects. Since the costs of solving design related problems rapidly increases as project time progresses, early identification and solving of problems are fundamental prerequisites for cost effectiveness.

Dynamic packing analysis deals with solving geometry conflicts, determining a collision-free assembly path for each product sub assembly. Traditionally done by time consuming manual analysis, using various general CAE-tools, dynamic packing is currently often carried out using automatic path planning.

However, at early project stages, product and process design are of an approximate nature and are subject to frequent design changes and restriction volume conflicts. This impedes the use of automatic path planning, as planning algorithms require the existence of a collision-free start and goal configurations as well as the existence of a collision-free path.

If no collision-free path exists between the given start and goal configurations the object or the surrounding geometry needs to be modified. One way to modify the geometry, without making far-reaching design assumptions is to shrink it. Shrinking can be defined as the process of translating every feature of an object in the normal direction into its interior. The geometry is shrunk to such an extent that path planning can obtain a collision-free path. Letting the original object follow this path will reveal intersecting geometries and hence support minimal design changes for a collision free assembly process.

1.1. Previous Work
Path planning has been a field of research since the seventies. Many different approaches have been suggested. For overviews see, [1] and [2]. In practice, there are two approaches that are especially widespread and used, due to their conceptual simplicity and implementability. The Probabilistic Roadmap Method, PRM was first described by [3], [4] and the Rapidly-Exploring Random Trees Approach; RRT was first described by [5]. Path planning approaches are however primarily aimed at finding a collision-free solution. The quality of the path, in terms of clearance to obstacles, smoothness and path length can be addressed in a post processing step [6]. Various approaches to reduce running times of PRM planners have been suggested. For instance, the Lazy PRM planner suggested by [7] minimizes the number of collision checks made during planning. Dilation of the collision-free sub set of a robot’s configuration space has been suggested to increase visibility in conjunction with narrow space sampling. Dilation of free space by permitting some penetration is treated by [8]. Penetration depth computation are however increasingly difficult with the complexity of geometries, which limits industrial relevance. Penetration depth computations are treated by [9] and [10]. Dilation of free space by shrinking of geometries is suggested in [11] and [12]. Path planning, aided by shrinking of geometries are previously also treated by [13] and [14]. The shrinking method in [12] is based on medial axis computation. The geometry is shrunk, around its medial axis. However, as computation of medial axis is difficult, because of its instability and algebraic complexity [15], industrial applicability is limited. The shrinking method in [11] utilizes a tetrahedralization of the model. The tetrahedralization is used to ensure that the shrunk model is contained within the periphery of the initial model. Shrinking of the model is carried out by translation of its vertices. The distance that the model can be shrunk depends on its triangulation and tetrahedralization. In [16] the authors present a method to shrink geometry based on vertex translation interleaved with retriangulation, we will later compare the proposed method to this one.

1.2. Proposed Method
This paper suggests a method to automatically support a minimal geometry design change that results in a collision free assembly path. The method supports dynamic packing cases when no collision-free path can be found. The principal idea is to shrink geometries, until a general path planner can find a collision-free path. Path clearance is then increased, using local path optimization techniques similar to [6]. As this path is run, using the original geometry, all colliding areas are accumulated and can be presented to the designer to support design changes. The method has the same outline as [16] with the main contribution of a new method to shrink geometry, described in section 2. To demonstrate the usefulness of the method we apply it to a real world problem in section 3.
2. Shrinking Algorithm

When performing shrinking directly on a triangle mesh small errors in the triangle mesh, which is not uncommon when working on real world problems, can severely affect the end result. Also, even on perfectly triangulated objects, changes such as that in topology can be hard to handle. By using an implicit representation of the geometry, independent of the original parameterization, many of these problems can be avoided.

The shrinking method described in this paper uses the signed distance field to represent the geometry. The signed distance field associated with a geometric model is defined as the distance from a given point to the closest feature on the model (Section 2.1). The distance is defined to be negative for points inside the model and positive outside. Shrunken versions of the geometry can then be defined by different isolevels in the distance field. Several techniques have been investigated to solve this nonlinear equation. In this paper we will present our most successful method which is based on adaptive sampling and contouring of the distance function in an octree structure (Section 2.2). It is showed that exploiting the Lipchitz continuity of the distance function makes for a much faster and less memory intensive algorithm than contouring of a general implicit function.

![Figure 1](image)

**Figure 1.** The quadtree, the 2D equivalent of an octree, used for adaptive sampling of the distance function. The signed distance function returns the shortest signed distance to the boundary, i.e. for the point \( p \), \( f(p) = -d \).

2.1. Distance functions

The distance function, \( f: \mathbb{R} \rightarrow \mathbb{R}^3 \), is defined as the shortest distance from a given point to the object. In order to get a volumetric representation of the object we extend the distance function such that points on the interior have a negative sign, one way to do this in practice is to use the angle weighted pseudo-normal of the closest primitive which is proven to point away from the point if and only if the point is an interior point as described in [17]. Once the distance function is constructed the original surface is given by the zero level set, \( \{ p: f(p) = 0 \} \). An offset surface at certain distances to the original mesh can defined as \( \{ p: f(p) = d \} \), where the constant \( d \) is the signed distance of interest.

2.2. Discretization and triangulation

In order to create a triangular approximation to a contour of the distance function we first set out to find a good discrete approximation to the distance function itself. An obvious start is to simply sample the function in a uniform cubic grid and use trilinear interpolation inside the cells to extend the function to the whole domain. A more refined approach is to sample the distance function in an adaptive way such that fast variations in the distance function will not get lost in the sampling. One such approach is the adaptively sampled distance fields (ADF) described in [18] where an octree data structure is used as the underlying adaptive grid in the sampling of the function, see Figure 1.

![Figure 1](image)

**Figure 1.** Octree structure used for adaptive sampling of the distance function.

Triangulation of signed distance fields and more general scalar fields have been the topic of research for quite some time [19]. One of the first and most widely used algorithms to address the problem by discretization of the distance function on a grid is the Marching cubes algorithm of Lorensen and Cline [20]. The marching cubes algorithm uses the sign changes between the sampled values in the corners of a cell to create a local triangulation inside the cell this is then repeated for each cell in the grid. The marching cubes algorithm suffered from several drawbacks including the generation of superfluous triangles and triangles of low quality. There were also cases where the algorithm generated cracks and holes due to ambiguous cube configurations. The marching cubes algorithm is also hard to extend to adaptive grids. Many papers have been written in attempt to address these drawbacks and extend the Marching cubes algorithm, including [21][22][23]. Another way to improve geometric quality is by adding smoothing [27] but this route will not be taken in this paper. By shifting focus to the dual of a uniform cubic grid triangle quality was increased and the ambiguities associated with the regular grid was avoided [24]. The same idea can be applied to the dual of an octree structure which leads to the Dual contouring algorithm [25]. In the Dual contouring algorithm an approximate solution of the non-linear equation is computed inside each cell intersected by the contour and the triangle mesh is connected to the neighbouring cells by a quadrilateral for each intersected edge. By adaptively sampling the signed distance function in an octree structure such as in [18] and then applying the dual contouring, a high quality adaptive triangle representation of the shrunken model can be obtained, Figure 2.
By exploiting some properties of the distance function and tailoring the ADF to the distance of interest, a particularly efficient algorithm can be obtained. We choose to use a top-down approach to build the octree i.e. we start at the root node containing the entire volume of interest and then divide it into eight equally sized children and then apply the same procedure recursively to each child. The distance function is Lipchitz continuous with constant 1, i.e.
\[ f(a) - f(b) \leq |a - b|, \forall a, b \in \mathbb{R}^3, \]
we can use this to formulate a criteria that enables us to abort the subdivision early if it is impossible for the function values inside the current cell to equal the distance of interest. A simple such criteria is
\[ d_c + \frac{l}{2} < d, \]
where \( d_c \) is the value of the distance function sampled in the centre of the cell, \( l \) is the length of the diagonal of the cell, and \( d \) is the amount we would like to shrink.

As in [18] we also use the fact that we can measure how well a child is approximated by the parent using trilinear interpolation. If the approximation error is sufficiently small the subdivision is halted. This means that we get smaller cubes, i.e. higher sampling rate, where the function is dominated by nonlinear terms.

### 2.3. Comparison

Here we will make a comparison with the shrinking algorithm presented in [16]. The method described in [16] starts with the original geometry and tries to shrink it step by step until the final distance is reached, the step size can become quite small which makes this expensive in computation time and prone to numerical errors. Our new method, based on distance fields, starts directly to find the level set of the shrunken geometry by solving the nonlinear equation i.e. the computation time to shrink an object with the new method is independent of the shrinking distance. The resulting geometry of the new approach can be quite different, for example we do not guarantee that the shrunken geometry has the same topology as the original mesh, on the other hand it respects the distance criteria better than the old algorithm which is in general impossible to fulfill without changing the topology, see Figure 3.

None of the algorithms require tetrahedralization or calculation of the medial axis. The new algorithm is even more independent of the original triangulation and somewhat more insensitive to errors such as small cracks in the triangulation. A strict comparison between the two methods is terms of computation time and accuracy is quite hard to do since they are based on different assumptions and approximations. One of the main advantages of the new method is the possibility to make a trade off between computation time and quality of the result. Both algorithms are dependent on the orientation of the triangles to be able to distinguish inside from outside so care must have been taken in the CAD model design to ensure this. Alternatively a method such as the one described in [26] can be used to classify the geometry into either surfaces or solids, and for the latter also into inside and outside.

### 2. Application

One motivation to use shrinking is, as described in [16], to get a suggestion how to redesign a part in a minimal way in order for the assembly or disassembly to be possible. Another use is when the part is flexible. Instead of modelling the part as a true flexible object one can model it as a rigid shrunk version of the original part with the assumption that the object is allowed to be deformed by at most the shrinking distance during the assembly.

Now we will examine a case where we would like to disassemble a storage box from the trunk of a car, Figure 4.
The storage box is made of foam which makes it possible to allow small deformations without damaging the part or the surrounding, an ideal situation in which to apply our new shrinking algorithm. The first problem encountered is that even in the initial assembled position the object is already in collision with the environment. By shrinking the object by 2 mm the part is no longer in collision at the initial position, but there is still no collision free path between the initial configuration and the goal configuration outside the trunk. By shrinking another millimeter the path planner succeeds in finding a collision free path for the shrunk object and we know that if we are able to deform the part by 3 mm during the disassembly the given path is acceptable. If this amount of deformation cannot be accepted then we run the path calculated for the shrunk object, using the original geometry, and note all the colliding triangles which then can be presented to the designers to support design changes, Figure 5.

Figure 5. The intersections (red curve) between the environment and the planning object are accumulated and used to support redesign

3. Conclusions

In this paper we presented a new method to shrink geometry in order to support design changes in an early stage of car development projects. The new method works as a complement to the method described in [16]. The new method is based on representation of the geometry using signed distance fields which has the advantage of being able to handle large shrinking distances, non connectivity in the input, furthermore it also gives higher degree of quality control and more accurately fulfils the distance criteria.

Acknowledgment

The support from the Swedish Governmental Agency for Innovation Systems (VINNOVA) and the Swedish Foundation for Strategic Research (SSF) ProViking II program is gratefully acknowledged. This work was carried out at the Wingquist Laboratory VINN Excellence and is part of the Sustainable Production Initiative and the Production Area of Advance at Chalmers University of Technology.

References

A classification of carrier and content of information

T. Fässberg, Å. Fasth, J. Stahre
Department of Product and Production Development, Chalmers University of Technology, Gothenburg, Sweden

Abstract: Mass customization is a driver for increased assembly complexity and the variety and complexity of products and parts require new and effective information and information flows to support final assembly operators. How to best present and convey information is however a difficult task. The aim of this paper is to develop a classification of carrier and content of information that can be used as a support for task allocation and design of new information systems for an assembly environment. The developed classification is mainly based on how different carriers and contents are being used in assembly and related to literature. In the design of new decision and information systems both carrier and content needs to be optimized and the concept of content and carrier needs to be contextualized in order to be useful in a task allocation and design process.

Keywords: Automation, Assembly, Carrier, Content, Design, Information

1. Introduction

Mass customization, individualization and personalization of products are drivers for an increased assembly complexity in today’s production systems. In the assembly environment the high variety of products and parts require new and effective information flows to handle the massive amounts of information. Final assembly operators have a need for fast and reliable information as well as better decision and action support. But how to best present and convey information in an assembly context is however a difficult task. Previous research have stated that the amount and content of information is a contributor to production complexity [1, 2]. It is of great importance that the right information is presented at the right time in the right way. Classifications have been put forward regarding what makes information of high quality [3]. Research has even indicated that proactive behaviour can be accomplished by analysing the type and amount of information to be presented and how to structure information [4]. The presentation of information can be broken down into two parts; carrier and content of information. Carrier concerns the medium of information e.g. paper, screens, and PDAs while the content concerns the mode of information e.g. text pictures, sound or movies. In the design of new decision and information systems both carrier and content needs to be optimized. However the concept of content and carrier needs to be contextualized in order to be useful in a task allocation and design process.

The aim of this paper is to develop a classification of carrier and content that can be utilized in a task allocation process to support the design of new information systems for an assembly environment. The developed classification is based on how different carriers and contents are used and the existing need on information systems.

2. Automation in an assembly context

Automation in an assembly context does not only consider mechanical tasks it also concerns cognitive support for control and information tasks. In an assembly context the support functions for operators are increasing. A transition has been made from tools such as electric screwdriver (that only provided mechanical assistance) to information support systems e.g. Pick By Light systems, which also provide cognitive support. Thus the scope of automation have widened through the use of information technology. Automation has an impact on the operators cognitive functions, thinking as well as doing [5]. But in what situations is this support needed and how is the information needed best presented?

2.1 Quality of information in assembly

The perceived complexity in an assembly system is in relation to the amount of information in the system. Information systems in an assembly context should aim to support operators reducing the perceived complexity. However, to design such information systems is not a trivial task. A information system can be divided into the following categories [6]:

- What information to present – contents, meaning
- How it should be presented – format, context and receiver characteristics.
- When it should be presented – timing in relation to the decision, whether it should be presented automatically or on request.

The role of intelligent decision support system is to support decision maker with better quality of information. According to Kehoe it is the quality rather than the quantity of information that is of importance [3]. Six qualitative criteria are presented for how to create efficient information: relevance, timeliness, accuracy, accessibility, comprehensiveness and format [ibid]. Hollnagel argues that quality is not necessarily a feature of the information but rather of the interaction with information [6].

On the other hand, in order to form decisions a certain amount of information is required. The information gap theory presented by Endsley claims that more data does not necessarily result in more information [7]. The problem with today’s system is not the lack of information rather to find what is needed when it is needed [ibid]. To ensure quality of information the quantity might have to be filtered. However, it can be ambiguous to decide which information to consider relevant [8]. The amount of information needed by the operator is however individual and dependent on their level of expertise [9]. There is a potential risk of cognitive overload if the operator is surrounded with a large number of information, which creates stress [10]. This highlights the importance to present quality information rather than quantity.
2.2 Cognitive automation in assembly

Cognitive automation (LoA \(_{\text{cog}}\)) could be described as the amount of technique and information provided to the operator in order to know what, how and when to do a specific task in the most efficient way (LoA \(_{\text{cog}}\) = 1-3 in the matrix, seen in figure 1). When the technique or machine is performing the task i.e. higher physical automation (LoA \(_{\text{phys}}\) = 5-7), the cognitive automation is mainly used for control and supervision (LoA \(_{\text{cog}}\) = 4-7).

It is equally important to acknowledge the cognitive dimension as the mechanical. Although, when companies redesign their system, they often only consider the mechanical Level of Automation (LoA), while the cognitive LoA is left to be solved afterwards. There is also a tendency at companies, when the mechanical level of automation is decreases so is the cognitive level [13].

Case studies shows that over 80 % of final assembly tasks are performed by operators based on their own experience [14] i.e. without any decision support (LoA \(_{\text{cog}}\) = 1). If more automation should be used in final assembly it is better that these systems provide information rather than decisions [8]. Meaning that it is better to support operators with good information support rather than telling what to do without explaining the rationale behind the decision. Humans actions are determined by their understanding of the situation not how the designer expects or assumes the user to view it [15]. This puts high demand on the system to be sufficiently transparent and adaptable to the user’s needs [6].

The second part is the areas on the right side, which are important areas to consider when performing a task allocation. Level of Competence (LoC) and Level of Information (LoI) are two important areas to consider, these could be further related to measuring the cognitive LoA in the current system [11] but also how the information is presented to the operators i.e. carrier and content of information.

2.3 Carrier and Content of information

Presentation of information can be broken down into two parts; carrier (how to present) and content (what to present) of information. Carrier concerns the medium of information e.g. paper, screens, or PDAs while the content concerns the mode of information e.g. text, pictures, sound or movies. This is a trivial definition but yet with a powerful potential when designing information systems since the carrier and content can be decoupled when analysing the system.

In order to investigate the number of information carriers and what they convey, two types of mappings can be done; one with focus on the exchange of information and one with focus on type of carrier [18]. In the design of new decision and information systems both carrier and content needs to be optimized, this could be done by measuring and analysing the companies cognitive automation [19].

3. Usage of Carrier and Content

This section will explore the concept of carrier and content and present different examples of carrier and content within different contexts.

3.1 Carrier of information

The environment or context in which the information system will be used in will have an impact on the design needs of the carrier. A short cyclic and high variety setting will pose different demands than a long cyclic and low variety environment. For instance the assembly operator working on a driven assembly line with a cycle time of 60 seconds or less does not have time to read instruction. Support is rather given by Pick By Light systems or similar to help the operator to identify which parts to assemble but not how to perform the task. The size of the work place and need of tools will also be important factors.
to consider. As product variety and complexity increases so does the need for more information.

Work instructions presented on a paper, as a bill of material or a drawing is a mobile although static way of presenting information. The drawback is that it is slow when frequent updates are needed or inconvenient when having many variants in the production. To better handle a variety of products and frequent updates digitalisation might be needed and presentation on monitors rather than paper.

If there is a need for mobility the carrier can be made mobile for example using hand held units. An experimental study has examined how the information source (carrier) affects the quality and productivity when presenting assembly instructions. The information source (carrier) used was an iPod touch and a static screen. The GUI that presented the information was a web-based system displaying text-based information (content). The study shows that quality is positively affected by mobility and shows indications of increased productivity [20]. In a dynamic context with a lot of movements, many variants and large action range, mobility is an important parameter. Even in quite restricted stations mobility can make a difference [21]. Another example of optimisation of carrier can be seen in research evaluating how display size have an impact on pointing performance [22].

3.2 Content of Information

Content is a wide concept containing the actual data or information that is to be conveyed to the recipient. The recipient in this context is an assembly operator. The content can be presented by different modes such as text, picture, movie or sound. Each of these modes has their benefits, drawbacks and area of usage. Text has been the most common mode to use in work instructions, but pictures have become more commonly used. However guidelines to help engineers to make these decisions are few.

As the variety of products increase or the complexity of the product itself increases more elaborate instruction are needed, which put higher demands on the content of information. In an assembly context much focus in on the development of information systems but not so much focus on the actual content. To address this a multimodal approach was developed making use of many different contents [23]. Antifakos et al. used accelerometers to enable individual instructions based on the action of the user [24]. This was to supply correct information depending on the situation each individual was facing. This shows that the content can be designed to fit the individual operators situation and competence. Augmented reality is another example of optimised content. Augmented assembly instructions can be provided in forms of pictures and text by tracking objects such as pens or fingers [25]. This allows a work instruction that dynamically changes in response to the users actions.

3.3 Effects of Content and Carrier

When mapping Kehoes [3] list of parameters influencing quality of information onto the concept of carrier and content, it is apparent that carrier and content address different parameters. In the case of carrier the design choices will affect the timeliness and format thus influencing the quality. It will have an impact on the accessibility however in Kehoes definition it is expressed that this parameter should not concern the medium transferring the information. While the content design will influence all parameters proposed to affect the quality of information. Manipulating the two parameters, content and carrier have an impact on Kehoes quantitative measure, thus affecting the effectiveness of the information system. When discussing carrier (how) and content (what) the when is missing. However different carriers can make it easier to control the when e.g. a mobile ICT tool is very beneficial in providing attention triggers and the information can be adjusted to fit the individual [21]. Further, since the quality of the interaction with the information can been altered by the choices of carrier and content, the quality of information can be improved be these choices.

The same LoA can have many different solutions of carrier and content. For example, LoAcog = 3 = Teching, can be a drawing on a paper or a text based list on a computer screen. Each solution has its benefits and drawbacks for a certain context. This implies that LoA alone cannot be used as decision criteria when designing cognitive support for assembly operators. It is evident that guidelines for what carrier and content to chose for a certain context are needed as help in a task allocation process. Figure 3 illustrates an example of a fixture being used as both carrier and content to secure quality for screwing tasks.

It is important to recognise that not only contextual parameters such as cycle time and station size need to be taken into account. It is also important to acknowledge that individual factors such as experiences and competence are interlinked with the need of information and when and how it is needed. The experienced operator often works by own experience (LoAcog = 1). Optimised carrier and content can be used to design smart cognitive support to increase that figure to ensure quality while supporting different experience levels of operators providing different needs of cognitive automation.

The development of cognitive automation might lead to an atrophy of mental skills [15]. Hence the design and use of cognitive systems must be made in awareness of its implications.

4. Conclusions

The results from the paper shown the importance to see cognitive automation as an enabler towards a more effective and competitive system by adapting the operators needs. The concept of dividing the possible solution into carrier and content makes it easier to adapt to the operators individual needs.
Further it gives an opportunity to give a more nuanced view of the information flow within an assembly context.

A start has been made in the development of a classification of content and carrier of information. Further development is needed in order to give better design suggestions of information systems. Further case studies and experiments are suggested to quantify the relationships between carrier, content, and quality and time parameters. The classification could then be incorporated in the presented concept model and used as a guideline when performing task allocation.

Acknowledgements

This work has been carried out within the Sustainable Production Initiative and the Production Area of Advance at Chalmers. The support is gratefully acknowledged.

References

Cost impact assessment of production program changes: a value stream oriented approach

J. Gottmann\textsuperscript{a,b}, W. Mayrhofer\textsuperscript{a,b}, W. Sihn\textsuperscript{a,b}

\textsuperscript{a}Vienna University of Technology Institute of Management Science, Austria
\textsuperscript{b}Fraunhofer Austria, Austria

Abstract: High capital production assets require an adequate workload to benefit from economies of scale. This and an ever increasing number of product variants often lead to enlarged batch sizes resulting in heightened work in process due to safety stocks and additional changeover. In times of economic volatility, changing production programs cause fluctuations in capacity demand along the value stream. Fixed costs have to be distributed over an ever changing amount of products – batch sizes and production costs are permanently altered. To assure the success of investment decisions, various assessment methods for new machines and their capacity such as the calculation of Net Present Value or Internal Rate of Return exist. These methods imply a predicted production program and associated costs. In contrast, follow-up costs along the value stream are often undertended in the calculation of future scenarios and planned measures. Possible impacts of a change of the production program (volume and variants) are mostly unknown.

The aim of the developed calculation model is the estimation of the flexibility of costs (elasticity) depending on these various cost drivers (units, variants and batch sizes). It supports the forecast of possible impacts regarding uncertain future developments and discloses that section of the value stream responsible for cost related effects and where necessary measures for improvement should be located.

Keywords: Production Costs, Flexibility, Product Variants

1. Introduction

The capability of a production system, to produce different product variants and different volume at an acceptable speed and cost, results in production flexibility [1]. However, an increasing product variety causes an increasing changeover effort on existing production equipment. To assure an adequate machine workload, higher lot sizes have to be formed. This, again, is contrary to the principles of Lean Production, which demand low work in process (WIP) for a short throughput time and hence require small batch sizes [2,3].

„All we are trying to do is shorten the time line...“
Taiichi Ohno, Toyota Production Chief after WWII

„The easiest of all wastes and the hardest to correct is the waste of time“
Henry Ford, Founder of Ford Motor Company

As another apparent issue, a change of production program or investments in the existing bottleneck can lead to a shift of the bottleneck in the value stream [4], which possibly changes the structures and efforts in support and logistics. Moreover, if not planned properly, investing into production equipment sometimes triggers a spiral that can be described as follows: higher assets need a higher workload – batch sizes are raised, throughput time worsens while simultaneously does flexibility of the production to accommodate to different product variants [5]. A shift of the bottleneck changes the behavior of current inventory and WIP within the value stream, as well as the corresponding allocation of thereby occurred costs and the level of costs.

Due to the complexity of the relationship between a changing number of product variants, its impact on the value stream and resulting costs, the estimation of these impacts and cost changes is difficult. Existing approaches to assess flexible production systems often focus on technical scope of producing different product variants at increasing quantity [6, 7, 8]. They evaluate production systems regarding an existing or predicted production program [8] or use predetermined cost factors without including their origination or shift due to changing production structures. [10, 11, 12].

However, the consideration of the whole value stream is essential to model altering conditions and bottleneck situations and hence facilitate the balancing of capacities. On this account, a proceeding that describes this correlation, considers all costs along the value stream and plots them against an altered production program (product variants and volume) is needed.
This paper describes the principles of an approach enabling the user to specify the developing of costs of each variant in the considered value stream, to reveal cost changes after a change of the production program and to localize sources of possible cost hikes along the value stream. To afford such an analysis of all relevant impacts, costs along the value stream are assigned according to their dependency on produced units, different variants and the connected batch sizes. The value stream thereby is divided in production processes and inventories (Figure 1). Moreover, support processes depending on the production can be included [5].

Compared to Activity Based Costing, which is a continuous calculation scheme, the presented method is used to forecast future costs with respect to possible scenarios in the production program. Thereby, potential over- and under loads of processes and employees are included in the cost calculation to foresee future additional costs or prevent shortfalls respectively [13].

2. Cost types along the value stream

Costs and expenses dependent on product variants are sited in the production area and its support activities. Expenses in production consist of wages, material and stock (current assets) and capital assets. The factor input needed for manufacturing is primarily dependent on the volume of manufactured product, the number of product variants, the resulting days of inventory and batch sizes (and connected number of batches). The relevant cost types along the value stream are depicted in Figure 2.

The cost units of the corresponding support processes are primarily composed of wages and salaries. The level of these expenses depends on several cost drivers, i.e. number of customer orders or timed transaction cycles.

According to the method of Activity Based Costing, all activities of these cost units can be divided in activity quantity induced (aqi) costs and activity quantity neutral (aqn) costs and will be related to the corresponding cost drivers [13]. Those cost drivers who are activated by the production process are once more the volume of produced goods, the amount of product variants and the amount of lots produced.

3. Connection of production units and product variants by the EPEI (Every Part Every Interval)

The basic concept of the described approach is the connection of production processes and inventories along the value stream and the calculation of variant-dependent costs with the help of the EPEI. If a value stream produces more than one product variant with the same production processes, usually changeover processes are needed. These changeover efforts must also be accomplished during the tact time, which represents the average customer’s request frequency.

In a best case scenario, process and changeover time for one specific part or good fit into the tact time and a one-piece-flow production is possible. Since changeover times in several industries are time consuming, this presumption is not always realistic. To meet the tact time anyway, production batches are formed to reduce the number of changeovers needed. Forming batches delays the production of other product variants, since the whole batch has to pass a process before another product variant can be started [2].

Hence not all variants may be produced in one day. This relationship can be described by the indicator EPEI. It specifies how long it takes to produce all product variants in their corresponding batch sizes. It considers the daily consumption, which has to be met, but also the necessary changeover times.
incurred for the several product variants. The EPEI is defined as
\[ EPEI_{\text{Process}} = \sum_{i=1}^{n} \frac{CO_i}{A_t - \sum_{i=1}^{n} PT_i \times DC_i} \]
\[ \text{with} \]
\[ i = 1, \ldots, n \quad \text{Amount of product variants} \]
\[ CO_i \quad \text{changeover time of variant} \, i \]
\[ PT_i \quad \text{process time of variant} \, i \]
\[ DC_i \quad \text{daily consumption of variant} \, i \]
\[ A_t \quad \text{technical availability (plan capacity)} \]
\[ A_t - \sum_{i=1}^{n} PT_i \times DC_i = \text{remaining time for changeover} \]

EPEI describes the relationship of necessary changeover time for all product variants with respect to available working hours for changeover efforts. This in turn is described by the available working hours for production per day minus the sum of process time needed to satisfy the daily consumption (Figure 3).

The EPEI depicts, that all product variants can be produced within a specific time period. This period demands the formation of batches, which secure the stocking up to meet the customer’s demand of every product variant. The level of the batch sizes relates to the necessary days of inventory and arises directly from the EPEI. The higher the EPEI the longer inventory must last to meet daily consumption (Figure 3). The EPEI describes the necessary changeover time for all product variants with respect to available working hours for changeover efforts. This in turn is described by the available working hours for production per day minus the sum of process time needed to satisfy the daily consumption (Figure 3).

The EPEI describes the relationship of necessary changeover time for all product variants with respect to available working hours for changeover efforts. This in turn is described by the available working hours for production per day minus the sum of process time needed to satisfy the daily consumption (Figure 3).

4. Conjunction along the value stream

The conjunction of the separate cost types along the whole value stream results from batch sizes and tied-up capital. The amounts contained in inventory arise from the batch sizes of previous and subsequent processes. It includes the value of the material of the product as well as value created in previous processes. Moreover, the additional purchased parts must be considered by their replacement price and their value must be accumulated over the following inventories to calculate their tied-up capital (Figure 4).

Every batch causes efforts in logistics for current inventory and warehousing as well as in corresponding support areas [13]. However, such effort may not be accumulated since no value was added. The number of batches plays an important role in calculating overall efforts within the value stream. The number of batches arises directly from produced batch sizes and the yearly requirement.

Bottlenecks govern batch sizes within the value stream based on its limited capacities, determining the smallest possible batch of each variant that enables meeting daily consumption. The same batch sizes should be assumed for all previous processes. If a previous process produced smaller batch sizes, this would increase changeover efforts without decreasing inventory space, since the following process still needs a higher batch size. It is only the waiting time that may be decreased, at the expense of the synchronicity of production. If a previous process produced larger batch sizes than the bottleneck, this would increase the inventory without generating additional capacity. If different batch sizes can’t be avoided, they should at least be a multiple of each other to ease production planning and control.

After the bottleneck, batch sizes can be smaller, because this raises flexibility towards the customer as production is able to switch faster to other product variants. Higher batch sizes after the bottleneck do not lead to a useful generation of capacity (because there is no higher throughput at the bottleneck) and overall decelerate the process towards the customer (Figure 4).

5. Approach

To calculate the described cost types and their conjunction, first the initial state must be identified. Therefore, process data

---

**Figure 4 Conjunction along the value stream**
defined by the value stream analysis has to be mapped for all existing product variants. The result is a production program with process and changeover times, corresponding shift models and the required tact time determined by the customer.

Further, all assets, wages, material costs and other expenses have to be compiled and assigned according to their dependency on produced units and different product variants. For the calculation of inventories in the value stream this means data generation of purchased parts and material regarding prices of retrieval, floor space required, container sizes, etc. as well as logistics efforts in time per batch.

The necessary data for production processes are divided in machine data and operator data and describe the process and changeover times for both. To cover the support areas, all activities in the relevant cost units as well as their cost drivers must be identified. General information has to contain the mentioned shift models, assets and direct costs. If the aim is a break-even-analysis, a revenue function must be dedicated.

With the help of this data basis the described cost types can be calculated and disposed to the different variants according to their different process and changeover times. If the capacity of one machine or operator is not utilized completely, all expenses have to be distributed to all goods produced by a utilization factor [13].

\[
J_{CU,j} = \frac{A_{k,j}}{CN_{PT,j} + CN_{CO,j}}
\]

\[
\text{with}
\]

\[
\begin{align*}
J &= 1 \ldots m & \text{process~j} \\
J_{CU,j} &= \text{factor~of~capacity~utilization} \\
A_{k,j} &= \text{capacity~available} \\
CN_{PT,j} &= \text{capacity~needed~for~process~time} \\
CN_{CO,j} &= \text{capacity~needed~for~changeover~time}
\end{align*}
\]

Finally the units produced and the number of variants (the production program) can be levered and an estimation of costs in a scenario of production program changes can help the user value existing or future production structures.

6. Summary and Conclusion

With the presented approach, scenarios for production program changes in existing or future production structures regarding their impacts on the level and the allocation of the costs along the value stream can be depicted.

If it is possible to picture interrelationships between product variants, production volume (units produced) and costs in 3 dimensions, the related revenues and critical production programs can be identified. If probabilities and corresponding cash flows are deposited, the results can be used for investment appraisals or discounted cash flow methods. Finally, this approach will allow an in depth comparison of different production structures.

The proposed approach will be the basis of a calculation tool that is developed. Currently the system specifications and the conceptual design of the calculation tool are in progress. For an extensive testing phase and in order to secure the viability of the approach an experimental setup of the solution is envisaged.

References
Discovering design structure matrix for family of products
M. Kashkoush, T. AlGeddawy, H. ElMaraghy
Intelligent Manufacturing Systems (IMS) Centre, University of Windsor, ON, Canada

Abstract: Challenges confronting designers and manufacturers to change their design and manufacturing capabilities, responsively and cost-effectively, are increasing. This is mainly attributed to the increasing products variety and design changes. Design knowledge is accumulated over generations of product variants, based on designers and manufacturers experience. It is informative and useful to discover, retrieve and use this embedded knowledge in planning for future designs. In this paper, an Integer Programming Knowledge Discovery model is developed to reveal dependency relationships (association rules) between different existing components of a given product family. Relationships are extracted in the form of a Design Structure Matrix (DSM); a system representation tool used for system decomposition and integration analysis. The developed model was successfully applied to a case study of a family of seven water boiling kettles. This paper introduces a novel and elegant approach for the automatic generation of the Design Structure Matrix (DSM) for a family of products.

Keywords: Product development, Assembly, Knowledge Discovery

1. Motivation
The proliferation of product variety and design changes adds to the challenges faced by designers and manufacturers to change their design and manufacturing capabilities responsively and cost-effectively. Existing product variants represent a set of physical and feasible solutions to the formulated design objectives. Each variant consists of a combination of parts and modules that enable products to perform the intended functions. Product components may be dependent; independent or mutually exclusive.

Design knowledge is accumulated over generations of product development based on designers and manufacturers’ experience and knowledge of design, manufacturing and assembly criteria. It would be very useful to discover, retrieve and use this embedded knowledge in planning for future products and variants. This would enable greater responsiveness to changes in products and more cost effective generation of new product variants and their manufacturing processes.

This paper proposes a new mathematical model for discovering the design knowledge, inherent in variants of a given product family, in the form of relationships among product components. Finding product components dependency relationships for multiple product variants is similar to the well-known field of Knowledge Discovery in Databases (KDD) [1] and particularly association rules extraction in computer science [2]. However, the proposed model is different since it generates definite dependency relationships instead of association confidence levels produced by other approaches.

The proposed Knowledge Discovery model is developed to extract dependency relationships, or association rules as called in other fields such as computer science, between different design features or components of already existing designs of a given product family. The Design Structure Matrix (DSM) is then constructed using such relationships. DSM, also known as Dependency Structure Matrix, provides a simple and efficient visual representation for a given system and supporting solutions for the decomposition and integration problems [3]. DSM is generated at present with input from users and designers.

2. Literature Review
2.1. Design Structure Matrix (DSM)
A Design Structure Matrix is a square matrix with identical columns and rows labels used for systems representation and analysis. A DSM represents the inter-relationships between different components of a given system in a visual, compact and analytically useful form. In a DSM, a given cell is marked or assigned a value if its row component depends on its column component. In the DSM shown in Fig. 1, element A, for instance, depends on elements B and D, and elements C and E are dependent on it.

![Figure 1. An example of a DSM of a system of five components](image)

The term Design Structure Matrix was first introduced by Steward [3], however, other DSM-like data representation forms used to be employed before under different names in various research fields (e.g. the “roof” of a Quality Function Deployment (QFD) matrix).

DSM also shares some similarity with the Design Matrix (DM) of Axiomatic Design (AD) introduced by Suh [4]. Unlike DSM, the DM captures dependencies between components or elements that lie in two different domains; such as functional requirements and design parameters domains. Some researchers have combined both tools, Axiomatic Design and Design Structure Matrix, in a single integrated procedure [5]. According to the review by Browing [6], DSM is classified into four types as follow:

1. Component-Based or Architecture DSM: For modeling system architectures based on interactions between the different components.
2. Team-Based or Organization DSM: For modeling organization structures based on interaction between the different groups of people.

3. Activity-Based or Schedule DSM: For modeling processes and activity networks based on information flow and dependences between the different activities.

4. Parameter-Based or Low-Level Schedule DSM: For modeling low level relations like those between design decisions and parameters or systems of equations.

The DSM type employed in this research is component-based. A typical application for this type is found in Pimmier and Eppinger [7] who utilized component-based DSM through an integration analysis methodology that clusters a given system components into groups based on different types of interactions; such as spatial, energy, information, and material interactions. These groups are then used to define the product architecture and the structure of the product development teams. The proposed methodology was applied to an automotive climate control system of 16 different components and useful results were obtained.

DSM is usually used to study single systems or products; few studies have used DSMs to analyze a family of products. Yu and Cai [8] used DSM to construct a Product Master Structure (PMS) for a family of products, which is a hierarchical structured form of representation for the entire components of a group of products with similar functions. PMS could be a useful supporting tool for rapid development of future product variants. Luh et al. [9] also employed DSM in a two stage approach to re-engineer product architectures and corresponding design process for a family of products and applied it to a case study of PLC product family of four variants.

An automated procedure for discovering Design Structure Matrix (DSM) for a family of products, or even a single product, does not exist in the literature to date.

2.2. Knowledge Discovery

Knowledge Discovery in Databases (KDD) is a very wide topic of interest primarily to Information and Computer Science researchers. However, finding useful patterns in a given set of data is a commonly encountered task in many other research fields. The term Knowledge Discovery in Databases (KDD) was first introduced by Piatetsky-Shapiro [1]. According to the introductory article of Fayyad et al. [10], KDD is defined as the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. A comprehensive survey of KDD can be found in [11].

The Knowledge Discovery conducted in this research could be classified under Dependency Modeling type of Knowledge Discovery [10]. It involves several sub-problems; for instance: mining sequential patterns [12], episodes [13], correlations [14], and association rules [15, 16]. Mining Association Rules, also known as Association Rules Learning and Association Rules Discovery, is the type of Knowledge Discovery employed in this paper.

Association Rules Discovery is concerned with finding interesting relations (associations) between variables (data items) in large databases. Such a topic was first studied by Hajek et al. [17] under the name “Hypotheses Determination”. However, the introduction of Association Rules Discovery is usually credited in literature to Piatetsky-Shapiro [15]. Association Rules Discovery was popularized particularly due to the prominent work of Agrawal et al. [2] who studied the problem of mining a large collection of dataset transactions for significant association rules between sets of products (items). An example of an association rule could be a statement that “87% of customers purchasing bread and butter also purchase milk (i.e. Eggs, Butter) \(\rightarrow\) (Bread), where the 87% is the confidence factor of the rule (i.e. rule significance).

Agrawal et al. and others who addressed this problem look for association rules that are supported with, or satisfying, some statistical measures with predefined values. The Knowledge Discovery problem used in this paper is different as the proposed model searches for definite association rules with 100% confidence.

3. Proposed Knowledge Discovery Model

Each variant of a given product family consists of a set of components (parts and/or modules) that are combined to perform some designated functions. The used design knowledge is accumulated over several product generations based on designers and manufacturers experience. Such an implicit knowledge, particularly dependency relationships among product components, is what is sought by the proposed Knowledge Discovery model.

A simple Integer Programming model of one main set of constraints is formulated to discover definite dependency relationships between components of a given family of products. The proposed model searches for the set of components on which each component depends. When component A for instance is said to “depend on” components B and E, it means that component A has never existed in any of the studied variants without components B and E. This is exactly what the proposed model looks for; the set of components that appear together in the entire set of studied product variants.

A given set of product variants is first decomposed into parts and/or modules (components for generality). Then a matrix of binary values as the one shown in Table 1 is constructed. Columns in this matrix represent components and rows represent product variants. If “1” is assigned to a given cell in that matrix, it means that the product variant in the corresponding row includes the component in the corresponding column. For instance, third product variant in Table 1 contains components no. 2, 4, 5, and 7.

<table>
<thead>
<tr>
<th>Variant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var. 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Var. 2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Var. 3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Var. 4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Var. 5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Var. 6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

An IDEF0 model for the proposed dependency relationships discovery approach is shown in Fig. 2. The input to the model is the design data matrix shown in Table 1, the controls are the objective function and constraints, the mechanism is to solve the model, and the output is the Design Structure Matrix.
The proposed Integer Programming model is applied to a product family of seven household kettle variants including a total of thirteen different components as shown in Table 2: container, two different types of handles, two types of temperature control, two types of on-off switches, two types of bases, and four types of lids. The container (1) is a common component across the entire studied family of kettles. Handles are either short (2) or long (3). Some kettles boil the water and automatically disconnect (4), while other kettles reach and maintain a user-defined temperature (5). Some kettle variants have the on-off switch at the top next to the lid (6), while others have it at the bottom under the short handle (7). Some kettles bases are attached to the containers (8), but there are also some bases which are detachable for ease of filling and pouring (9). The openings to fill the kettle with water might have a hinged lid (10), hinged releasable lid with a pop-up mechanism (10 and 11), detachable lid (12), or even a fixed lid (13). Possible variants of each set of components are illustrated in Fig. 3. Underlined features in that figure are for the second kettle variant, which is the one used for demonstration.

The model parameters are the elements of the input design data matrix. An element $x_{ij}$ in that matrix takes the value “1” if product variant $i$ includes component $j$. The total no. of products is $u$, and the total no. of components is $v$. The proposed model has one type of decision variables which are the elements of the targeted DSM. Thus, $Z_{ij}$ is a binary variable that takes the value “1” if the component $k$ depends on component $j$.

The objective function of the proposed Integer Programming model maximizes the total value of all elements of the produced DSM (equation (1)) which means maximizing the value of individual DSM cells (every decision variable). Accordingly, the model initially assumes that a relationship exists between every two components.

$$\text{Maximize } \sum_{k=1}^{v} \sum_{j=1}^{v} Z_{kj} \quad (1)$$

However, in addition to binary constraints (equation (3)), this objective is controlled by a single set of constraints (equation [2]), that does not allow any decision variable $Z_{ij}$ (an element in the produced DSM) to take a non-zero value if its corresponding column component $k$ does not always appear with its corresponding row component $j$ in the input design data matrix. Therefore, only DSM cells corresponding to dependent components are guaranteed to have a value of “1”.

$$Z_{kj} \left( \sum_{i=1}^{u} x_{ik} - \sum_{i=1}^{u} x_{ik} x_{ij} \right) = 0 \quad k = 1, \ldots, v, \quad (2)$$

$$j = 1, \ldots, v$$

$$Z_{kj} \in \{0,1\} \quad k = 1, \ldots, v, \quad (3)$$

$$j = 1, \ldots, v$$

It is expected that all diagonal elements in any DSM produced by the optimization model will take a value of “1”, which represents self-inference of each component, hence they can be ignored.

4. Case Study
data where it could be observed that those two components are
the only components that always appear/disappear together.

Table 2 Design data representation matrix

<table>
<thead>
<tr>
<th>Kettle Variant</th>
<th>Container</th>
<th>Handle</th>
<th>Temp. Control</th>
<th>On-Off Switch</th>
<th>Base</th>
<th>Lid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>A. Corded Basic</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>B. Cordless Basic</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. Corded Control</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>D. Whistle</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E. Cordless Easy Open</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F. Cordless Simple</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G. Cordless Classic</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 4. The DSM generated by the proposed model for the components of the studied family of kettles

In addition to the typical application mentioned in Section 2 for component-based DSMs, a very useful application would be to translate it to the corresponding master liaison graph for entire family of kettles. A master liaison graph [18] could be used for analyzing integration and modularization of a given product family, supporting mass customization and delayed product differentiation strategies for efficient variety management.

5. Conclusions

This paper has introduced an elegant approach for the automatic generation of component-based Design Structure Matrix (DSM), instead of constructing it manually through visual inspection. A new Integer Programming Knowledge Discovery model was formulated and used to discover dependency relationships between different components of a given product family and generate the corresponding DSM. The proposed model was successfully applied to a family of kettles of 7 variants and 13 different components. The model is straightforward and easy to program and use, however, the significance or the reliability of the discovered dependency relationships relies on the size and accuracy of available data.

The next step of this ongoing research work is to use the discovered DSM in constructing a master liaison graph for the entire family of products. This would be useful in analyzing integration and modularization opportunities.

References

Enhanced mixed integer linear programming for flexible job shop scheduling

V. Roshanaei, H. ElMaraghy, A. Azab
Intelligent Manufacturing Systems (IMS) Centre, University of Windsor, Ontario, Canada

Abstract: This paper proposes a novel Mixed Integer Linear Programming (MILP) formulation for Flexible Job Shop Scheduling Problems. Flexible job shops normally have a certain number of multi-purpose machines executing jobs of different processing requirements. The proposed MILP model is constructed based on the Wagner’s definition of binary variables and is an enhanced version of one of the most popular MILP models in literature. The developed model has fewer constraints and decision variables. Accurate size dimensionality measurement and comparative studies have been performed. A standard benchmark has been used to demonstrate the computational efficiency and effectiveness of the proposed model. The results confirm the computational superiority of the new model.

Keywords: Scheduling, Flexible Job Shops Scheduling Problem (F-JSSP), Mixed Integer Linear Programming (MILP)

1. Introduction
Thirst for increased productivity in the modern business world has spurred manufacturing practitioners to seek every opportunity for cost reduction and profit generation. Production scheduling has long been recognized as a tool for enhancing productivity. The importance of scheduling as a logical enabler in manufacturing systems has increased recently due to the growing consumer demand for variety, reduced product life cycles, changing markets with global competition and rapid development of new processes and technologies [1]. These economic and market pressures stress the need for minimizing inventory while maintaining customer satisfaction of production and delivery. This requires efficient, effective and accurate scheduling. The scheduling problem studied in this paper is a special case of FMS referred to as Flexible Job Shop Scheduling problem (F-JSSP) which is usually encountered in industries with high product variety and medium demand for each product.

1.1. Problem definition and representation
The F-JSSP extends the regular job shop scheduling problem (JSSP) by assuming that each job can be feasibly executed by more than one machine out of a subset of eligible machines. This adds another decisional complexity to the classical JSSP known as routing flexibility or feasible assignment. The addition of the optimal machine selection intensifies the NP-hardness of the classical JSSP decision which only takes care of operations sequencing. The incorporation of routing flexibility into classical JSSP can create two essentially different F-JSSPs [2]:
- **Totally F-JSSP (TF-JSSP)** where each operation can arbitrarily be executed on any available multi-purpose machine in the shop.
- **Partially F-JSSP (PF-JSSP)** where each operation can be processed on a subset of available machines in the shop. Therefore, not every machine is capable of processing all operations. Any infeasible assignment of operations to machines creates serious problems on the shop floor. The PF-JSSP is more difficult to solve as opposed to the TF-JSSP for the same number of machines and jobs [2]. Therefore, to facilitate the solution, the PF-JSSP is transformed to the TF-JSSP by adding “infinite processing times” referred to as big M to the incapable machines. Scheduling problems are represented by a standard triplet \((\alpha / \beta / \gamma)\). According to [3], the PF-JSSP can be denoted by \(FJc|M_i |C_{\text{max}}\). The first symbol indicates the type of shop which is F-JSSP, while the second symbol indicates machine eligibility. Finally the third symbol denotes the objective function which is make-span. The problem studied in this paper encompasses both TF-JSSP and PF-JSSP.

1.2. Assumptions
- Optimal singular process plan is determined a priori for each part type. i.e., no process plan flexibility is considered.
- Certain operations can be processed on more than one machine, i.e., there exists routing flexibility.
- Jobs are independent and no priorities are assigned.
- Pre-emption or cancellation of jobs is not considered.
- Each machine can only process one job at a time.
- Each job can be processed by only one machine at a time.
- Infinite in-process storage is allowed.
- Processing times are deterministic and include set-up, operations, transportation and inspection (approval).
- All jobs are inspected a priori i.e., no defective part is considered.
- The orders volume is known a priori and jobs are simultaneously available at time zero.
- Breakdowns are not considered.

1.3. Objective
The purpose of this study is to develop an integrated scheduling solution technique capable of both assigning and sequencing all operations to all available machines so that maximum completion times of operations (make-span) is minimized for both TF-JSSP and PF-JSSP. It could be inferred that make-span optimization incorporates machines utilization through reduction of idle times on all machines and ensures even workload distribution among work centres [4]. A Mixed Integer Linear Programming (MILP) formulation that perfectly captures the details of the problem is proposed.

2. Literature survey:
Many MILP formulations of production scheduling were inspired by Wagner’s seminal paper [5] employing various objective
functions, assumptions and solution methodologies. To the best of authors’ knowledge, only one position-based MILP model with make-span objective function exists which has been proposed by Fattahi, et al [6] for F-JSSP. This issue is even evident in the exhaustive literature survey of Ozguven, et al. [7].

2.1. Motivation and contribution:
In this study, an enhanced version of the MILP model proposed by [6] has been presented to develop an effective position-based MILP formulation with improved computational efficiency and solution effectiveness. The main idea behind a position-based model is operations of different jobs are scheduled on positions on different machine tools of a job. The decision variable of the model is binary that takes a value of one if operation l of job j (Ojl) is scheduled to position f on machine i.

3. Proposed MILP
The used notations are presented below.

**Notations**
- j Index for jobs where (1 ≤ j ≤ n)
- i Index for machines where (1 ≤ i ≤ m)
- l Index for operations of job j where (1 ≤ l ≤ n)
- e(j,l) Takes value 1 if machine i is processes Ojl and 0 otherwise.
- f Index for positions (order) on machine i where (1 ≤ f ≤ fj)
- Pj,l Processing time of operation Ojl on machine i.
- M A very large positive number

**Decision Variables:**
- Xj,l,i,f Binary variable taking value 1 if Ojl is processed in position f of machine i, and 0 otherwise.
- Sj,l Continuous variable for starting time of Ojl
- Bj,f Continuous variable for the beginning time of each order-position

**Min Cmax**

\[
\begin{align*}
\sum_{i} \sum_{f} X_{j,l,i,f} &= 1 \\
\sum_{j} \sum_{l,i} X_{j,l,i,f} &\leq 1 \\
\sum_{j} X_{j,l,i,f} &\leq e(j,l) \\
S_{j,l} &\geq S_{j,l-1} + \sum_{f} \sum_{i} P_{j,l,i} X_{j,l,i,f} \\
B_{j,f} &\geq B_{j,f-1} + \sum_{l} \sum_{i} P_{j,l,i} X_{j,l,i,f} \\
B_{j,f} &\geq S_{j,l} + M(1 - X_{j,l,f}) \\
B_{j,f} &\geq S_{j,l} - M(1 - X_{j,l,f}) \\
C_{max} &\geq S_{j,l} + \sum_{f} \sum_{i} X_{j,l,i,f} P_{j,l,i} \\
B_{j,f}, S_{j,l} &\geq 0 \\
X_{j,l,i,f} &\in [0,1] \\
x_{j,l,i,f} &\in [0,1]
\end{align*}
\]

Constraints set (1) ensure that each operation is assigned to one and only one position of all available machines. Constraints set (2) demonstrate the fact that some capabilities of all machines may not be fully utilized. Constraints set (3) ensure that each operation is processed on one of the eligible machines that have been determined a priori. Therefore, the first 3 sets of constraints ensure operations are feasibly assigned to their respective machines. Constraints set (4) preserve the precedence relationships between the starting times of operations of a job. Constraints set (5) express the fact that each operation is scheduled to slot/position. It would ensure the earliest time an operation is scheduled on a machine is when the slot on the machine is available and it is the operation’s turn to be scheduled. These two constraints are Either-Or constraints. Constraints set (8) takes care of calculating the last completed operation for each job which is the maximum completion time of operations on all available machines. Constraints set (9) reveals that the continuous variables representing starting times of operations and positions are invariably positive. Constraints set (10) demonstrate the binary nature of the decision variables. In the proposed MILP, certain efforts are made to reduce numbers of binary variables which exercise substantial impact on the efficiency of MILPs. As a result, the binary decision variable (yj,h,i) is removed from the MILP proposed by [6], which substantially increases computational efficiency and solution effectiveness of the proposed MILP. Getting rid of this binary decision variable reduces the number of binary decision variables of the proposed MILP by (nm²) as compared to the MILP proposed in [6]. Furthermore, the constraints sets (7), (9) and (10) were succinctly condensed to constraints sets (2) and (3). The foregoing consolidation markedly reduces the number of constraints.

4. Computational evaluations:
It is true that the parametric superiority of mathematical models can be inferred from MILPs building blocks, which are binary, and continuous decision variables and number of constraints. But the common practice for making analogies between the performances of MILPs is to compare the computational efficiencies of the combined MILPs building blocks. This is because in some cases, one of the MILPs may possess fewer numbers of decision variables and higher numbers of constraints, but the other MILP may have higher numbers of decision variables yet fewer numbers of constraints. Two tables have been designed to report the key performance measures of the MILP proposed in [6] and the proposed MILP. Table 1 addresses the number of binary integer and continuous decision variables and also the number of constraints for the MILP proposed in [6]. Table 2 demonstrates the statistics regarding the building blocks of the proposed MILP. Table 3 summarizes the errors resulting from each MILP with respect to key performance measures using relative performance deviation (RPD). RPD measures the deviation of each MILP with respect to the best performing MILP.

\[
\text{RPD} = \frac{\text{MILPsolution} - \text{MILPbest}}{\text{MILPbest}} \times 100\% \quad (\text{Equation 11})
\]
CPLEX Optimizer 12.1.0 is used. As shown in Table 3, all the elements of the proposed MILP are superior to the MILP proposed by in [6].

Furthermore, the proposed MILP can produce feasible integer solutions for 5 larger size instances and can solve the benchmark up to 8 jobs and 7 machines versus other model in the literature.

**Table 4. Performance of the proposed MILP**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MFJS1</td>
<td>2.2.2</td>
<td>40</td>
<td>32</td>
<td>25%</td>
<td>26</td>
<td>13</td>
<td>50%</td>
</tr>
<tr>
<td>MFJS2</td>
<td>2.2.2</td>
<td>32</td>
<td>24</td>
<td>33%</td>
<td>24</td>
<td>11</td>
<td>118%</td>
</tr>
<tr>
<td>MFJS3</td>
<td>3.2.2</td>
<td>72</td>
<td>60</td>
<td>20%</td>
<td>36</td>
<td>17</td>
<td>112%</td>
</tr>
<tr>
<td>MFJS4</td>
<td>3.2.2</td>
<td>84</td>
<td>60</td>
<td>40%</td>
<td>38</td>
<td>17</td>
<td>100%</td>
</tr>
<tr>
<td>MFJS5</td>
<td>3.2.2</td>
<td>84</td>
<td>72</td>
<td>16%</td>
<td>38</td>
<td>19</td>
<td>100%</td>
</tr>
<tr>
<td>MFJS6</td>
<td>3.3.2</td>
<td>189</td>
<td>135</td>
<td>40%</td>
<td>50</td>
<td>25</td>
<td>100%</td>
</tr>
<tr>
<td>MFJS7</td>
<td>3.3.5</td>
<td>225</td>
<td>162</td>
<td>39%</td>
<td>55</td>
<td>28</td>
<td>96.5%</td>
</tr>
<tr>
<td>MFJS8</td>
<td>3.3.4</td>
<td>216</td>
<td>162</td>
<td>39%</td>
<td>55</td>
<td>28</td>
<td>96.5%</td>
</tr>
<tr>
<td>MFJS9</td>
<td>3.3.3</td>
<td>243</td>
<td>162</td>
<td>50%</td>
<td>56</td>
<td>28</td>
<td>100%</td>
</tr>
<tr>
<td>MFJS10</td>
<td>4.3.5</td>
<td>300</td>
<td>240</td>
<td>25%</td>
<td>66</td>
<td>33</td>
<td>100%</td>
</tr>
<tr>
<td>MFJS11</td>
<td>5.3.5</td>
<td>720</td>
<td>495</td>
<td>45%</td>
<td>99</td>
<td>49</td>
<td>102%</td>
</tr>
<tr>
<td>MFJS12</td>
<td>5.3.7</td>
<td>840</td>
<td>585</td>
<td>43%</td>
<td>106</td>
<td>66</td>
<td>92.7%</td>
</tr>
<tr>
<td>MFJS13</td>
<td>6.3.7</td>
<td>1260</td>
<td>846</td>
<td>49%</td>
<td>131</td>
<td>66</td>
<td>98.5%</td>
</tr>
<tr>
<td>MFJS14</td>
<td>6.3.7</td>
<td>1617</td>
<td>1176</td>
<td>37%</td>
<td>149</td>
<td>78</td>
<td>91%</td>
</tr>
<tr>
<td>MFJS15</td>
<td>7.3.7</td>
<td>1617</td>
<td>1176</td>
<td>37%</td>
<td>149</td>
<td>75</td>
<td>98%</td>
</tr>
<tr>
<td>MFJS16</td>
<td>8.3.7</td>
<td>2184</td>
<td>1512</td>
<td>44%</td>
<td>174</td>
<td>88</td>
<td>97.7%</td>
</tr>
<tr>
<td>MFJS17</td>
<td>8.4.7</td>
<td>3584</td>
<td>2496</td>
<td>43%</td>
<td>219</td>
<td>111</td>
<td>97.3%</td>
</tr>
<tr>
<td>MFJS18</td>
<td>9.4.8</td>
<td>4896</td>
<td>3096</td>
<td>58%</td>
<td>256</td>
<td>123</td>
<td>108%</td>
</tr>
<tr>
<td>MFJS19</td>
<td>11.4.8</td>
<td>7040</td>
<td>4532</td>
<td>55%</td>
<td>308</td>
<td>148</td>
<td>108%</td>
</tr>
<tr>
<td>MFJS20</td>
<td>12.4.8</td>
<td>8832</td>
<td>5376</td>
<td>64%</td>
<td>346</td>
<td>161</td>
<td>115%</td>
</tr>
</tbody>
</table>

*: Feasible solution, IF: Integer Feasible and LB: (Lower bound)

The branch & cut tree may be as large as $2^n$ nodes, where $n$ is the number of binary integer variables. A problem containing only 30 binary integer variables could produce a tree having over 1 billion nodes. As an example, for MFJS10 instance (Table 3), the proposed MILP searches $2^{468}$ fewer nodes which leads to 83% CPU time savings (from 3600 seconds to 600 seconds). The proposed MILP enjoys fewer numbers of constraints by 37.6%. This difference would be even more evident as the size dimensionality of the problem increases.
The performance differences between the two MILPs in key performance measures including number of binary and continuous variables and also number of constraints are easily noticeable. For example, the MILP proposed in [6] possesses 40.2% higher number of binary variables which is the most influential factor in MILPs performance evaluation. The proposed MILP is superior to the MILP in [6] by enjoying 100.3% fewer continuous decision variables. The combined effect of having fewer binary and continuous variables as well as constraints is reflected in the computational time savings of the MILP in [6] and the proposed MILP with 37,308 second and 2,775.09 seconds on the standard benchmark respectively. This reduction is equivalent to CPU time saving of 92.56%.

5. Conclusions and future work

This paper developed an effective position-based MILP model. Solutions obtained by the proposed MILP are computationally superior to previously proposed MILP and is capable of solving the F-JSSP in problems closer to the size of real-life problems. An enhanced position-based Mixed Integer Linear Programming formulation was developed. It proves to be more efficient and accurate when compared with the other MILP model in the literature. The size dimensionality of the proposed MILP model was computed. Two factors are critically important for practitioners while solving optimization problems: 1) solution effectiveness, and 2) speed with which an algorithm achieves the intended results. The proposed MILP was applied to a standard benchmark in the literature [6] in order to demonstrate these factors. It proved its superiority in every aspect.

As an avenue for future research, it is suggested that transportation time be considered between machines in the shop floor because when a job completes its operations, it has to be transported from the current machine to the next scheduled machine.

References:
Allocation of maintenance resources in mixed model assembly systems

W. Guo\textsuperscript{a}, J. Jin\textsuperscript{a}, S.J. Hu\textsuperscript{b}

\textsuperscript{a} Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI, USA
\textsuperscript{b} Department of Mechanical Engineering, The University of Michigan, Ann Arbor, MI, USA

Abstract: Mixed-model assembly systems (MMASs) have been well recognized for their ability to handle increased variety for mass customization. With multiple products to produce, the prioritization of maintenance work-order becomes more crucial and challenging. This paper developed a quantitative method to systematically determine maintenance work-order priorities in a MMAS using on-line production information. The effect of product mix ratio on maintenance decisions is studied in detail. Based on the proposed approach, one can search for an optimal maintenance plan that will lead to improved productivity within the optimization horizon.

Keywords: Maintenance, Mixed Model Assembly, System Value

1. Introduction

In today's marketplace, where customers demand high product variety and short lead time, manufacturing systems have evolved from the traditional dedicated assembly lines to today's mixed-model assembly systems (MMASs). MMASs have been well recognized for their ability to handle increased variety resulting from mass customization.

A MMAS typically consists of basic operation stations and variant operation stations used to perform different tasks required for individual product types. In this way, a variety of products can be produced in a MMAS. Although MMASs can provide better flexibility to meet customer demands with short production lead time [1], the system configurations at an MMAS become highly complex [2]. However, there is little study on the maintenance prioritization decision for MMASs. Since the product values for different variants are different and the production routes for each product can vary significantly in a MMAS, the product variety and their demand ratio will significantly affect the priorities of executing maintenance work-orders, especially when there are more maintenance work-orders than available people or resources. Therefore, the prioritization of maintenance work-order becomes even more crucial and challenging in a MMAS.

The topic on maintenance resources allocation has been widely studied in the recent decades, and the importance of maintenance prioritization has been well recognized [3]. Although there exists lots of research in the literature [4] in addressing the problem of maintenance prioritization, most practical approaches rely on expert experience and knowledge, intuitions, or trial and error, e.g., analytic hierarchy process [5], weighted average system reliability index method [6], etc.. However, in today's industries, the production and assembly systems become very complex and fast-evolving; this makes human judgment and heuristic rules less effective and even insufficient in maintenance prioritization. Therefore, an effective quantitative measure is needed to evaluate the production system performance that can be used for prioritizing maintenance work-orders.

Since the machines are operated in a highly dynamic manner that varies over time, the dynamic status of a production system needs to be considered in the maintenance decision. Such online information can be used to enable a more efficient and cost-effective maintenance decision making [7]. To capture the real-time production information, the buffer contents are very important for maintenance decision making because the finite buffer capacity makes it crucial to prioritize maintenance work in order to achieve a balanced workflow and high profit. Recently, a system value approach was proposed in [8] for maintenance prioritization in a single-product system with the incorporation of online information including buffer contents.

This paper focuses on developing a quantitative method to systematically determine maintenance work-order priorities in a MMAS with the integration of on-line production information (e.g. station failure status and buffer contents). Specifically, this paper extends the existing system value method to a MMAS for assigning priorities to the maintenance work-orders using online information. For this purpose, the effect of the mix ratio between different product variants on the maintenance decisions will also be studied in detail. Based on the proposed approach, one can search for an optimal maintenance plan that will lead to improved productivity within the optimization horizon.

1.1. Review of System Value Approach

Intuitively, the performance of any production system can be directly expressed as the total amount of value that has been added during a certain production time period. The total value added consists of finished products as throughput and unfinished work-in process at the current instant. These intermediate parts in the system also have contributions to the overall system throughput later on. The value of a part, called part value, varies with the changes in system layout and the part’s location in the system. The “no shut off” rule assumes that when one machine in the system fails, all other working machines will continue to process the parts available to them. Therefore, even during the downtime of a machine in the system, work can still be performed on some intermediate parts and values can be added to them.

The system value is defined as the summation of all part values existing in the system at a given moment \( t \). For a simple production system with a single product type being produced, the system value [8] at a given time \( t \) is mathematically defined as

\[
W(t) = \sum_{i=1}^{n} v_i C_i(t)
\] (1)
where \( n \) is the number of stations in the system, \( C_i(t) \) is the number of parts held in station \( i \), and \( v_j \) is the part value for a part in station \( i \). The system value defined in (1) represents the work done by the production system on the parts existing in this system (including the raw parts, finished products, and work-in-progress). Although this approach can quantitatively evaluate the effects of different alternatives with the consideration of online production information, it was established to handle the maintenance prioritization in systems that produce only one product type and it could not be directly applied to MMASs. We need a quantitative measure to effectively consider the multiple product types and their different values in the system in order to make an effective maintenance prioritization decision for MMASs.

2. Methodology

2.1. Proposed System Value Index for MMASs

A quantitative measure for describing the effects of any given maintenance priority decision in a MMAS is defined in this section. The system value approach reviewed in Section 1.1 is extended to multiple product types for a MMAS, which will be used as a performance index in determining maintenance work-order priorities.

At a MMAS, some machines will operate on more than one type of parts and their corresponding buffers will store multiple types of parts. Since the various parts may have different values by passing through different routes, a different set of part values should be used for each product type in the system. Therefore, the extended definition of the system value for a MMAS can be expressed as

\[
W(t) = \sum_{i=1}^{n} \sum_{j=1}^{m} (v_{ij}C_{ij}(t))
\]

where \( W(t) \) is the summation of all part values existing in the system at a given time \( t \), \( n \) is the number of stations in the system, \( m \) is the number of different part types the system is capable to produce, \( v_{ij} \) is the part value of part type \( j \) at station \( i \), \( C_{ij}(t) \) is the buffer content level of part type \( j \) at station \( i \) at time \( t \).

To consider the different profits made by different products that are produced in the system, \( w_j \) is defined by part \( j \)'s unit price \( p_j \) minus its unit cost \( c_j \), i.e., \( w_j = p_j - c_j \), to describe part \( j \)'s contribution to overall system profit. The part value \( v_{ij} \) represents the accumulative added value to part \( j \) up to station \( i \) when it passes through its quickest production path from the beginning of the production process. Therefore, when a part \( j \) is currently at station \( i \), \( v_{ij} \) is a percentage of \( w_j \) since the part \( j \) is an unfinished part at station \( i \). This percentage describes part \( j \)'s completeness during its production in the system, with “1” indicating a finished product and a “0” indicating a raw part. Therefore, the product completeness can be defined as the ratio of \( \sum_{k=1}^{K} T_{k,j} / \sum_{k=1}^{K} T_{k,j} \), where \( T_{k,j} \) is the cycle time of machine \( k \) on part \( j \)'s production path. Part \( j \)'s completeness is increased along the production flow and will reach maximum 100% at the end of the production system. Therefore, the part value \( v_{ij} \) at any station \( i \) can be generally expressed as

\[
v_{ij} = \left( \sum_{k=1}^{K} T_{k,j} / \sum_{k=1}^{K} T_{k,j} \right) \times w_j.
\]

Now, the part value at any station in a MMAS is defined for a given point in time \( t \). When the real-time content level of any part type at any station is obtained, the total system value \( W(t) \) can be calculated and serve as a quantitative measure of the dynamic system status at any given time \( t \).

2.2. Evaluation of Priority Effects through Simulation

Since an analytical evaluation of priority effects of the maintenance actions using the defined index of system value is hard to achieve, computer simulations are used to perform the evaluation in this study.

For a predefined simulation time length \( s \), all system statuses are collected between the starting time \( Ts \) and at the ending time \( Te = Ts + s \). The system statuses are then transformed into the system value format using the methods described in Section 2.1 and yielding system values: \( W(Ts) \) and \( W(Te) \). The difference expressed as \( V = W(Te) - W(Ts) \) describing the value created during the time period \( [Ts, Te] \), is then a quantitative measure representing the effects of a given maintenance order priority. In this work, we used ProModel [9] to realize such discrete-event simulations to track the part flow and collect data.

With sample size \( N \), the output is a vector \( V \) instead of a single value. This vector \( V \) is used to determine the most profitable maintenance priority in optimization by exploring its average value, confidence interval, and distribution.

2.3. Validation of Methodology

In this section, simulation experiments with some simple scenarios are carried out in order to validate the methodology addressed above and discover the importance of mix ratio on the maintenance priority in MMASs. To explore the effects of different maintenance priorities through simulation, a MMAS is established for producing two types of chairs. As shown in Figure 1, chairs A (wheeled chairs) and B (no-wheel chairs) are produced in an assembly system with a mix ratio \( p = A:B \).

![Figure 1. Modules of chairs and their assembly processes](image)

Figure 1. Modules of chairs and their assembly processes

As shown in Fig. 1 and Fig. 2, S1 and S2 machines provide common operations for assembling back rest, lumbar, frame and seat, and S21 and S22 are parallel machines for production line balancing for S2. S4 and S5 provide variant operations for assembling the rod, pneumatic cylinder, wheel base, and wheel for chair A; S6 and S7 provide variant operations for assembling the arm and leg for chair B. When a raw work-piece comes into the system, it goes through S1 and S2 (either S21 or S22), then the intermediate part of Type A goes to S4 and S5, and the intermediate part of Type B goes to S6 and S7, respectively. The
mix ratio \( p \), which indicates the workload for Type A and Type B, is an important parameter affecting maintenance priority decision. To explore the significance of the mix ratio, a set of mix ratio values including 2:8, 3:7, 4:6, 5:5, 6:4, 7:3, and 8:2 are studied (i.e., The Product A’s percentage is equal to 0.2, ..., 0.8, respectively).

The followings are assumed in simulation: (i) Inter-arrival times of starting parts are independent and identically distributed (i.i.d) random variables that follow a Poisson distribution. (ii) There is a finite buffer with the capacity of 50 units located between any two consecutive machines. The incoming queue serves as an infinite buffer and the initial buffer content is zero for all buffers. (iii) Machine processing times are i.i.d random variables that follow a Poisson distribution with the cycle time of 3min. (iv) Each product type has a unit price and a unit cost. To start with a simple simulation case, assume Products A and B to have the same selling price and the same unit cost.

To illustrate the impact of mix ratio, assume S4 and S7 suffer machine failures at almost the same time and follow the same distribution, while all other machines have no failures. One maintenance personnel is in the system to perform all the repairs and maintenance. He starts working on a failed machine immediately after the failure occurs, and he can only work on one machine at a time. An operator “->” is defined to describe the sequence of maintenance work-orders. S4->S7 indicates the maintenance personnel works on S4 first and then on S7.

In simulation, we choose the production length to be almost one work shift, e.g., \( S = 8 \) hours, since the maintenance decision-making is usually a short term decision process. \( Ts = 30\text{min} \) is used as a warm-up period, and sample size \( N = 500 \). We experimented with a case that S4 and S7 suffer machine failures every 30min, and the repair of each machine takes 9min (denoted as Model A), and the results are shown in Fig. 3, including average system value and 95% confidence intervals. Given a certain mix ratio, the maintenance priority leading to a higher value should be preferred since it indicates a more effective system.

Intuitively, maintenance work should be performed on S7 first if we have a lot more requiring of Product B than Product A. This intuition is confirmed in Fig. 3, i.e., when the mix ratio goes towards the extreme case of 2:8, it is more desirable to have the S7->S4 strategy to generate a higher system value. The same conclusion is held for the mix ratio of 8:2 when more products of Type A are to be produced than that of Type B.

Furthermore, Figure 3 shows a narrow intersection where maintenance priority has little influence on the system value. Therefore, this range of the mix ratio serves as the decision threshold for switching the maintenance priority between S7 and S4. The location of the maximum average system value, the slopes of the increasing and decreasing trends, and the intersection may all vary with different downtime scenarios and system configurations.

Given a system configuration and its downtime scenario, the presented methodology can quantitatively measure the efficiency of maintenance works and determines work-order priorities accordingly under various product mix ratios, which will be addressed in detail with a case study in the next section.

3. Case Study

In this section, we will explore the effect of the product mix ratio on the decision threshold with a more realistic scenario, which assumes S4 and S7 to have identical random failure arrival rates following a Poisson distribution. It is also assumed that the maintenance person needs 5min of preparation after he receives a failure signal. Therefore, if S4 and S7 both fail within the 5min preparation time, the maintenance person needs to decide a priority for the work-orders. Otherwise, he just simply follows the first-come-first-serve strategy.

The same system configuration as shown Fig. 2 is used in the simulations. Assumptions on production time, part arrival rate, and buffer capacity remain the same. Four experimental scenarios with different repair times (deterministic/exponentially distributed, 9min/12min) are carried out (denoted as Model A1, A2, A3, and A4, respectively).

Table 1. Simulation conditions for Model A1, A2, A3, and A4

<table>
<thead>
<tr>
<th>Machine Center</th>
<th>Downtime Frequency</th>
<th>Repair Times (4 scenarios)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S4, P(30), decision time: 5min</td>
<td>A1</td>
<td>A2</td>
</tr>
<tr>
<td>S7</td>
<td>9</td>
<td>E(9)</td>
</tr>
</tbody>
</table>

Table 1.

![Figure 4. Results from Model A1](image)

![Figure 5. Results from Model A2](image)

![Figure 6. Results from Model A3](image)

![Figure 7. Results from Model A4](image)

Figure 3. Results from Model A

Figure 3 also shows the decreasing trend caused by the finite buffer and machine blockage during production as the mix ratio moves towards the two extreme cases, 2:8 and 8:2.
3.1. Simulation Results

Figures 4-7 show the simulation results of the average system value and its 95% confidence interval (CI) for Models A1-A4, respectively with 500 replications in the simulations.

The effect of the mix ratio is reflected by the curve trends, location of the peak points, slope of the step-wise lines, and the location and width of the intersection. Furthermore, the results of Model A1 and A2, (or A3 vs. A4) show how the distribution of repair times affects the effect of the mix ratio on the maintenance action. Comparing the results of Model A1 and A3 (or A2 vs. A4), it further indicates how the length of repair times affects the effect of the mix ratio on the maintenance action.

Optimal product mix ratio

Figures 4-7 show that when the mix ratio goes towards the extreme case of 2:8, it is more desirable to have the S7->S4 strategy leading to a much higher average system value, i.e., maintenance work should be performed on S7 first if we need to produce more Product B than Product A. Similarly, S4->S7 strategy achieves a higher average system value when the mix ratio goes towards 8:2.

Interaction between downtime and peak position

Figures 4-7 show that the value of the mix ratio leading to the maximum system value varies at different downtime scenarios. Comparing the peak positions in Fig. 4 to Fig. 6, or Fig. 5 to Fig. 7, it can be seen that S7->S4 has its peak position shifted from 4:6 to 3:7 when repair times are increased. In the S7->S4 strategy, longer repair times cause more A to be blocked and thus push the peak position to shift to a mix ratio level with fewer A to produce. It is noted that the peak positions for S4->S7 also have a tendency to shift to a lower A mix ratio when repair times are increased, although the tendency is shadowed by the large variances when comparing Fig. 5 to Fig. 7.

Sensitivity of the product mix ratio on system value

In Figures 4-7, the slope of increase/decrease represents how significant the mix ratio affects the production performance in terms of the system value. Comparing Fig. 4 to Fig. 6 where the latter comes from a scenario with longer repair times, repair times affect the system performance in both the average system values and confidence intervals by directly affecting buffer contents and machine blockage time. With a longer repair time, the slope becomes steeper, which indicates that the changes in the product mix ratio lead to larger differences in system performance. Hence, the longer repair times we have, the more important the mix ratio becomes.

Decision robustness of maintenance priority

The randomness in the system causes the intersection of two strategies, where there is not much difference on the system value generated by different maintenance priorities. Therefore, the intersection region and its mix ratio range can reflect the robustness of the maintenance decision-making process. Comparing Fig. 4 to Fig. 6, or Fig. 5 to Fig. 7, there is a location shift of the intersection when repair times are increased. Comparing Fig. 4 to Fig. 5, or Fig. 6 to Fig. 7, the intersection becomes wider as the exponentially distributed random repair times brought larger variances, indicating that a wider mix ratio range is covered in the intersection with a more robust decision region. When a mix ratio falls into the intersection region, it is less important to determine which maintenance priority works better. From the above figures, we conclude that with different system configurations, parameters, and downtime scenarios, the intersection may shift and its width may vary.

Summary of results

When the product mix ratio is between 5:5 and 8:2, the S4->S7 maintenance strategy is preferred for all models A1, A2, A3, and A4 since it generates the maximum average system value; while the S7->S4 strategy shows that the maximum system value occurs when the mix ratio is between 2:8 and 5:5. As the repair time increases, the peak positions may shift and the mix ratio plays a more important role in affecting the system performance, and the intersection between the two strategies is also affected. Moreover, different distributions of repair times do not affect the average system values or their curve shapes, but affect the variances and width of the intersection.

4. Conclusions and Future Work

In this paper, the system value method for assigning priority to maintenance work-orders using on-line production information is extended to account for mixed model production and serves as a base for evaluation of maintenance priorities. Then an example assembly system with a set of predefined downtime scenarios is carried out via simulations. The priority effects are evaluated and the simulation results are analyzed in detail. It is also shown that utilizing on-line production information to support maintenance decisions through the system value method provides us much insight into the complex mixed model systems. Future research work will focus on defining the decision thresholds for mixed model systems to help us determine maintenance priority assignment. The proposed method will also be scaled up to apply on problems from larger assembly areas.

Acknowledgement

This work is supported by a grant from the National Science Foundation under CMMI 0825438.

References

Product architecting for personalization

C. Berry, H. Wang, S. J. Hu
Department of Mechanical Engineering, The University of Michigan, Ann Arbor, MI, USA

Abstract: Personalization is an emerging manufacturing paradigm whereby customers can tailor products to their individual needs while maintaining high production efficiency. This paradigm necessitates "Personalized Product Architecting" for determination of customizable/personalizable product modules and cost-effective assembly and fabrication methods. This paper presents a method for identifying product architectures and appropriate manufacturing processes to achieve personalization considering functional utility and manufacturing cost. Results of ergonomic experiments and conjoint analysis are used to build functions relating manufacturability, efficacy, and utility for the architecting. A case study based on a shoe insole is conducted to illustrate the architecting method for personalized products as well as process/system development for data acquisition and manufacturing.

Keywords: Personalization, product architecting, product line design, customization, personalization resolution

1. Introduction

The manufacturing industry has experienced a number of paradigm shifts from craft production, to mass production, and to customization [1] (Figure 1). However, the highly competitive global market and the increasingly diverse customer needs for products necessitate the development of new methods and technologies addressing the extreme diversity of market desires. As such, the next manufacturing paradigm can be envisioned towards Personalization, in which firms will be capable of manufacturing products individually tailored for, and by each consumer [1]-[4]. Personalized products will differ from others by moving past the previous design and production model considering a producer/consumer relationship and instead utilizing a collaborative relationship between consumers and manufacturers. A successful incorporation of the consumer into the product design and manufacturing process is thus the main challenge to realizing Personalization.

![Figure 1. Historical manufacturing paradigm volume-variety relationship [1](image)](image)

Development and manufacturing strategies in three areas are seen as necessary to producing these user-oriented products. The first is an open product platform that allows for product compatibility/interchangeability of its functional features or components with standard mechanical, electrical, and information interfaces. The product platform is taken in its typical context as the set of products based on similar design and components, and the resulting product offerings represent the product family [5]. The architecture of the product platform should be formulated by considering the trade-offs between the perceived needs of the market or individuals, the perceived value of the product family to the market, and the manufacturing costs to the firm of producing the family. The architecting process is thus a key enabler towards successful product development for product customization and personalization. A second need is the development of manufacturing machines and systems for fabricating and assembling personalized product features or modules. The third need is a network or cyber-infrastructure system facilitating the design, review, analysis, and production of these products, incorporating consumers as appropriate in the process. This paper focuses on the first area, i.e., the product architecting, and makes an initial attempt to address related process/system development.

The concepts encompassing product family design, product family architecture, and platform-based product development have been explored extensively by prior research [6]. Two approaches including top-down and bottom-up methods have been observed in industry examples for the implementation of product family design conducted by firms [7]. Enumeration of possible combinations of modules in a product family was also discussed [8]-[9]. Product family architecture was studied to map product functionality to technological feasibility and manufacturability through functional, technical, and physical views [10]. In addition, modularity is noted as an important feature to successful product architectures, and product modularity can be approached in a number of ways [11]-[17].

Extensive work has also been performed in marketing and management research regarding the optimization of a firm's product line and product portfolio planning. A review of the product line design has been provided, noting the ubiquitous need for firms to consider customer input and guidance as early as possible in the design and marketing process [18]. One such method, conjoint measurement, has been applied as a means for producers to provide "transformation of subjective responses into estimated parameters" [19]. An extension to discrete level conjoint analysis is a methodology for developing utility functions for representing a product attribute on a continuous range [20]. The use of conjoint analysis was evaluated as a tool to design product platforms through the simultaneous planning of two products as opposed to designing them individually [21]. Adaptive conjoint analysis was applied in a process structured to enable a "Design by Customers" process which allows customers to find base products relating to their
explicit utility needs [22] in product customization. Two methods based on “buyer’s welfare” and “seller’s welfare” problems were considered for product line selection [23]-[24]. An approach to product portfolio planning was developed to incorporate engineering costs by means of a shared surplus relating customer expectations of quality and cost [25].

Through the reviewed literature, it can be found that there is a lack of (1) an understanding of interactions between personalized and customized attribute levels in product line designs and (2) a method for determining whether and at what resolution(s) to offer personalized options in the context of a product platform. This paper proposes a product architecting method to determine customized and personalized attributes in product modules, and how specific to make the personalized attributes to each consumer. The method is demonstrated for a prototypical personalized shoe product including identification of shoe suitable modules with appropriate personalization resolutions. The experiments for data acquisition in the product architecting and associated manufacturing process for personalization are also discussed.

The organization of this paper is as follows. Section 2 proposes a personalized product architecting method which optimizes product lines with personalized attributes by considering utility of product modules to consumers, product efficacy, and assembly and fabrication cost. The method of personalized product architecting based on a case study of shoes. An example of process and system development for data collection and manufacturing is described in Section 3. Section 4 summarizes the paper and future research.

2. Personalized product architecting

This section develops a method of personalized product architecting. It was envisioned that a personalized product will typically have an open architecture and may consist of three types of modules (Figure 2) [2]: common modules that are shared across the product platform; customized modules that allow customers to choose, mix and match; and personalized modules that allow customers to create and design. Module differentiation is the key toward successful design of personalized products.

![Figure 2. Personalized products may consist of three types of modules [2].](image)

The architecting method is demonstrated by a case study based on casual leather shoes.

2.1 Algorithm of product architecting for personalization

Assume a modular product platform, consisting of multiple modules \( m=1,2,\ldots,M \) and levels for these modules \( l=1,2,\ldots,l_m \). The product line is considered as the selection of these module levels and their relation to the product attributes they address is used to evaluate the line. Each level, including levels representing personalized module designs, can be described by the variable \( x_{ml} \) which is equal to 1 when included in the product line and equals 0 otherwise. The architecting problem is to design the entire product line by determining the choice-menu of module levels so that consumers may choose any combination of these levels to design their product. The resultant line can include common modules having only one level worth offering to the market, customized modules having multiple levels worth, and personalized modules having one level of consumer-specific design.

A concept of personalization resolution is also developed to evaluate manufacturing costs and efficacy for personalized attributes. The concept of resolution can be taken as a level of consumer-specificity at which the manufacturing system for the personalized attribute will be designed and set to operate. Denote the resolution as \( r_{mi} \) representing the resolution for level \( l \) of module \( m \) and can be set to any value within a prescribed continuous range deemed feasible by the firm. The variable cost of producing the personalized module levels can be derived as a function of the resolution, designated as \( c_{ml} = f(r_{mi}) \). Such a variable cost-resolution relationship can be envisioned as an approximation of cost with relation to manufacturing considerations such as the number or complexity of tools, fixtures, machining routines, operations involved.

The module utility as a function of resolution can be determined as follows. The market can be represented as a set of rationally derived segments of consumers with similar tastes and habits. Segments or individuals are denoted as \( i=1,2,\ldots,I \), with segments having population \( q_i \). It is also assumed that a method of conjoint analysis has been used to derive utility values for each of the discrete, non-personalized module levels as a dollar amount representative of the segment or individual’s willingness-to-buy, or reservation price, denoted \( u_{iml} \). A utility-efficacy function of the module for each segment \( u_{iml} = f(c_{ml}) \) can be fitted to data of attribute part-worth against average performances of offerings. Combining with testing results of module performance which reflect resolution-efficacy relationships \( e_{ml} = g(r_{mi}) \), one can obtain a function \( (u_{iml} = f(c_{ml})) \) relating personalization resolution to consumer value for each of the personalization methods and each of the market segments/individuals.

Personalized architecting includes evaluation of the tradeoff between manufacturing costs and perceived value of a product to determine optimal personalization parameters, and how the personalization of the product can influence the offerings of related product features. The decision-making involves choices and preferences of consumers, fixed and variable production costs, and product pricing. A flow chart describing the procedure is shown in Figure 3 noting personalized product considerations in red.

Mathematically, the product architecting can be formulated as a type of welfare problem with the purpose of maximizing a dollar value surplus between the utility the market segments assign the product and the price at which the product can be purchased. In this formulation, each module is considered separately, directly comparing module utility \( (u_{iml}) \) to module price \( (p_{ml}) \), fixed \( (f_{ml}) \) and variable \( (c_{ml}) \) production costs. In contrast to typical integer programming implementations of conventional buyer’s and seller’s welfare problems [23], the personalized resolution based terms \( r_{ml} \) introduced for
considering personalization result in a mixed-integer programming problem. The objective function is taken as the value surplus, the difference between the aggregate market’s willingness-to-buy for the product line and the sales price at which the product line may be offered.

The mathematical representation can be

$$\max \sum_{i=1}^{I} \sum_{m=1}^{M} \sum_{l=1}^{L} (u_{iml} - p_{ml})x_{iml}$$

Constraints for the optimization procedure are shown

$$\sum_{l=1}^{L} x_{iml} = 1, \forall i, m$$

$$x_{iml} \leq y_{ml}, \forall i, m, l$$

$$\sum_{i=1}^{I} \sum_{m=1}^{M} \sum_{l=1}^{L} q_{i}(p_{ml} - c_{ml})x_{iml} - \sum_{i=1}^{I} \sum_{m=1}^{M} \sum_{l=1}^{L} f_{ml}y_{ml} \geq P$$

Constraint 2.1 requires that each customer or segment specify a single level choice for each module for purchase. Constraint 2.2 introduces $y_{ml}$ as a dummy variable indicating that level $l$ of module $m$ will be produced, which represents inclusion of the fixed manufacturing costs $f_{ml}$ associated with production system development and implementation as well as representing the availability of the module level on the choice menu. Constraint 2.3 then considers the profit of the firm over the entire product family, by considering the sum of the profits (sale price minus variable production cost) for all incremental module units produced and subtracting the total fixed manufacturing costs associated with all modules in production. $P$ can be chosen as different levels for the total profit to the firm of the product line, and thus the firm can investigate market satisfaction for differing profit targets.

Linearization should be performed to find solutions to the original objective function, by simplification of the non-linear constraint 3 of the original problem into a number of equivalent linear constraints. Alternatively, a simplified formulation can be a direct comparison between the total value yielded by consumer’s selection of module offerings and the total cost necessary to produce the line. A difference in the formulation is that the task of defining the sale price at which the modules are offered is neglected, instead simply comparing reservation price to variable and fixed production costs. In this case, the objective function can be

$$\max \sum_{i=1}^{I} \sum_{m=1}^{M} \sum_{l=1}^{L} (u_{iml} - c_{ml})x_{iml} - f_{ml}y_{ml}$$

The objective function is subject to constraints 2.1 and 2.2. A numerical example of implementation of the formulation to a personalized module offering is shown in Section 2.2.

2.2 Case study: Architecting for personalized shoes

This section presents a numerical case study for personalization of casual leather shoes to demonstrate the proposed architecting algorithm. As shown in Figure 4, a leather shoe consists of an “aesthetics” module (i.e., upper) and a functional (protection) module which distributes the user’s weight in a manner that is comfortable and not detrimental to the biomechanics of the user during walking or standing. An insole is the most critical component in the functional module. It is often a combination of a thin piece of paper board or other fiber and usually is less costly. Therefore, the insole has the potential to be personalized to offer the maximum utility to consumers economically. For simplicity of illustration, this case study only focuses on optimal selection of customized and personalized levels of the leather shoe insole.

Assumptions regarding the market characteristics and segmentation are made to provide a sample problem consisting of eight market segments with equal population of 100,000 consumers, and with assumed analysis performed as to their preference for eight discrete module offerings for an insole. The module levels reflect common options for insole products, and all utility, cost, and performance values are based on dollar-value assumptions for the purpose of carrying out an example architecting procedure. Efficacy values are based on assumed averages for performance of the discrete module offerings using typical peak plantar pressure measurements. Table 1 shows the defined attribute levels, the utility and efficacy data from conjoint analysis, and fixed cost.

<table>
<thead>
<tr>
<th>$f_{ml}$</th>
<th>$c_{ml}$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>50.2</td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td>48.75</td>
<td></td>
</tr>
<tr>
<td>45000</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>65000</td>
<td>52.5</td>
<td></td>
</tr>
<tr>
<td>20500</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>45000</td>
<td>55.75</td>
<td></td>
</tr>
<tr>
<td>50000</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>45000</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>50000</td>
<td>55</td>
<td></td>
</tr>
</tbody>
</table>

Figures 5a-b provides efficacy against perceived utility to approximate each segment’s utility-efficacy function to measure
of the value associated with performance as well as provided efficacy against personalization resolution. The physical explanation of the resolution will be discussed in Section 3. For this implementation case both functions are simplified as linear relationships.

Figure 5. (a) Insole module utility vs. efficacy (b) Efficacy vs. resolution

The optimization program (3) and (2.1-2) for the insole module example was implemented to provide direction for satisfying a market with optimal value at low cost. Results in Table 2 shows the optimal menu to be a combination of discrete module levels representing the sheet EVA, molded EVA, and soft molded EVA with a personalized module representing production at a resolution of 3 mm \( r_{m} = 3, r_{n} = 0 \). The personalized module is offered to four market segments \( i=2, 3, 4, 8 \) whose utility-efficacy relationships signify their increased value for personalization. The yielded results signify the concept behind inclusion of personalized and customized module variants coexisting for a product. The personalized insole is currently offered to the market segments having steeper utility-efficacy functions, noting their higher value for product efficacy, and that they may be willing to cover higher production costs. Sensitivity analysis could be performed on expanded data sets for these or other case studies to examine how the architecting process responds to market preference changes or manufacturing costs.

Table 2 Optimization results

<table>
<thead>
<tr>
<th>( i )</th>
<th>( \tau_{y} )</th>
<th>( \tau_{x} )</th>
<th>( \tau_{z} )</th>
<th>( \tau_{y} )</th>
<th>( \tau_{x} )</th>
<th>( \tau_{z} )</th>
<th>( \tau_{y} )</th>
<th>( \tau_{x} )</th>
<th>( \tau_{z} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</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>
</tr>
<tr>
<td>2</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>
</tr>
<tr>
<td>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</td>
<td>0</td>
</tr>
<tr>
<td>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>
</tr>
<tr>
<td>5</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>
</tr>
<tr>
<td>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</td>
</tr>
<tr>
<td>7</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>
</tr>
<tr>
<td>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>
</tr>
<tr>
<td>9</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>
</tr>
<tr>
<td>10</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>
</tr>
</tbody>
</table>

3. Data acquisition and manufacturing for personalization

Using the example of personalized shoe insole, this section discusses process/system development for manufacturing and data collection in the product architecting. The process includes anthropometric data acquisition, data manipulation, manufacturing methods, and performance evaluation.

3.1 Anthropometric Data Acquisition

Data collection began with a desktop laser scanner attempting to capture plantar foot surfaces when supported in a stationary position. Orthotic impression foam was used to minimize collection error due to subject movement. The laser-scanning was performed using a Brown and Sharpe coordinate measuring machine with an integrated Nikon Metrology single-beam red-light laser scanner. Data collection in this manner resulted in 3-dimensional data in stereolithography format typical of rapid prototyping and scanning software, which represent geometries as a number of triangular mesh surfaces. Filtering algorithms embedded in the scanning software were used to reduce redundant points to yield reasonable files sizes (>700,000 surface triangles). A typical data set representative of filtered scan data is shown in Figure 6a.

3.2 Data Manipulation

The data sets were then modified through alignment, editing, and sub-sampling procedures. Alignment and point-cloud editing were performed in the CAD program Rhinoceros 4.0, and sub-sampling in 3-Matic point-cloud/meshing software. The alignment was performed to orient the foot along the x-, y-, and z-coordinate axes such that any tilting or rotation of the foot during impression is corrected. Further data reduction was performed after the observation that the relatively high-density point-cloud data resulted in extremely rough contours of the foot. Sub-sampling of the data (<30,000 surface triangles) was then performed with the intent of producing a much smoother surface still representative of the user’s geometry, an example of which can be seen in Figure 6b.

Figure 6. (a) Footprint measurement (b) Smoothed User Geometry

3.3 Manufacturing methods for Personalized Insoles

The obtained plantar surface of the foot is sectioned at intervals to provide a contoured surface for the insole (Figure 7a). These contours can then be machined automatically from foam sheet material using a laser-cutting machine to produce a stack-up to be joined by adhesives similar to an existing rapid-prototyping technology for laminated paper models, Laminated Object Manufacturing (LOM) [26]. Alternatively, the surfacemating design could be CNC machined similar to custom-orthotics production.

Prototyping of example insoles was performed on a number of data sets and at varying contour thickness of 1.5, 3, and 4.5 mm. Materials used include varying firmness EVA foams. The sectioned curves from obtained data were used to machine insoles on a Universal X2-660, 120 Watt carbon dioxide laser cutter. Machine programming was generated using BobCAD-CAM software. A resulting surface, generated at a resolution of 1.5 mm, is shown for a data set in Figure 7b.

Figure 7. (a) Personalized insole design (b) Prototype Personalized Insole

3.4 Performance Evaluation of Personalized insoles

Quantification of the improved product performance yielded by insole personalization is to evaluate improvement over non-personalized insole products as well as determination of the resolution-efficacy relationship for the process.
Performance measurement has begun with the use of the F-Scan pressure measurement system (TekScan, Inc.) to quantify standing pressure distributions resulting from foot-insole contact (Figure 8). The F-Scan system consists of thin sensor films than can be used in-shoe to measure forces, pressure distributions, and time-variable gait characteristics and is typically used in podiatric applications for diagnosing foot or gait problems and enhancing the efficacy of intervention procedures.

Figure 8. F-Scan Pressure Measurement System
(http://www.med.nyu.edu/rehabengineering/facilities/stressmeasure.html)

Preliminary data are shown in Figure 9 for a personalized insole produced at 1.5 mm resolution and a comparative thickness flat EVA sheet, observing the increased contact area provided by the personalized insole as well as the drastically reduced high-pressure concentration at the heel location. Data collection will help define the resolution-efficacy function.

Figure 9. Pressure Distribution (a) EVA Sheet (b) Personalized Insole

The layered, fully conforming design facilitates the definition of the personalization resolution as the height between layers of the insole. Resolution in this sense can be related to material handling costs and machining routine operation times, with smaller resolutions requiring increased machining times as well as potentially increased stock material waste. Resolution for milling the surface can be defined relating tool-path parameters for step-down or provided surface roughness. The exterior of the insole can be assumed to be produced with the shape of the last to ensure a good fit between the formed upper and the insole, reducing issues between the newly introduced personalized module and the interfaces between it and other product modules.

4. Summary and future work
Personalized Products and Production represents the pinnacle of manufacturing paradigm evolution towards increasing consumer satisfaction through the availability of high efficacy personalized products offered at affordable prices. In this paper, Personalized Product Architecting is implemented as the first step for the personalization paradigm. The architecting process begins by considering a product as a decomposition of its components into a modular platform and relating the modules to their product functionalities. Through the conjoint analysis, a product’s value for personalization can be identified by examining market preferences for existing products based on their efficacy. Personalized module designs are implemented with a measure of personalization resolution defining the degree of specificity to each individual’s tastes and preferences for product functionality. Fixed and variable manufacturing costs are determined for the discrete module designs of the product, and manufacturing costs are estimated as a function of resolution for personalized modules. By comparing the perceived market value and resulting costs for the overall product family, the best combinations of common, customized and personalized product modules can be determined for offering to the market. Different from the product architecting for customization, the framework results in a mixed-integer programming problem solving for both continuous and binary decision variables with respect to module variants included in a product line.

The architecting framework is demonstrated using a case study of a product line for a personalized casual shoe with options of personalized insole offered to consumers. Based on the shoe insole example, the paper also describes data acquisition and manipulation, manufacturing methods, and evaluation.

Further work on the shoe case study will be accomplished in a number of areas. First, the architecting and manufacturing for the personalized insole must be validated through the suggested biomechanics studies for evaluating the resolution-efficacy function. Techniques for incorporating this insole into the shoe product can then be considered, to examine including this option into the last library of a shoe producer and evaluate the production processes affected. The case study can then be examined for the flexibility it requires in the actual processes for product design and automated manufacturing. The introduction of anthropometric data into the product design requires application of efficient data manipulation and integration into the assumed computer-aided design model of the product, as well as rapid generation of machining programs and assembly instructions. Lastly, efficient linearization methods or heuristic approaches should be developed for solving the mixed-integer problem. The developments of these methods are necessary across all other personalizable modules such as the shoe last and upper as well as other products such as chairs and helmets.

5. Reference


Automatic creation of virtual manikin motions maximizing comfort in manual assembly processes

R. Bohlin, N. Delfs, L. Hanson, D. Högberg, J.S. Carlson

a Department of Product and Production Development, Chalmers University of Technology, Gothenburg, Sweden
b Fraunhofer-Chalmers Research Centre for Industrial Mathematics, Gothenburg, Sweden
c Virtual Systems Research Centre, University of Skövde, Skövde, Sweden

Abstract: Effective simulation of manual assembly operations considering ergonomic load and clearance demands requires detailed modeling of human body kinematics and motions, as well as a tight coupling to powerful algorithms for collision-free path planning. The focus in this paper is a unified solution that automatically creates assembly motions for manikins taking kinematic constraints, balance, contact forces, collision avoidance and comfort into account.

The manikin used in this work has 162 degrees of freedom - six exterior fictitious joints determine the position of the lower lumbar and the remaining ones are interior joints. The inverse kinematic problem leads to an underdetermined system allowing us to pick a solution that maximizes a scalar valued comfort function. The comfort function offers a generic way to give preference to certain poses while avoiding others, typically by considering joint limits, forces and moments on joints, and magnitude of contact forces. In order to avoid collisions, poses close to collision are penalized. The method is implemented and demonstrated on two challenging assembly operations taken from the automotive industry.

Keywords: Advanced biomechanical models, Ergonomics, Optimization Algorithm.

1. Introduction

Although the degree of automation is increasing in manufacturing industries, many assembly operations are performed manually. To avoid injuries and to reach sustainable production of high quality, comfortable environments for the operators are vital [7]. Poor station layouts, poor product designs or badly chosen assembly sequences are common sources leading to unfavourable poses and motions. To keep costs low, preventive actions should be taken early in a project, raising the need for feasibility and ergonomics studies in virtual environments long before physical prototypes are available.

Today, in the automotive industries, such studies are conducted to some extent. The full potential, however, is far from reached due to limited software support in terms of capability for realistic pose prediction, motion generation and collision avoidance. As a consequence, ergonomics studies are time consuming and are mostly done for static poses, not for full assembly motions. Furthermore, these ergonomic studies, even though performed by a small group of highly specialized simulation engineers, show low reproducibility within the group.

Effective simulation of manual assembly operations requires detailed modeling of human body kinematics and motions, as well as a tight coupling to powerful algorithms for collision-free path planning. The focus in this paper is a unified solution that automatically creates assembly motions for manikins taking kinematic constraints, balance, contact forces, collision avoidance and comfort into account. This paper extends the work presented in [5].

2. Related work

The problem of generating collision-free motions for digital humans has been extensively studied in computer animation and robotics. See for example [6,10,14,20] for approaches aiming at reachability and assembly tasks. Many manikin models presented in the literature use a foot or the lower torso as a root of a hierarchical model of links. Usually, the positioning of the root is handled differently than for the other links. In [3] and [12], the root has a fixed position with respect to a global coordinate system, which restricts the inverse kinematic calculations in an undesired way. Models without the ability to change root usually set the root to the lower torso [18]. This, however, gives complications for the balance control.

One way to handle the fixed root is to introduce a high level algorithm that changes the tree hierarchy and/or repositions the root when needed [2,18]. In [15], an external root with six degrees of freedom is implemented with the purpose of matching motion capture data, but the inverse kinematic calculations do not utilize this extra freedom fully.

3. Manikin model

In this section we present the manikin model we base our work upon, and the method we propose to solve the balance, positioning, contact force, collision avoidance and comfort problems simultaneously in a unified way. The work is based on [5], and parts of the method are related to the resolved motion-rate method presented in [2].

3.1. Kinematics

Our manikin model is a simple tree of rigid links connected by joints. Each link has a fixed reference frame and we describe its position relative to its parent link by a rigid transformation \( T(\mathbf{q}) \). Here \( \mathbf{q} \) is the value of the joint between the link and its parent. Each joint has one degree of freedom, so a wrist, for example, is composed by a series of joints and links.

To position the manikin with respect to some global coordinate system, we introduce an exterior root at the origin and a chain of six additional links denoted exterior links – as opposed to the interior links representing the manikin itself. The six exterior links, with three prismatic joints and three revolute joints respectively, mimic a rigid transformation that completely
specifies the position of the lower lumbar. In turn, the lower lumbar represents an interior root, i.e., it is the ancestor of all interior links. Note that the choice of the lower lumbar is not critical. In principal, any link could be the interior root, and the point is that the same root can be used through a complete simulation -- no changes in the tree hierarchy are needed.

Now, for a given value for each of the joints, collected in a joint vector \( \theta = [\theta_1, \ldots, \theta_n]^T \), we can calculate all the relative transformations \( T_1, \ldots, T_n \), traverse the tree beginning at the root and propagate the transformations to get the global position of each link. We say that the manikin is placed in a pose, and the mapping from a joint vector into a pose is called forward kinematics. Furthermore, a continuous mapping \( t \rightarrow \theta(t) \), is called a motion.

### 3.2. Balance, contacts and positioning constraints

In order to facilitate the generation of realistic poses that also fulfills some desired rules we add a number of constraints on the joint vector. These kinematic constraints can be defined by a vector valued function \( g \) such that

\[
g(\theta) = 0
\]  

must be satisfied at any pose. Finding a solution to (3.1) is generally referred to as inverse kinematics.

One important part of \( g \) ensures that the manikin is kept in balance. The weight of its links and objects being carried, as well as external forces and moments due to contact with the floor or other obstacles, must be considered. The sums of all forces and moments are

\[
g_{\text{cm}}(\theta) = m g + \sum_{j=1}^{n} f_j
\]

\[
g_{\text{tor}}(\theta) = m \times (m g) + \sum_{j=1}^{n} (p_j \times f_j + t_j)
\]

where \( m \) is the total body mass, \( \mathbf{g} \) is the gravity vector, \( m \) is the center of mass, \( f_j \) and \( t_j \) are external force and torque vectors at point \( p_j \), \( j = 1, \ldots, M \). Note that the quantities may depend on the pose, but this has been omitted for clarity.

In general, external forces and moments due to contacts are unknown. For example, when standing with both feet on the floor it is not obvious how the contact forces are distributed. In Section 3.5, the unknown forces and moments are included when finding inverse kinematics solutions.

Another common type of constraints restricts the position of certain links, either relative to other links or with respect to the global coordinate system. Typical examples of such constraints keep the feet on the floor, the hands at specific grasp positions and the eyes pointing towards a point between the hands. Positioning constraints are also expressed in the form (3.1).

### 3.3. Comfort function

When generating realistic manikin poses and motions, it is essential to quantify the ergonomic load. To do so, we introduce a scalar comfort function

\[
h(\theta, f)
\]

capturing desired ergonomics aspects. The comfort function is a generic way to give preference to certain poses while avoiding others. Typically \( k \) considers joint limits, forces and moments on joints, magnitude of contact forces etcetera. The joint loads are key ingredients when evaluating poses from an ergonomics perspective [19]; Research shows that real humans tend to minimize the muscle strain [16], so by normalizing the load on each joint by the muscle strength good results can be achieved.

### 3.4. Collision avoidance

The contacts described in Section 3.2 are intended ones, and contribute to the force and moment balance. Other contacts, from now on denoted by collisions, are undesired. The comfort function offers a convenient way to include a simple, yet powerful, method penalizing poses close to collision. In robotics this method is generally known as Repulsive Potential [9,11]. The idea is to define a barrier of thickness \( d_c \), say, around the obstacles decreasing the comfort towards -infinity near collision. One possibility is to add the following to the comfort function:

\[
h_{\text{coll}}(\theta) = -\eta \begin{cases} \frac{1}{2} \left( \frac{d(\theta)}{d_0} \right)^2 & \text{if } d(\theta) \leq d_0 \\ 0 & \text{if } d(\theta) > d_0 \end{cases}
\]

where \( d(\theta) \) is the minimum distance between the manikin and the obstacles, and \( \eta \) is a positive scaling factor.

This method does not address the problem of escaping an already occurring collision. The idea is merely that if the manikin starts in a collision-free pose, then the repulsive potential prevents the manikin from entering a colliding pose.

### 3.5. Unified solving

In this section we present a method for a unified treatment of balance, contact forces, position constraints, ergonomics and collision avoidance. Often in practice, the number of constraints is far less than the number of joints of the manikin, allowing us to consider ergonomics aspects and maximizing comfort when choosing one solution.

Let \( x' = [\theta^T, f^T] \) be the unknowns, i.e. the joint variables and the unknown forces and moments. Then we formulate the problem as follows:

\[
\begin{align*}
\text{maximize } & h(x) \\
\text{while } & g(x) = 0
\end{align*}
\]  

We solve this nonlinear optimization problem by iterative linearization; Let \( x_c \) be the current state. Then, for small \( \Delta x \)

\[
g(x_c + \Delta x) \approx g(x_c) + J(x_c) \Delta x \]

where \( J(x_c) = \partial g / \partial x \mid_{x_c} \) is the Jacobian at \( x_c \). In order to satisfy \( g(x_c + \Delta x) = 0 \) while increasing \( h(x) \), we project \( \nabla h(x) \) onto the null space of \( J \), thus taking the step

\[
\Delta x = -J^* g(x_c) + \lambda (I - J^* J) \nabla h(x)
\]  

(3.3)

where \( J^* \) is the Penrose pseudo inverse and \( \lambda \) is a scalar chosen such that the step stays within a reasonable domain where the linearization is assumed to be valid.
The step $\Delta x$ is also truncated so that each value stays within its joint range, and that the manikin remains collision-free. The latter is guaranteed if $|\Delta x| < K d$ where $d$ is the minimum distance to the obstacles and $K$ an appropriate constant.

A consequence of the iterative procedure and the limited step size is that all poses in the sequence $x_1, x_2, \ldots$, as well as poses linearly interpolated in between pairs of successive points, are feasible – a property we will use later. Thus, it is not a matter of global optimization, rather finding a feasible sequence of poses to a local comfort maximum.

### 3.6. Scaling issues

Of the many solutions to $0 = J(x) \Delta x + g(x)$, the pseudo inverse, through $g(x) + J(x) = 0$ in (3.3), gives the one with minimum $L_2$ norm. Consequently, there is a weighting choice between prismatic joints, revolute joints, forces and moments. For a manikin, a typical length scale is one meter, and force and moment ranges correspond to the total weight of the manikin.

### 3.7. Following paths

Our framework for defining kinematic constraints, comfort and for finding stationary solutions to (3.2), can now be extended to include motions. Let $g(x,t)$ and $h(x,t)$ be time dependent constraint and comfort functions respectively, and $t_1, t_2, t_3, \ldots$ an increasing sequence of (fictitious) time samples. Then, for a known initial state $x(t_0)$, we get a motion $x(t), x(t_1), x(t_2), \ldots$ by letting $x(t_k)$ be the solution to

$$\begin{cases} \text{maximize } h(x,t_k) \\ \text{while } g(x,t_k) = 0 \end{cases}$$

for the initial solution $x(t_0), k = 1, 2, \ldots$.

Note that even if the sample density tends to infinity, the mapping $t \to x(t)$ may be discontinuous for two reasons: (i) $g(x,t)$ and $h(x,t)$ may be discontinuous, e.g. when constraints are added or deleted, (ii) even if $g(x,t)$ and $h(x,t)$ evolves continuously, local maxima may disappear causing jumps to another local maximum. However, the intermediate steps in the iterative solving of (3.2) are always feasible, and so by tracing and inserting the solver steps into the sequence whenever needed, a sequence of arbitrarily small increments can be obtained.

### 4. Results

The applicability of the manikin model and the unified solving is demonstrated by two real assembly cases from the automotive industry. The goal is to generate feasible motions for the manikin assembling objects into a car. The solutions were generated in two phases; First, collision-free assembly paths for the objects were created using the automatic path planner in the software package Industrial Path Solutions (IPS) [8]. See also [9,11] for general overviews of the path planning problem. Then the hands were attached to certain grasp positions forcing the manikin to follow the object along the assembly path, while the ergonomic status was monitored and assessed based on an ergonomic standard, in our case SES [17].

Common in both cases is that the feet were allowed to slide on the floor and rotate around the z-axis – the remaining three degrees of freedom were locked. This gives comfortable foot positions along the motion, allowing for an appropriate choice if fixed feet are desired. Moreover, the comfort function penalized joint values near their limits and high joint moment.

#### 4.1. CEM Box

The first case is a one-hand assembly of an electronic device called CEM Box. The other hand gets support by the car body by a constraint locking five degrees of freedom, allowing the palm to rotate around the surface normal. When moving the CEM Box along the assembly path, the kinematic constraints $g(x,t)$ and the comfort function $h(x,t)$ locks the right hand to the moving object while keeping the balance, maximizing comfort and avoiding collisions. The manikin’s collision avoidance includes static geometry as well as the CEM Box, with the exception of the hands that are forced to be in contact.

The initial path for the CEM Box, as well as the manikin’s motion, was computed in about 30 seconds each on an ordinary PC. Snapshots are shown in Figure 1. The risk assessment according to SES [17] is low. The most critical part is the position of right hand.
4.2. Tunnel bracket
The second case is a two-hand assembly of a tunnel bracket. Note that the initial path planning problem for the tunnel bracket is very challenging due to the handbrake lever, the gearshift and the dashboard, allowing no more than a few millimeters of clearance during a long part of the assembly path. As a result, the tunnel bracket must make a complicated motion with numerous adjustments that of course also propagate to the manikin.

The initial path for the tunnel bracket was computed in about 90 seconds, and the manikin’s motion shown in Figure 1 was computed in about 30 seconds. The SES identifies a possible risk due to the position of the back and the hands.

5. Discussion
Even though pose prediction is a commonly used term, the primary goal when simulating assembly motions is, in many cases, to prove the existence of at least one feasible motion. Then the human is likely to find one as well, perhaps even a better one. If, however, no ergonomically acceptable motion can be found for the manikin, then actions must be taken.

In general, assembly studies and ergonomics evaluations are relevant to do for each potential operator. In [4,13], the authors suggest methods for selecting a set of manikins representing anthropometric diversity in a population, and together with the process presented here, a set of manikins can be analyzed automatically. In order to find comfortable poses, it is important to add as few kinematic constraints as possible, and thereby give the solver the largest possible freedom. In the test cases, for example, both feet are allowed to slide and rotate on the floor.

The decoupled approach for automatic collision-free motion planning presented here, where paths for the objects to be assembled are generated first and the collision avoiding manikin is following the paths in a second phase, has its limitations. (i) the initial object paths may be impossible to follow for the manikin due to its kinematic structure, (ii) even though the object is collision-free, there may not be clearance enough, e.g. for the manikin’s hands, (iii) when the manikin follows an object path, it may get caught in traps formed by the geometry and the repulsive potential. One of our future goals is to bridge this gap and take the manikin into account when planning the initial path.

6. Conclusion
By introducing a chain of exterior links and joints that positions the lower lumbar, we are able to reposition the manikin without a series of re-rooting operations, and we maximize the comfort using as few kinematic restrictions as possible.

Furthermore, we present a unified solution that automatically creates assembly motions for manikins taking kinematic constraints, balance, contact forces, collision avoidance and comfort into account. The framework is general, and the comfort function could be extended to include other information in order to generate even more favorable poses.

The applicability in practice is demonstrated by two challenging assembly operations taken from the automotive industry.

4. Acknowledgement
This work was carried out within The Swedish Foundation for Strategic Research (SSF) ProViking II program and the Wingquist Laboratory VINN Excellence Centre, supported by the Swedish Governmental Agency for Innovation Systems (VINNOVA). This work is part of the Sustainable Production Initiative and the Production Area of Advance at Chalmers University of Technology.

References
[17] Scania Ergonomics Standard (SES) - Produktion utg.1 2009-03.
An assembly decomposition model for subassembly planning considering imperfect inspection to reduce assembly defect rates

J. Ko \textsuperscript{a,b} and E. Nazarian \textsuperscript{b}
\textsuperscript{a} Ajou University, Korea, \textsuperscript{b} University of Nebraska-Lincoln, Lincoln, NE, USA

Abstract: The assembly decomposition is to divide the assembly to subassemblies that are to be joined in the final assembly processes. The assembly decomposition decision strongly affects the effectiveness of a product assembly in terms of quality, sequence and supplier selection. This paper presents an assembly-decomposition model to improve product quality. Mixed-integer programming is used to partition the liaison graph of a product assembly. The mixed-integer programming model takes into account the defect rates in components and assembly tasks. The defect rate of the final assembly product is to be minimized considering type II errors in subassembly inspection. A numerical example is presented to demonstrate the methodology, and this numerical study shows that assembly decomposition strongly affects the final assembly defect rate. The developed assembly decomposition method is expected to enhance the decision making in assembly planning.

Keywords: Assembly, Decision making, Design, Disassembly, Quality

1. Introduction

The assembly decomposition problem is to divide the product assembly to constituting subassemblies. These subassemblies are produced individually in the subassembly lines of a final assembler or by the suppliers for the final assembler. These subassemblies are joined together to make the final product in the main assembly line of the final assembler.

The subassembly decision affects the quality of the final assembly. Each subassembly has unique quality characteristics depending on the constituting components and their joining processes. In addition, because some subassemblies are inspected before the final assembly, different subassembly decisions will lead to inspection of different subassemblies. Thus, subassembly decisions will affect how the quality characteristics of components and joining processes propagate to the final assembly.

In spite of the extensive research on assembly design and quality, only limited number of research papers studied the effect of assembly decomposition on assembly quality [1-6]. Moreover, little research directly considered the inspection error with the assembly decomposition. This paper presents an assembly decomposition model to minimize assembly defect rates considering imperfect inspection.

This paper presents an assembly decomposition methodology based on a graph-theoretic method. This paper uses the liaison graph of a product assembly to represent the assembly and subassemblies. This study uses mixed-integer programming (MIP) to model the partition of the assembly graph. The MIP model considers the defect rates of components and assembly tasks. The model also considers type II errors in subassembly inspection. The objective function of the model is to minimize the defect rate of the final assembled product. A numerical example is presented to demonstrate the effect of assembly decomposition on the final assembly defect rate.

2. Definition of the assembly decomposition problem

In this study, the final assembly, subassemblies and subassembly structure are represented by graphs. The assembly is represented as liaison graph $G = (V, E)$ in which vertices ($V$) correspond to the components in the assembly and edges ($E$) corresponds to the assembly tasks to join the components. For example, see Figure 1. A subassembly is defined as a collection of components and all the assembly tasks between them. In the graph theoretic representation, a subassembly is equivalent to a connected subgraph that is separated by edge cuts from the other parts of the assembly graph. For instance, in Figure 1, nodes $j$ and $m$ and edge $(j, m)$ represent a subassembly $s_2$. The edges that are cut to generate subgraphs (subassemblies) represent assembly tasks to be performed for joining the subassemblies represented by the subgraphs.

This study assumes that the assembly decomposition is performed through three layers. The top layer represents the final assembly, the mid-layer represents assembly decomposition to subassemblies, and the bottom layer represents the components of each subassembly including single-component subassemblies. In Figure 1, the final assembly is denoted by $s_0$ and is decomposed into subassemblies $s_1$ and $s_2$ in $i = 1$.

In this study, subassembly decomposition is conducted by generating feasible partition sets of the graph representing the subassembly. In each decomposition, a set of subassemblies are generated by satisfying connectivity, precedence and other subassemblies constraints. The maximum number of subassemblies is given. Precedence relations of assembly tasks are also known in advance. Dimensional quality and degrees of freedom are not considered in this model. For these issues, see [1-2].

3. Subassembly defect rate and effect of imperfect subassembly inspection

Assume that the defect rates are known for the components and assembly tasks joining the components. Also assume that the component and assembly task defects are
independent of each other. Although a more detailed analysis can be conducted using reliability modeling techniques such as

\[
\delta_i = \text{Defect rate of component } i; \text{ a real number } \in [0, 1) \\
\delta_{ij} = \text{Defect rate of assembly task } (i, j); \text{ a real number } \in [0, 1)
\]

Sets

\[
V = \text{Set of all components (represented by vertices) in the final assembly} \\
E = \text{Set of all assembly tasks (represented by edges) in the final assembly; } (i, j) \text{ if there is an assembly task between components } i \text{ and } j \text{ in } V \\
A = \text{Collection of all subassembly sets, each of which can be defined by feasible assembly decompositions of the final assembly; union of all } U's
\]

Variables

\[
u_i^s = \text{Inclusion of component } i \text{ in subassembly } s; \text{ binary;} \\
u_i^s = 1 \text{ if and only if subassembly } s \text{ includes component } i, 0 \text{ otherwise.} \\
\gamma_{ij}^s = \text{Inclusion of performed assembly task } (i, j) \text{ in subassembly } s; \text{ binary;} \\
\gamma_{ij}^s = 1 \text{ if and only if subassembly } s \text{ includes performed assembly task } (i, j), 0 \text{ otherwise.} \\
p^s = \text{Nonempty subassembly; binary variable; 1 if and only if subassembly } s \text{ is nonempty, 0 otherwise.} \\
r_i^s = \text{Imaginary sink component for connectivity counting purpose; binary; 1 if and only if component } i \text{ is a sink at subassembly } s, 0 \text{ otherwise.} \\
a_i^s = \text{Imaginary flow for connectivity counting purpose; real; } a_i^s \text{ represents imaginary flow from component } i \text{ to } j \text{ at subassembly } s. \\
\gamma^s = \text{Indication of a single-component subassembly (vertex singleton); binary}; 1 \text{ if subassembly } s \text{ has only one component, 0 otherwise.} \\
\Delta(s) = \text{Defect rate of subassembly } s; \text{ a real positive number } \in [0, 1)
\]

4.2 Optimization formula

The basic structure of the optimization model is shown as follows:

\[
\begin{align*}
\text{Min } & \Delta(s_0) \\
\sum_s u_i^s & = 1 \quad \forall i \in V \\
y_{ij}^s & \leq u_i^s \quad \forall i, j \in V : (i, j) \in E \\
y_{ij}^s & \leq u_j^s \quad \forall i, j \in V : (i, j) \in E \\
y_{ij}^s & \geq u_i^s + u_j^s - 1 \quad \forall i, j \in V : (i, j) \in E \\
\sum_i u_i^s & \leq B \cdot p^s \quad \forall s \in A \\
\sum_i u_i^s & \geq p^s \quad \forall s \in A \\
\sum_i p^s & \leq \text{Sub}_{\text{max}} \quad \forall s \in A \\
\sum_s p^s & \geq \text{Sub}_{\text{min}}
\end{align*}
\]

4.1 Nomenclature

<table>
<thead>
<tr>
<th>Indexes and Parameters</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B ) = A large number</td>
<td></td>
</tr>
<tr>
<td>( \text{Sub}_{\text{min}} ) = Minimum number of subassemblies</td>
<td></td>
</tr>
<tr>
<td>( \text{Sub}_{\text{max}} ) = Maximum number of subassemblies</td>
<td></td>
</tr>
<tr>
<td>( i, j ) = Component</td>
<td></td>
</tr>
<tr>
<td>( s ) = Subassembly</td>
<td></td>
</tr>
<tr>
<td>( s_0 ) = Final assembly</td>
<td></td>
</tr>
<tr>
<td>( \beta ) = Probability of accepting a defective subassembly in inspection (type II error); ( \beta \in [0, 1] )</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Illustration of the layers and subassemblies in assembly decomposition. In (a), circles represent components, and straight lines between circles represent assembly tasks. Subassemblies are represented by shaded areas. In (b), rectangles represent subassemblies and circles at the bottom represent components. In (b), the highest layer \( l = 0 \) indicates the final assembly. Tasks \((i, j)\) is a cut edge in layer \( l = 1 \).

fault tree analysis, without loss of generality this study assumes a simple case. Thus, the following equation represents a simple defect rate calculation for subassembly \( s \), \( \Delta(s) \equiv \sum_i \delta_i + \sum_{(i,j)} \delta_{ij} \)

where \( \delta_i \) and \( \delta_{ij} \) represent the defect rates of component \( i \) and assembly task \((i, j)\), respectively. This subassembly defect rate is an approximated linear form ignoring higher order terms for small defect rates such as in higher than 3-sigma quality levels.

The defect rate of the final assembly depends on subassembly structure and subassembly inspection as well as the components and assembly tasks. Assume that an inspection is conducted between the times when a subassembly is constructed and it is joined for the final assembly. Since the inspection is not perfect, there is a possibility of accepting a nonconforming subassembly. This probability, known as a type II error probability and denoted by \( \beta \), is incorporated in determining the defect rate of the final assembly. For the subassembly structure in Figure 1, \( \Delta(s) = \alpha + \delta_{\text{com}} + \beta\alpha + \delta_{\text{ins}} + \delta_{\text{sub}} \). In general, the defect rate of the final assembly with the inspection of non-single component subassemblies is calculated as:

\[
\Delta(s) = \sum_s \beta (1 - \gamma_i) \cdot \Delta(s) + \gamma_i \cdot \Delta(s),
\]

where \( \gamma_i \) is a binary variable indicating a single-component subassembly.
\[ \sum_{j \in V} q_j^s - \sum_{j \in V} q_j^s \geq u_j^s - \text{Card}(V) \cdot r_j^s \quad \forall i \in V, \forall s \in A \]

\[ \sum_{i \in V} r_i^s = 1 \quad \forall s \in A \quad (5) \]

\[ \sum_{j \in V} q_j^s \leq (\text{Card}(V) - 1) \cdot u_j^s \quad \forall i \in V, \forall s \in A \]

\[ u_i^s, x_i^s, r_i^s, p_i^s, y_i^s \in \{0,1\} \quad \forall i, j, \forall s \in A \]

\[ q_j^s, \Delta(s) \geq 0 \quad \forall i, j, \forall s \in A \quad (6) \]

Equation set (1) represents the objective function to minimize the final assembly’s defect rate that can be calculated by applying the procedure in Section 3 recursively or other quality evaluation methods. Constraint set (2) restricts that each component is assigned to exactly one subassembly. Constraint set (3) ensures that a subassembly containing components i and j must also include the assembly task joining i and j. Constraint set (4) controls the number of subassemblies. The first and second equations determine if a possible subassembly includes any components. The third and fourth equations restrict the desirable maximum and minimum number of subassemblies generated. Constraint set (5) ensures each subassembly is a connected graph. This constraint is an adaptation from a contiguity equation in [7]. Constraint set (6) defines the ranges and types of variables.

5. Numerical experiments

5.1 Case study description

This section describes numerical case studies to demonstrate the effectiveness of the proposed models in the assembly decomposition problem. The list of the 14 components and their defect rates are shown in Table 1. The 18 assembly tasks and their precedence diagram are shown in Figures 2 and 3. The assembly task defect rate is shown in Table 2. These defect rates were generated based on a 6-sigma quality level. Type II error probability (\( \beta \)) for this study was set to 10%.

The final assembly defect rate is calculated by the procedure shown in Section 3, and the resulting problem was solved by a generic MIP solver CPLEX®.

5.2 Result and discussion

The optimal solution shows improved defect rates in the final assembly. For illustration, the paper compares the optimal solution, one of the feasible solutions, and an optimal solution with an increased task F17 defect rate in Figures 4 and Table 3.

![Figure 2](image)

**Figure 2.** Liaison graph and assembly tasks. Circles represent components and straight lines between circles represent assembly tasks.

![Figure 3](image)

**Figure 3.** Task precedence diagram. F19 is a dummy task representing the end of the product assembly process.

For the case of Figure 4-(a), some components with high defect rates (e.g. components H3 and F in subassembly s1) are included in subassembly s1 to be inspected, and this earlier screening of non-conforming parts results in reduced possible problems later in the final assemblies. For the case of Figure 4-(b), these two components are not involved with any earlier screening processes. In addition, some assembly tasks with considerably high defect rates (e.g. tasks (W4, C) and (W4, H3) across subassembly boundaries) are performed in the final assembly process, resulting in the possible defects of the final product assembly. For the case of the optimal solution with the increased task F17 defect rate to 0.01% in Figure 4-(c), due to this increased task defect rate, this task is included in subassembly s1 to reduce the defect rate of the final assembly. All these led to that the optimal solutions that have defect rates less than 60% of the other case.

![Table 1](image)

**Table 1.** Component defect rates.

<table>
<thead>
<tr>
<th>Name (components)</th>
<th>Defect rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.005%</td>
</tr>
<tr>
<td>H1</td>
<td>0.009%</td>
</tr>
<tr>
<td>H2</td>
<td>0.001%</td>
</tr>
<tr>
<td>H3</td>
<td>0.008%</td>
</tr>
<tr>
<td>E1</td>
<td>0.005%</td>
</tr>
<tr>
<td>E2</td>
<td>0.01%</td>
</tr>
<tr>
<td>E3</td>
<td>0.001%</td>
</tr>
</tbody>
</table>

![Table 2](image)

**Table 2.** Assembly task defect rates.

<table>
<thead>
<tr>
<th>Name (associated components)</th>
<th>Defect rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 (E1,E3)</td>
<td>0.003%</td>
</tr>
<tr>
<td>F2 (E3,W4)</td>
<td>0.005%</td>
</tr>
<tr>
<td>F3 (E1,E2)</td>
<td>0.01%</td>
</tr>
<tr>
<td>F4 (W1,W4)</td>
<td>0.009%</td>
</tr>
<tr>
<td>F5 (E2,W4)</td>
<td>0.006%</td>
</tr>
<tr>
<td>F6 (W1,H1)</td>
<td>0.003%</td>
</tr>
<tr>
<td>F7 (E2,H2)</td>
<td>0.007%</td>
</tr>
<tr>
<td>F8 (W2,H1)</td>
<td>0.008%</td>
</tr>
<tr>
<td>F9 (H1,H2)</td>
<td>0.003%</td>
</tr>
</tbody>
</table>

![Table 3](image)

**Table 3.** The comparison of the assembly defect rates for Figure 4 cases.

<table>
<thead>
<tr>
<th>Case</th>
<th>s0</th>
<th>s1</th>
<th>C</th>
<th>F</th>
<th>H3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>272</td>
<td>1720</td>
<td>50</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(b)</td>
<td>479</td>
<td>1490</td>
<td>50</td>
<td>30</td>
<td>80</td>
</tr>
<tr>
<td>(c)</td>
<td>286</td>
<td>1760</td>
<td>-</td>
<td>30</td>
<td>-</td>
</tr>
</tbody>
</table>

The results of the numerical examples suggest that the defect rates of the final assembly are highly affected by subassembly structures. The reason is that in the optimal structure the impact of component defects on the final
assembly defect rate is reduced by giving a chance to inspect less reliable subassemblies including high defect rate components and tasks early.

(a): optimal subassembly structure

(b): a feasible subassembly structure

(c): optimal subassembly structure with the increased task F17 defect rate (0.01%)

Figure 4. Comparison of the different subassembly structures.

6. Conclusions
This study presented a new approach for evaluating the effect of subassembly structures on the quality of the final product assembly. A graph-theoretic approach for partitioning the liaison graphs of a product assembly was presented. The evaluation of the final assembly defect rate considered the defect rates of components and assembly tasks in subassemblies as well as type II errors in subassembly inspection. The results of this study suggest that the defect rate of the final product can be managed better by proper subassembly planning.

Acknowledgement
This study was partially supported by the US National Science Foundation Grant #1100960.

References