MANUFACTURING SYSTEM

Edited by **Faieza Abdul Aziz**

INTECHOPEN.COM

Manufacturing System
Edited by Faieza Abdul Aziz

Published by InTech
Janeza Trdine 9, 51000 Rijeka, Croatia

Publishing Process Manager Marija Radja
Technical Editor Teodora Smiljanic
Cover Designer InTech Design Team

First published April, 2012
Printed in Croatia

A free online edition of this book is available at www.intechopen.com
Additional hard copies can be obtained from orders@intechopen.com

Manufacturing System, Edited by Faieza Abdul Aziz
p. cm.
ISBN 978-953-51-0530-5

Contents

Preface

This book attempts to bring together selected recent advances, tools, application and new ideas in manufacturing systems. Manufacturing system comprise of equipment, products, people, information, control and support functions for the competitive development to satisfy market needs.

The book contains 21 chapters which covers the very latest developments and the most up-to-date techniques and systems in the following areas:

- Manufacturing Systems Design, Development and Operations
- Intelligent Manufacturing Systems
- Scheduling Systems
- Production Systems and Process Optimization
- Supply Chain Management and Logistics
- Digital Manufacturing Systems
- Control System Design
- Knowledge Management Systems
- Manufacturing Strategy

Manufacturing System provides a comprehensive collection of papers on the latest fundamental and applied industrial research. The book will be of great interest to those involved in manufacturing engineering, systems and management and those involved in manufacturing research.

Editor in Chief
Engr. Dr. Faieza Abdul Aziz
Universiti Putra Malaysia

Co-Editors
Dr. Azmah Hanim Mohamed Ariff, Universiti Putra Malaysia
Dr. Azfanizam Ahmad, Universiti Putra Malaysia
Ir.N.Jayaseelan, The Institution of Engineers, Malaysia
Dr. Qaisar Ahsan, Center of Excellence for Science and Applied Technology, Islamabad, Pakistan

Dr. Eng. Zakria Mahrousa, University of Aleppo, Syria
Mohd Shukri Ibrahim, Universiti Putra Malaysia
Awanis Romli, Universiti Malaysia Pahang

1

Particle Reduction at Metal Deposition Process in Wafer Fabrication

Faieza Abdul Aziz, Izham Hazizi Ahmad,
Norzima Zulkifli and Rosnah Mohd. Yusuff
Universiti Putra Malaysia
Malaysia

1. Introduction

Metal Deposition or metallization process is one of the processes in fabricating a wafer. A wafer is a thin slice of semiconductor material, such as a silicon crystal, used in the fabrication of integrated circuits and other micro-devices. Due to the nature on the process, it creates lot of particles, which would impact the next process if it were not removed. Particle deposition on the wafer surface can cause the circuit to malfunction; leading to a loss of yield. Cleaning process needs to be done after metal deposition process in order to remove the particles

Metal deposition, which has been constructed by several metal layers, allows the flow of current between interconnections. Each metal layers consist of three types of metal films such as Ion Metal Plasma Titanium (IMP Ti), Titanium Nitride (TiN) and Aluminum. The metal deposition started after the wafer has completed the "Tungsten Chemical Mechanical Polishing" (CMP) process.

Metal layers are deposited on the wafer to form conductive pathways. The most common metals include aluminium, nickel, chromium, gold, germanium, copper, silver, titanium, tungsten, platinum and tantalum. Selected metal alloy also may be used. The metal layer is shown in Figure 1 and the interconnection between metal layers is shown in Figure 2.

The deposited metal(s) offers special functionality to the substrate. Typically, the metal aqueous solution is employed for the wet metal deposition process due to the consideration of its low cost and operation safety.

Metallization is often accomplished with a vacuum deposition technique. The most common deposition processes include filament evaporation, electron- beam evaporation, flash evaporation, induction evaporation and sputtering. There are also two types of wet metal deposition processes – electrolytic and electro-less plating.

Sputtering and evaporation are well established as the two most important methods for the deposition of thin films. Although the earliest experiments with both of these deposition techniques can be traced to the same decade of the nineteenth century (Grove, 1852; Faraday, 1857), up until the late 1960s evaporation was clearly the preferred film-deposition technique, owing to its higher deposition rates and general applicability to all types of materials. Subsequently, the popularity of sputter deposition grew rapidly because of the need to fabricate thin films with good uniformity and good adhesion to the substrate surface (demand driven by the microelectronics industry) as well as the introduction of radio-frequency (RF) and magnetron sputtering variants.

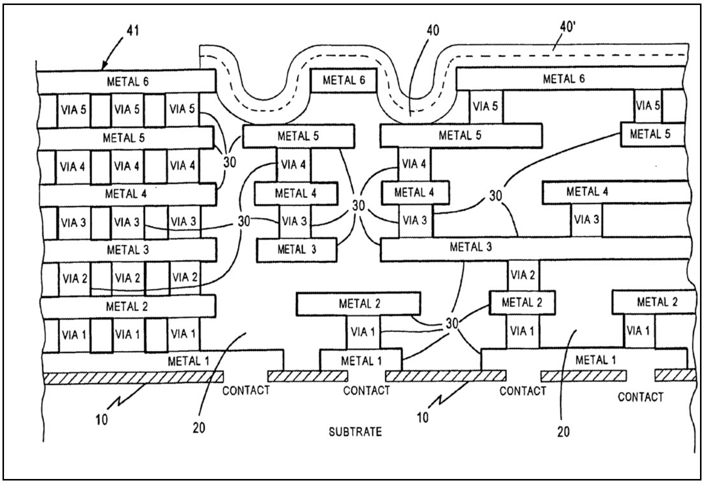


Fig. 1. The Metal layers

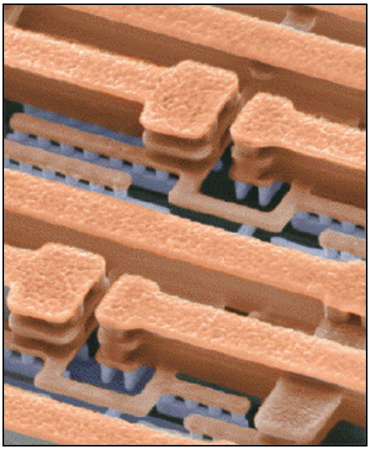


Fig. 2. The interconnection between metal layers

In this chapter, a thorough investigation was carried out to improve shut down event problem at Metal Deposition process during wafer fabrication. Particle contamination on wafer surface can cause the circuit to malfunction and leading to machine shut down. Data of shutdown event versus sputter target life showed that the rate of machine shutdown increased by the increment of sputter target life. The sputter target life was further investigated to determine the appropriate sputter target life to be used in order to avoid particles generation during metal deposition process.

2. Particles

Particles can be defined as "suspension of solid or liquid mass in air". Particles can originate from a variety of sources and possess a range of morphological, chemical, physical and thermodynamic properties. The particles could be combustion generated, photo-chemically produced, salt particles from sea spray or even soil-like particles from re-suspended dust. Particles may be liquid; solid or could even be a solid core surrounded by liquid.

Particles are represented by a broad class of chemically and physically diverse substances. Particles can be described by size, formation mechanism, origin, chemical composition, atmospheric behavior and method of measurement. The concentration of particles in the air varies across space and time, and is related to the source of the particles and the transformations that occur in the atmosphere. Some of the more generalized characterization of particles is:

i. Primary and secondary particles: A primary particle is a particle introduced into the air in solid or liquid form, while a secondary particle is formed in the air by gas-to-particle conversion of oxidation products of emitted precursors.
ii. Particle characterization as per size: Particle can be classified into discrete size categories spanning several orders of magnitude, with inhalable particles falling into the following general size fractions- PM_{10} (equal to and less than 10 micrometre (µm) in aerodynamic diameter), $PM_{2.5-10}$ (greater than 2.5 µm but equal to or less than 10 µm), $PM_{2.5}$ (2.5 µm or less), and ultra fine (less than 0.1 µm).
iii. Particle characterization depending on requirements of study: Some of the particle components/ parameters of interest to health, ecological, or radiative effects; for source apportionment studies; or for air quality modeling evaluation studies are particle number, particle surface area, particle size distribution, particle mass, particle refractory index (real and imaginary), particle density and particle size change with density, ionic composition (sulphate, nitrate, ammonium), chemical composition, proportion of organic and elemental carbon, presence of transition metals crustal elements and bioaerosols

2.1 Particle contamination

Particle contamination can be defined as the act or process of contaminating by particulates. Particle contamination is problematic for many industries. They can appear unexpectedly mixed in solids, liquids and gases. Particles can be from many sources i.e.- metals, biological (skin, hair etc), polymers, building dusts etc. They all have different characteristics and properties such as shape, size and chemistry, which assist in identification. Scanning Electron Microscope (SEM) and Energy-dispersive X-ray spectroscopy (EDX) coupled with optical microscopy provides a powerful machine for unambiguously identifying such particles. The technique is frequently coupled with Fourier transform infrared spectroscopy (FTIR) when identifying the source of organic contamination (Stephen, 2010).

2.1.1 Particle contamination in semiconductor

Deposition of aerosol particles on semiconductor wafers is a serious problem in the manufacturing of integrated circuits. Particle deposition on the wafer surface can cause the circuit to malfunction, leading to a loss of yield. With the circuit feature approaching 1 µm in size of one-megabit memory chips, particle control is becoming increasingly more important (Benjamin et al., 1987). Particle contamination during vacuum processing also has a significant impact in Very Large Scale Integration (VLSI) process yield (Martin, 1989) and has motivated most manufacturers to adopt particle control methods base on sampling inspection.

According to Bates (2000), semiconductor memory chips are very sensitive to the particles because the circuitry is so small. In a typical clean room manufacturing environment, particles are deposited on the wafer surface by sedimentation, diffusion, and/ or electrostatic attraction. Sedimentation usually occurs for large particles, particularly those larger than 1 μm in diameter, whereas diffusion occurs for small particles below 0.1 μm in diameter.

In the intermediate size range, both sedimentation and diffusion may occur and must be considered. When particles are electrically charged, enhanced deposition can take place. The rate of particle deposition on a wafer surface depends on both the size of the particle and their electrical charge. In addition, the deposition rate is also influenced by the airflow around the wafer, which in turn are affected by the size of the wafer, the airflow velocity, and the orientation of the wafer, with respect to the airflow. Although the mechanisms of particle deposition on semiconductor wafers are reasonably well understood and approximate calculations have been made (Cooper, 1986; Hamberg, 1985), no detailed quantitative calculation has been presented.

2.1.2 Particle contamination in wafer processing

As the chip density increases and semiconductor devices shrink, the quality of fabrication becomes more crucial. The composition, structure, and stability of deposited films must be carefully controlled and the reduction of particulate contamination in particular becomes increasingly crucial as device sizes shrink and densities increase. As the devices grow smaller, they become more sensitive to particulate contamination, and a contaminant particle size that was once considered acceptable may now be a fatal defect. Voids, dislocations, short circuits, or open circuits may be caused by the presence of particles during deposition or etching of thin films. Yield and performance reliability of microelectronic devices may be affected by the mentioned defects (Alfred, 2001).

Often, the process gases will react and deposit material on other surfaces in the reactor besides the substrate. The walls of the processing chambers may be coated with various materials deposited during processing, and mechanical and thermal stresses may cause these materials to flake and become dislodged, generating contaminated particles. In processing steps that use plasma, many ions, electrons, radicals, and other chemical "fragments" are generated. These may combine to form particles that eventually deposit on the substrate or on the walls of the reactor (Alfred, 2001). Particulate contamination also may be introduced by other sources, such as during wafer transfer operations and backstream contamination from the pumping system used to evacuate the processing chamber.

In plasma processing, contaminated particles typically become trapped in the chamber, between plasma sheath adjacent to the wafer and plasma glow region. These particles pose a significant risk of contamination, particularly at the end of plasma processing, when the power that sustains the plasma is switched off. In many plasma-processing apparatuses, a focus ring is disposed above and at the circumference of the wafer to enhance uniformity of processing by controlling the flow of active plasma species to the wafer, such as during a plasma etch process. The focus ring, and the associated wafer clamping mechanism, tends to inhibit removal of the trapped particles by gas. Thus, there is a need to provide a reliable and inexpensive process to remove such particles from the wafer-processing chamber (Alfred, 2001).

Similarly, in chemical vapor deposition and etching, material tends to deposit on various parts of the apparatus, such as the susceptor, the showerhead, and the walls of the reactor, as the by-products of the process condenses and accumulates. Mechanical stresses may cause the deposited material to flake and become dislodged. These mechanical stresses are often caused by wafer transfer operations, but may also be caused by abrupt pressure changes induced by switching gas flow on and off and by turbulence in gas flow. Thus, process by-products at the end of the processing stage must be flushed from the chamber to prevent them from condensing and accumulating inside the chamber.

Typically, the flow of the processing gas is shut off at the end of a processing stage, whereupon the pressure in the chamber rapidly falls to zero as the vacuum pump continues to run. Idle purge may be used; in which purge gas is introduced into the chamber at intervals while no processing is taking place. Nonetheless, pressure spikes occur with the cycling of gas flow, causing disruption of particles, which may then contaminate the wafer surface. This limits the particle reduction benefits from the idle purge. A large portion of device defects is caused by particles disrupted by pressure change during wafer loading and moisture on the pre-processed wafer surface (Alfred, 2001).

Three types of particle contamination can be defined, which are under the deposited film as shown in Figure 3, in the deposited surface as shown in Figure 4 and deposited Film as shown in Figure 5. Particle under the deposited film will cause the surface of the wafer to become dirty. The particle may come from the previous process. Particle in the deposited surface will cause gas phase nucleation, leaks into the system, contamination in gas source/flow lines and sputter off walls. The particle may come from the gas phase nucleation, system leak or contaminated gas line. Particle on the deposited film will cause film build-up on the chamber walls. The source may come form the process chamber or from the wafer handling.

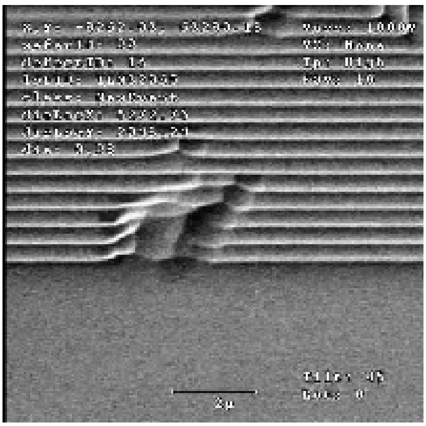


Fig. 3. Particle under the deposited Film

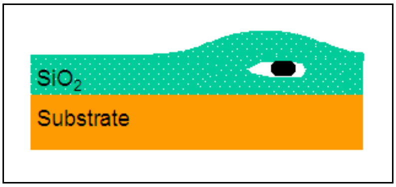


Fig. 4. Particle in the deposited surface

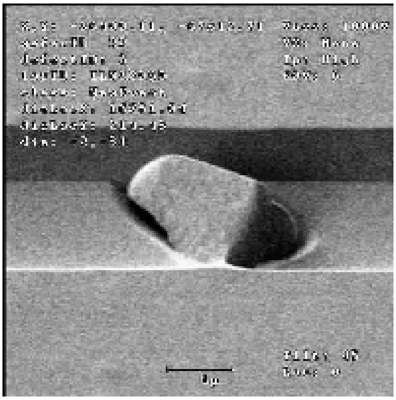


Fig. 5. Particle on the deposited Film

Example of TiN particle transformation is shown in Figure 6. From this figure, the particle was dropped on the wafer's surface. The source of the particle may come either form previous process or from current process. After the deposition process done, the particle will be covered underneath the metal layer, which cause the damage of the interconnection.

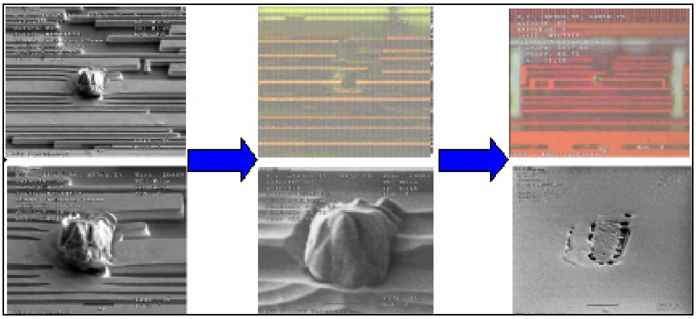


Fig. 6. The transformation of TiN particle

The particles entrained in the load lock air volume by turbulence during pumping are either carried in through the handling of the wafer, generated within the camber from causes such as wear or residual from previous pumping and venting cycles. Particles are removed as they are drawn out during pumping or as they are carried out of the surface of the wafers. Additional particles may bind to the walls of the chamber or machining to tightly that they are agitated free by subsequent pump/ vent cycles (Peter Bordon, 1990).

An equilibrium background level is reached because the number of particles carried out by pumping and deposition on the wafer surface is proportional to the number of particles entrained into the gas volume. For example, if the number of particles entrained doubles, twice as many particles land on the wafer and twice as many flow out the pump line (Bordon, 1990). The effectiveness of these mechanisms has long been recognized. For example, it is a common practice to pump/ vent clean process chambers in high- current ion implanters and other process machines or to run getter wafers after a chamber has been contaminated.

Low levels of particulate contamination can be obtained in process gas systems by using careful system design, high-quality compatible materials, minimum dead legs and leak rates, careful start-up and operating procedures, etc. Low particle levels can also be obtained in gas cylinders through careful selection of cylinder materials, surface treatment and preparation, and through close attention to gas fill system design and operation (Hart, et al., 1994).

Particle levels in flowing gas systems may be steady or (as in machine vent lines) cyclic over time. In machine feed lines, the gas is usually well mixed and particles are uniformly distributed. However, particle levels in gas cylinders can vary by orders of magnitude over time due to such effects as liquid boiling, gravitational settling, and diffusion to internal surfaces. Such effects may also produce non-uniform particle distributions, including stratification, in gas cylinders (Hart, et al., 1995). Levels of suspended particles in filled cylinders can be measured with a high-pressure Optical Particle Counters (OPC). Data obtained directly from cylinders show that careful attention to quality can result in low cylinder particle concentrations.

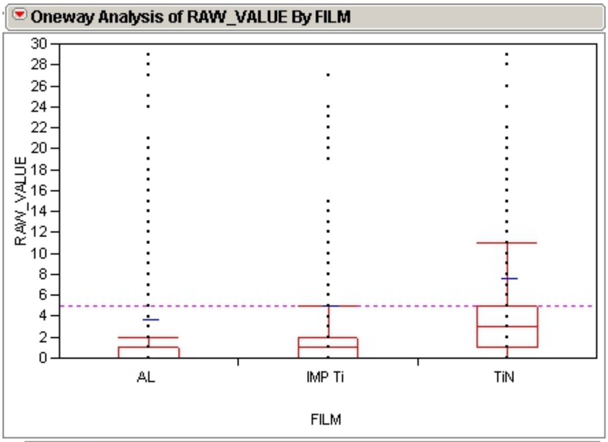
Cylinder and bulk gases are frequently reduced in pressure with an automatic regulator before entering the flowing distribution system. Automatic regulators may produce increased particle levels (through regulator shedding, impurity nucleation, and condensational droplet formation) that are sometimes followed by system corrosion (Chowdhury, 1997) or suspended nonvolatile residue formation. Gases are therefore filtered after pressure reduction and before entering the distribution system. Ceramic, metal, or polymer membrane filters are selected for compatibility with the process gas. Such filters can produce a low particle level as well as a low degree of variability in contamination over time.

CNC data for particles as small as 0.003µm in O2 and H2 can also be obtained using an inert gas CNC with a special sample dilution device developed by Air Products (McDermott, 1997). These data showed that membrane filters can be used to produce high-cleanliness gases to 0.003µm in large-volume gas systems. Well-designed distribution systems should contribute a minimum of additional particulate contamination to the flowing gas.

As the particle may impact the wafer quality, which result in wafer scrap, corrective and preventive action must be made immediately to stop the particle contamination from becoming catastrophic. Thus, a systematic problem solving method is needed to solve the issue.

2.2 Particle failure

In this step, the types of particle failure were studied. Data from Daily Particle Qualification process was analyzed. TiN particle have highest standard deviation ~4.0 compared to Aluminium (Al) and Ion Metal Plasma Titanium (IMP Ti) as shown in Figure 7. This showed that TiN particle performed the most inconsistent compared to Al and Imp Ti particle in the Metal Deposition process. Data for each films particle qualification was obtained and then Pareto Analysis was made. From the Pareto Chart, TiN defect has the highest failure rate compare to other films as shown in Figure 8.



Quantiles

Level	Minimum	10%	25%	Median	75%	90%	Maximum
AL	0	0	0	0	1	3	29
IMP Ti	0	0	0	1	2	5	27
TiN	0	0	1	3	5	8	30

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
AL	2033	1.19085	2.49566	0.05535	1.0823	1.2994
IMP Ti	1192	1.89262	3.07781	0.08915	1.7177	2.0675
TiN	2070	3.64589	3.92741	0.08632	3.4766	3.8152

Fig. 7. Comparison between Al, IMP Ti and TiN

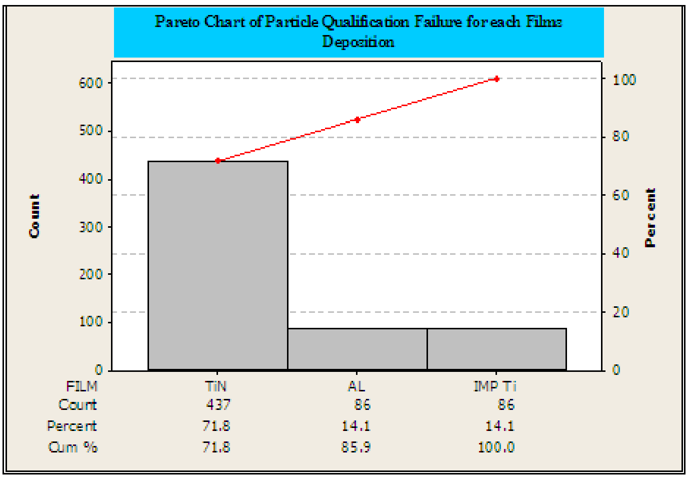


Fig. 8. Pareto Chart of Particle Qualification Failure for each Films Deposition

2.3 Chamber configuration and wafer processing sequence

Chamber configuration for metal deposition machine is shown in Figure 9. From below chamber configuration, the wafers which inside the cassette are placed at cassette Load lock, which consist of Load lock A (LLA) and Load lock B (LLB). Wafer will pass through form Buffer camber to transfer chamber in Chamber A. Then, wafer will be cooled down in Chamber B. At the end of the process, wafers will be vented to Atmosphere condition in Load lock A or Load lock B, depending to which Load lock the wafers origin.

Wafers in a production pod will be pumped down to vacuum condition from atmospheric pressure in load lock A or load lock B, depending on where the lot is placed. Degas and notch alignment occurred in Chamber E and F. The deposition process begins with IMP Ti deposition in Chamber C. The wafers will move into the transfer chamber in Chamber A. Metal deposition will occur in Chambers 1, 2, 3 and 4. Depending on the application, multiple metal films can be stacked without breaking the vacuum. After the deposition process, wafer will be cooled down in Chamber B and vented back into the atmosphere in the production pod at Load lock A or B.

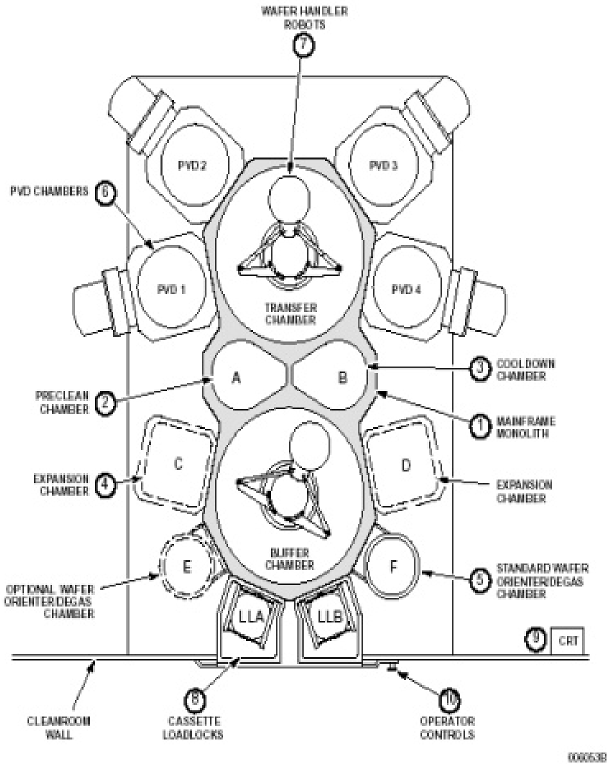


Fig. 9. Chamber Configuration of the Metal Deposition Machine

3. Cause and effect analysis diagram

A case study has been conducted in one wafer fabrication company. Downtime reduction at Metal Deposition process is being focused in this work. In average, three cases of Metal Deposition machines have been shutdown every week. Shutdown criteria is based on more than "10-area count per wafer" or well known as "adders". The machine will be shut down if the post scan result shows particle increase more than 10 adders.

Brainstorming session has been done with the team members and root causes have been identified and classified under six main sectors, which are machines, material, methods, measurements, environment and personnel. Figure 10 is the fish bone diagram for cause and effect analysis.

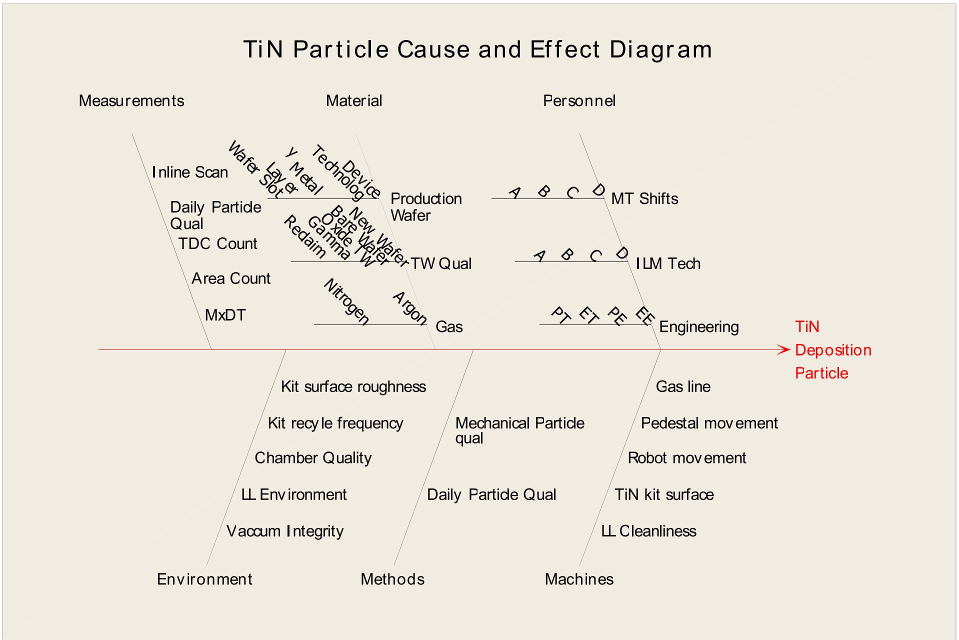


Fig. 10. The fish bone diagram of cause and effect analysis

3.1 Measurement systems

Particle measurement is performed by using the SP1 machine as shown in Figure 11. SP1 is the machine, which is measuring the particle by scanning and counting the existing particle on the wafer. Besides that, SP1 also is capable to show the wafer map, which can tell either the particles are clustered on the wafer, saturated, mild signature and others.

Percentage of Gage Repeatability and Reproducibility (GRnR) was done on SP1 in the production floor, which is used for particle scanning purpose. Gage RnR study is conducted to determine the measurement system variability in term of Repeatability and Reproducibility. TiN particle count is our KPOV that causing machine shutdown by ILM due to metal deposition particle existence in the machine.

Based on the GRnR study, the result proves that the SP1 is capable to measure the TiN Dep particle counts cine the total GRnR is less than 30%. The result of GRnR study is shown in Figure 12.

Fig. 11. The SP1 machine which measure the particle

Gage R&R

Source	VarComp	%Contribution (of VarComp)
Total Gage R&R	0.67500	6.88
Repeatability	0.38333	3.91
Reproducibility	0.29167	2.97
Day	0.14722	1.50
Day*Slot	0.14444	1.47
Part-To-Part	9.13519	93.12
Total Variation	9.81019	100.00

Source	StdDev (SD)	Study Var (6 * SD)	%Study Var (%SV)
Total Gage R&R	0.82158	4.9295	26.23
Repeatability	0.61914	3.7148	19.77
Reproducibility	0.54006	3.2404	17.24
Day	0.38370	2.3022	12.25
Day*Slot	0.38006	2.2804	12.13
Part-To-Part	3.02245	18.1347	96.50
Total Variation	3.13212	18.7927	100.00

Number of Distinct Categories = 5

Fig. 12. The result of GRnR study for SP1

3.2 In-line monitoring and systematic machine excursion monitoring

There are three methods of inspecting and measuring the particle, which are using production wafers through Systematic Machine Excursion Monitoring (STEM), production wafer that went to In-line monitoring process (ILM) flow and test wafers which is being used during machine qualification process.

In- line monitoring (ILM) is a process to detect any defect in real time. It is done in many ways, such as in line inspection, upon request from user, from production lots which go to ILM flow and also Systematic Machine Excursion Monitoring (STEM) lots. Machine-related defect excursions are controlled by systematically checking process machines. Production wafers are being used for STEM purpose. Each of Metal Deposition machine need to do STEM activity once every two days. STEM is a process where the lot which is already completely processed from one machine, will be held for ILM scan. The scan is done to check for any defects that may be caused by the processing machine at previous process. STEM will provide faster detection and containment of the defect excursion. All major process machines are monitored in a systematic manner.

For STEM activity, Manufacturing Technician will hold the lot for ILM Technician after run through the metal deposition machine. ILM Technician will scan four wafers/ machine using Complus or AIT machine. If the scan result shows particle signature and above the control limit (more than 10 counts), they will shutdown the whole machine and the machine owner need to verify the shutdown prior to release the machine back to production.

For production wafers, there will be about 30% of the WIP will go to ILM inspection step. This is the random sampling in line scanning that has been designed to detect any defects along the process of fabricating the wafers from first process until end on the process. It has been designed in the process flow, where lots that are needed for this sampling will have ILM inspection flow compare to the other 70% of the lots that do not have ILM flow. Lots that have ILM flow will arrive at ILM inspection step after completing metal deposition process. ILM technician will scan the lot and if found particle and above the control limit (more than 10 counts), machine will be shutdown and same verification need to be done prior to release the machine back to manufacturing.

For qualification process, bare wafers or known as test wafers is used to check the machine's condition and performance. Qualification process is done based on schedule. Basically, every metal deposition machines need to perform qualification process once everyday. This is to ensure the machine is fit to run and not causing any defect later. Qualification process is carried out by manufacturing technician using SP1.

The qualification process is started with the pre particle measurement. Qualification wafer (bare wafer) will be selected and pre particle measurement is done using SP1. After pre-measurement is completed, the wafer will go inside the machine and process chamber for machine and chamber qualification purpose. After the process is completed, the wafer is again brought to SP1 for post particle measurement. The differences between pre particle value and post particle value will determine the machine and chamber's condition. For qualification process, the control limit is tightened to five count only. If particle is found more than five count, chamber will be shut down and pending verification from machine owner is needed prior to release the machine back to production.

Example of the pre particle and post particle measurement is shown in Figure 13. In Figure 13, two wafer maps were shown, which are pre particle wafer map and post particle wafer map. In Pre Particle wafer map, two particle counts were detected as circled. In post particle wafer map, four particle counts were detected. Two count were the existing particles and the other two were new particles, which were detected during post particle measurement. From Figure 13, the adders were two counts (post particle value- pre particle value).

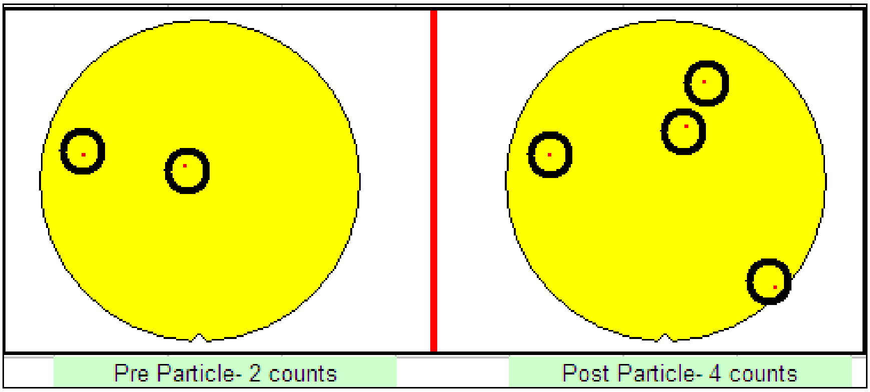


Fig. 13. Pre Particle and Post Particle Wafer Map

3.3 Possible root cause

From Ishikawa Diagram, possible root causes will be screened out to get the actual causes. From actual causes, potential corrective and preventive actions will be determined and implemented.

3.4 Personnel

There are four shifts in the studied company, which are shift A, B, C and D. Manufacturing technician (MT) for every shift is responsible to perform the daily qualification job daily. Since there are four shifts running in the production, the level of experiences between shifts to shift differs. The level of experiences of the MT is very important since they need to perform the qualification process. Experienced MT will know and easily catch the particle issue inside the chamber by looking at the qualification result, but less experienced MT may take some times. Study has been done to check the Manufacturing Technician's efficiency. Their year of services and also certification were referred. The data were obtained from Human Resources Certification Record. Level of certification is from one to three. Level one is the minimum certification level, while level three is the maximum level of certification.

Study has been done to check the ILM Technician's efficiency. Their year of services and also certification were referred. The data were obtained from Human Resources Certification Record.

Process technician, process engineer, equipment technician and equipment engineer also play their roles during machine shutdown. When machine detected unwanted particle and need to be shutdown, process technician normally will follow the Out of Control Action Plan (OCAP) in order to release the machine back to manufacturing group as soon as possible. Process engineer also will take a look on the issue and do analysis and then come out with the release plan. Equipment technicians and engineers need to ensure proper maintenances job been carried out as per checklist. This is to ensure the cleanliness of the machine after Preventive Maintenance (PM) was done. Study has been done to check the Equipment and Process Technician's efficiency. Their year of services and also certification were referred. The data were also obtained from Human Resources Certification Record.

Beside MT, In-Line monitoring (ILM) technician also play big responsibility to determine the particle rate. It is important to have a proper scanning and analyzing of the STEM lot, so that the decision to shutdown the machine is base on real issue.

3.5 Material

For production wafers, different technologies will give different impact of the particle. This is mainly related to the process recipes, which different devices will have different process recipe, thus the deposition rate and thickness will be different from one device to others. Study has been done to see the relationship between particle issue and technologies. From the shutdown event, list of lots that have been scanned was obtained. From the list, product technologies were segregated and the relationship between them with the particle is studied.

Beside the technology, the metal layers also have impact to the particle issue since more metal layers means more times the lot will go to metal deposition process and the chances for expose to particle issue is more. Example for lot with four layer metal will go four times metal deposition process compare lot with five metal layers, will go five times metal deposition process. Study has been done to see the relationship between particle issue and the metal layers. From the same list from shutdown event, metal layers were obtained to see if there is any relationship between metal layers and particle.

Test wafer also have some impact to the particle issue. For new test wafer, the performance is better compared to wafers that sent to rework and reused. This is because the rework wafer normally will have remaining particle, which can not be removed due to saturated at the surface of the wafer and needs stronger cleaning recipe to remove them. Brand new wafers normally will have a lot less particle. In this study, the incoming particle for 50 lots of new test wafers was measured using SP1 to get the potential incoming particle. From here, any existing particle from test wafers itself that may contaminate the process chamber later during qualification process can be seen.

3.6 Method and measurement

Correct methods, which are used during both particle and mechanical qualification, were studied and observed. Judgment was made base on observation across all four shifts on the procedures during the qualification process.

The particle measurement is done based on In-line scan and during particle qualification process. For inline particle scanning, it is done after the lot has completed the metal deposition process. The job of in-line scanning is known as Systematic Machine Excursion Monitoring (STEM), which been done once in two days. Lot will be on hold for in line monitoring (ILM) scan.

Four wafers will be scanned for each machine to check for particle performance. The wafers will be scanned using Scanning Electron Microscope (SEM) machine. If particle signature exists, ILM personnel will notify the Metal Deposition machine owner to check for the machine's health. If the particle level exceeds the limit, which is more than 10 particle counts, the machine will be shut down and need to follow procedures in order to bring the machine back up to the production.

For production lots that go to ILM flow, the lot will be scanned for scratches and particle. If scratches or particle are found to exceed the limit, which is more than 10 particle counts, machine will be shutdown and need to follow the procedure as well. Wafers will be scanned using Complus or AIT machine also.

Qualification process is a process to check for the machine's performance, so that it always performs same as the baseline. One of the important factors in qualification is the particle performance. Particle value is measured based on the different value between pre particle measurement and post particle measurement. Differences of both values will determine the particle existence inside the machine. If particle count is more than five area count/ wafers, machine will be shutdown and need to be followed up by machine's owner before release back to production.

3.7 Machine

Machine is the main focus of the particle issue. This is due to the mechanical movement such as pedestal and robot movement inside the machine that can generate particle. Beside that, gas line also can create particle.

Load lock cleanliness is also very important since this is the place where the lot is transferred into the machine from its base. Load lock is a chamber that is used to interface a wafer between air pressure and the vacuum process chamber. According to Borden (Borden, 1988), Wu (Wu, et al., 1989) and Chen (Chen, et al., 1989), in the absence of a water aerosol, the dominant source of wafer contamination is the agitation of particles during the pumping (venting) of the entry (exit) load lock.

In this study, 100 lots were selected to check the particle level in the cassette. Since wafers are inside the vacuum state inside the cassette, particle inside the cassette need to be measured. It was measured using mini- environment tester. The cassettes were opened in Wafer Start room and the particle was measured for all the 100 lots. The particle count that obtained from the testing is captured.

To study for particle during wafer handling and robot movement, mechanical qualification process was carried out. Before it was done, chamber was cleaned first to eliminate the potential source of particle coming from process chamber. One lot, which consists of five test wafers, was selected. Wafers inside the cassette were arranged in slot 5, 10, 15, 20 and 25. Pre particle measurements were obtained for all the five wafers. The lot then was vented inside the load lock and also into the deposition module. Without running the deposition process, the lot was moved out back into the load lock and cassette. Post particle measurement was done to check for the adders. This cycle is repeated for 10 times for the entire machine and data is captured and analyzed.

Particle in gas line also was focused in this study. Particle in gas line is measured by referring to the data that is obtained from the particle sensor. The particle sensor is mounted at the gas line as shown in Figure 14. This is to ensure any particle in the gas line can be detected and the amount of particle entered to the process chamber can be monitored and recorded.

Sputter target also been studied to check the correlation between sputter target life and also shutdown. Example of sputter target that been used inside metal deposition machine is shown on Figure 15. The event of shutdown and the usage of target life is captured and analyzed.

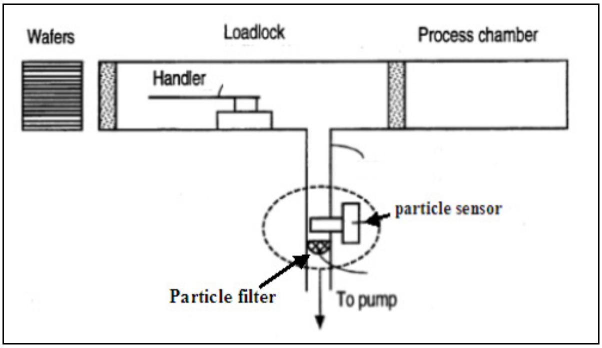


Fig. 14. Particle Sensor that Mounted at Gas Line

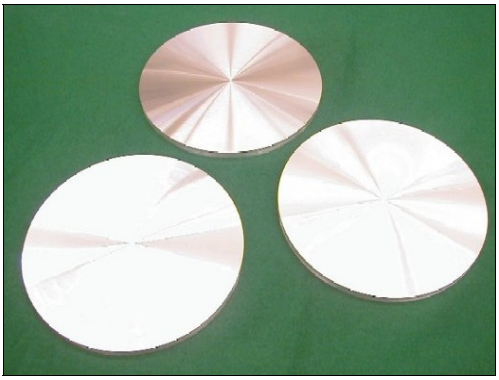


Fig. 15. Example of Sputter Target

3.8 Environment

The machine environment also has been studied to see any contribution to shutdown event. Load lock environment has been checked. Airborne particle measurement was conducted over metal deposition machines to collect the particle count. One hundred lots were prepared in this study. Wafers inside the cassette were arranged in slot 5, 10, 15, 20 and 25. Pre particle measurements were obtained for all the five wafers. The lot then was vented inside the load lock and left for five minutes. After five minutes, lot was moved out back from load lock move into cassette. Post particle measurement was done to check for the adders.

4. Results and discussion

From the fish bone diagram, all the possible causes have been screened out and verified to find the actual true causes. From actual true causes, corrective actions and preventive actions will be defined, identified and will be implemented to eliminate the particle issue.

4.1 Personnel

Verification have been made to people who directly working at metal deposition machine and related to the shutdown. Summary of the possible causes, which related to personnel, is shown in table 1.

Causes		Verification & Validation Process	Result	True/ False
Personnel	Manufacturing Technician (MT) in shift	1. To check the level of experiences of the MT 2. To check the capability of MT to perform qualification job correctly	1. Base on the study, all shift have dedicated MT > 2 years of experience to handle Metal Deposition tool 2. All the Manufacturing Technicians also capable to perform qualification job corectly	FALSE
	In Line Monitoring (ILM) Technician in Shift	To check the capability of ILM Technician to handle to scanning tool and catch the particle	Base on the study, all shift ILM Technician also having > 2 years of experience and capable to handle the scanning tool	FALSE
	Engineering Personnel (Process/ Equipment)	1. To check the capability of shift Equipment Technician (ET) and Equipment Engineer (EE) to perform Preventive Maintenance (PM) job efficiently. 2. To check the capability of Process Technician (PT) and Process Engineer (PE) to follow up on the ILM Shutdown issue to avoid re-shutdown	1. All shift ET/ PT and Engineers (PE/ EE) are well trained and have experiences > 3 years in average. 2. Equipment & Process team have their own checklist to be followed and verified by Section Head.	FALSE

Table 1. Summary of verify possible root causes related to personnel

4.1.1 Manufacturing technician (MT)

A validation and verification have been done to check the level of experiences of the MT and also the capability of MT to perform qualification job correctly. From the study, all shifts have dedicated MT more than two years of experience to handle Metal Deposition machine due to the criticality of metal deposition process. All the Manufacturing Technicians are also capable to perform qualification job correctly based on the checklist. The dedicated MT is summarized in table 2.

Shift	Person in Charge	Date Join	Years of Experiences	Certification level
A	1. NEDUMARAN A/L MEGAWARNAM	2004	> 5 years	3
	2. LIYANA HANIM BINTI AKBAR	2006	> 3 years	3
	3. MUHAMAD TERMIZI BIN AHMAD TAJUDIN	2006	> 3 years	3
B	1. ROZI BIN MD HASSAN	2003	> 6 years	3
	2. NOOR JANNAH BINTI MAHADZIR	2007	> 2 years	3
	3. ANUAR BIN MAT ISA @ ABDUL AZIZ	2007	> 2 years	3
C	1. MOHD ASRIZAL BIN AHMAD	2003	> 6 years	3
	2. KASMINI BINTI TUKOL	2007	> 2 years	3
	3. MAIMUNAH BINTI HASHIM	2007	> 2 years	3
D	1. MOHD YUSRI BIN YUSOF	2007	> 2 years	3
	2. MOHD IZHAM BIN MOHD IZHAR	2004	> 5 years	3
	3. ZAIDA BINTI AHMAD	2007	> 2 years	3

Table 2. Manufacturing Technicians in-charged of Metal Deposition Machine

4.1.2 Inline monitoring (ILM) technician

Based on the study, all shift ILM Technicians are also having more than two years of experience and capable to handle the scanning machine and captured defect images. The shift ILM Technician is summarized in table 3. Conclusion can be made that all the MT who handle the medal deposition machines are capable to perform qualification process and mistake that can lead to particle generation is almost zero.

Shift	Person in Charge	Date Join	Years of Experiences	Certification level
A	REDZUAAN BIN ABDUL RAHIM	2002	>7 years	3
	BALAKRISNAN A/L A.MUNIANDI	2007	> 2 years	3
B	RUZAINI B. ADZHA	2006	> 3 years	3
	CHAREN A/L KHAN	2007	> 2 years	3
C	SUNTHARA MURTHI S/O RAMAN	2003	> 6 years	3
	ERUAN BIN ABU SEMAN	2007	> 2 years	3
D	VASANTHAN A/L VELOO	2007	> 2 years	3
	IBRAHIM BIN IDRIS	2005	> 4 years	3

Table 3. In-line monitoring (ILM) Shift Technicians

4.1.3 Engineering personnel (process/ equipment)

Verification and validation made to check the capability of shift Equipment Technician (ET) and Equipment Engineer (EE) to perform Preventive Maintenance (PM) job efficiently. Also validation made on the capability of Process Technician (PT) and Process Engineer (PE) to follow up on the ILM Shutdown issue to avoid re-shutdown due to incorrect qualification job done prior releasing machine during shutdown.

The summary of PT is shown in table 4 and summary of ET is shown in table 5. Based on the verification, all shift ET/ PT and Engineers (PE/ EE) are well trained and having experiences to perform their job efficiently. Equipment and Process team have their own checklist to be followed and verified by Section Head during performing PM activities and also releasing the machine from shutdown.

Shift	Process Technician	Date Join	Years of Experiences	Certification level
A	NOR AZELINA BINTI ISMAIL	2002	>7 years	3
	HARYANI BINTI ABDULLAH	2007	> 2 years	3
B	NORMALA BINTI NAPIAH	2006	> 3 years	3
	MOHD SYUKRI BIN CHE HASSAN	2007	> 2 years	3
C	KHAIRUL ANWAR BIN ABU BAKAR	2003	> 6 years	3
	CANITTHA A/P IEKIN	2007	> 2 years	3
D	PUTERI SURINAEDAYU BINTI MEGAT ISMAIL	2007	> 2 years	3
	NOR ADILA BINTI ABDUL RASHID	2005	> 4 years	3

Table 4. Shift Process Technicians for Thin Film Metal Module

Conclusion can be made that PT who work at metal deposition process is capable to perform machine recovery as per procedure during the event of ILM shutdown. Particle generation during recovery or re-occurrence of shut down due to wrong recovery is zero. Conclusion also can be made that all ET who working with metal deposition machines are capable to perform preventive maintenance jobs effectively and mistake that can lead to particle generation is almost zero.

Shift	Equipment Technician	Date Join	Years of Experiences	Certification level
A	MOHAMMAD RIDZAL BIN ABDULLAH	2002	>7 years	3
	MOHD KHADAFI BIN MAHAMOD	2007	> 2 years	3
B	AYUB BIN AHMAD	2006	> 3 years	3
	FRANCIS SELVAN A/L SINNAYAH	2007	> 2 years	3
C	MOHD ABDUL WAFI BIN AHMAD NADZIR	2003	> 6 years	3
	MOHD AZHUZAIRI BIN ABDULL AZIZ	2007	> 2 years	3
D	KHAIRUL HYFNI BIN NOORDIN	2007	> 2 years	3
	MOHD FAHMI BIN MOHD TAIB	2005	> 4 years	3

Table 5. Shift Equipment Technicians for Thin Film Metal Module

4.2 Material

Validation and verification were made in order to study the relationship between materials used in metal deposition process, with the shutdown rate related to particle, as shown in Table 6. The number of shutdown event (weekly) versus output (wafer move out from equipments) is shown in Figure 16. In average, three machines will be shutdown for every 27,900 wafer output from metal deposition process.

4.2.1 Relationship between shutdown event and output (weekly)

By using Minitab, correlation test between shutdown event and output has been conducted. The Pearson correlation between shutdown and output is 0.005, which means no correlations between both variables.

Regression analysis between shutdown event versus output was made using Minitab. The result of R square (R^2) is zero, which means no relation between shutdown event and output.

Causes		Verification & Validation Process	Result	True/ False
Material	Production Wafers	To check the relationship between shutdown and moves	1. Correlations test between shutdown and moves had been done. The Pearson correlation between shutdown and move is 0.005, which means no correlations between both variables. 2. Regression analysis between shutdown versus Moves was made using Minitab as shown in Figure 4.3. From the result, the R square (R2) is zero, meaning no relation between shutdown and moves.	FALSE
		To check the relationship between shutdown and moves by technologies	1. Pearson Correlation test was done between shutdown and each of the moves by technologies as shown in Figure 4.6. 2. From the Pearson Correlation test, conclusion can be made that no correlation is exist between shutdown and technologies. 3. Base on the regression analysis, no significant relationship between the technologies and the shutdown event. The R2 also showed 0%, which meaning no relationship between the shutdown and technologies	FALSE
		Relationship Between Shutdown and Technologies (STEM Failed)	1. From the Pearson Correlation shown in Figure 4.8, conclusion can be made that no strong correlation between shutdown and technologies. 2. Regression analysis was made between the shutdown and technologies as shown in Figure 4.9 and no strong evidence to conclude that the shutdown is influenced by the technology.	FALSE
		Relationship Between Shutdown and Metal Layers (LM)	1. From Figure 4.11, no strong evidence to say that the shutdown is influenced by the four layer metal lots 2. From Regression analysis in Figure 4.12, also indicates no strong correlation between shutdown and layer metal.	FALSE
	Test Wafers	Relationship Between Incoming Particle from New Test Wafer and Shutdown	All the 50 lots of new test wafer were passed the incoming particle screening. From this study, conclusion can be made that the particle generation is not coming from the new test wafers that were used during qualification process.	FALSE

Table 6. Summary of possible root causes related to materials used in metal deposition

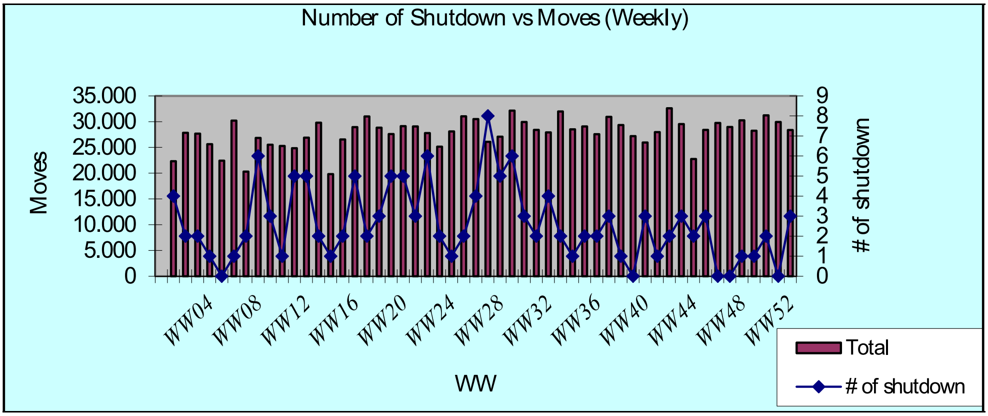


Fig. 16. Number of Shutdown Event versus Machines Output (Weekly)

4.2.2 Relationship between Incoming particle (from new test wafer) and shutdown event

Incoming particle screening was performed for 50 lots of new test wafers. Histogram was generated as shown in Figure 17. From the results, the mean is 1.18 with standard deviation of 1.24. Out of 50 lots of new test wafer that have been measured, 23 lots resulted in zero count of incoming particle, 6 lots showed one count of incoming particle, 10 lots showed two count and 11 lots showed three counts of incoming particle. Since the specification for the incoming particle is five count, all the 50 lots of new test wafer passed the incoming particle screening. From this study, conclusion can be made that the particle generation is not caused by the new test wafers that were used during qualification process

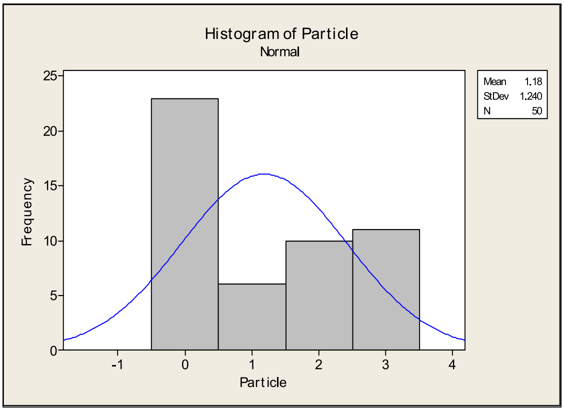


Fig. 17. Histogram for Incoming Particle in New Test Wafer

4.3 Method and measurement

Method of executing the qualification process were observed and summarized. Since the entire MT whose handle metal deposition machines were at level three, therefore they are considered as competent to perform the task efficiently. This is proved by the qualification data that is available in the spreadsheet and also in the CIM system. By looking at the shutdown trend, conclusion can be made that STEM is effective to detect particles that generated at Metal Deposition process.

4.4 Machines

The relationship between shutdown event and machine was studied and the summary of the result is shown in table 7.

Causes		Verification & Validation Process	Result	True/ False
Machine	Vacuum Cassette	To check the particles that may enter a cassette from the cleanroom	Particle and mechanical qualification was done to check the particle existance. Result was clean and no particle was found during wafer loading from production pod to the loadlock	FALSE
	Wafer Handler/ Robot movement	To check the particle that may be created during wafer transfer from vacuum cassette to deposition module.	Particle and mechanical qualification was done to check the particle existance. Found that particle was created during wafer loading from vacuum cassette to deposition module	TRUE
	Gas Line	To check the particle in the gas line	Base on the particle data which is obtained from the particle sensor, no partilce can escape through the particle filter.	FALSE
	Sputter Target	To check the relationship between Sputter Target Life with ILM Shutdown	Base on the correlation analysis, there is relation between Sputter Target Life and ILM Shutdown	TRUE

Table 7. Summary of verify possible root causes related to Metal Deposition Machine

4.4.1 Relationship between shutdown event and particle in vacuum cassette

Results of particle existence in vacuum cassette are shown in Figure 18. From the bar chart, 82 lots detected zero particle count, 13 lots showed particle with one count and five lots showed two counts of particle. Conclusion can be made that particle almost not exist and can be considered as negligible in vacuum cassette, as the production is running under clean room environment of class one category.

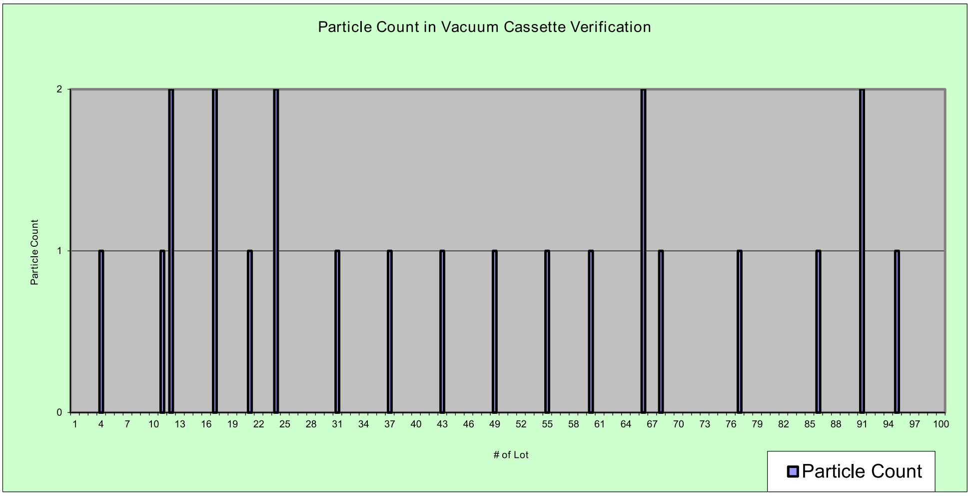


Fig. 18. Bar Chart for Particle Existence in Vacuum Cassette

4.4.2 Relationship between shutdown event and particle generation during mechanical movement

Results for mechanical qualification process are tabulated in table 8. From the data, bar chart was generated as Figure 19. From the bar chart, each of the machines showed particle generation inside the load lock. From the data, conclusion can be made that particles can be generated during wafer handling and mechanical movement.

Tool	Particle Count (adders) for mechanical Qualification Process for all Metal Depositon Tool									
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Metal Dep 01	8	1	5	11	2	1	5	5	9	4
Metal Dep 02	12	3	8	2	2	4	12	5	9	8
Metal Dep 03	3	7	7	14	9	1	7	11	1	7
Metal Dep 04	15	1	7	2	10	2	5	8	2	9
Metal Dep 05	6	7	14	7	5	4	12	6	12	8
Metal Dep 06	3	2	1	3	5	13	9	3	7	11
Metal Dep 07	13	9	4	13	4	1	7	2	15	7
Metal Dep 08	5	3	9	4	2	7	9	3	1	5
Metal Dep 09	4	7	5	4	3	13	1	9	5	9

Table 8. Result for Mechanical Qualification Process for all Metal Deposition Machines

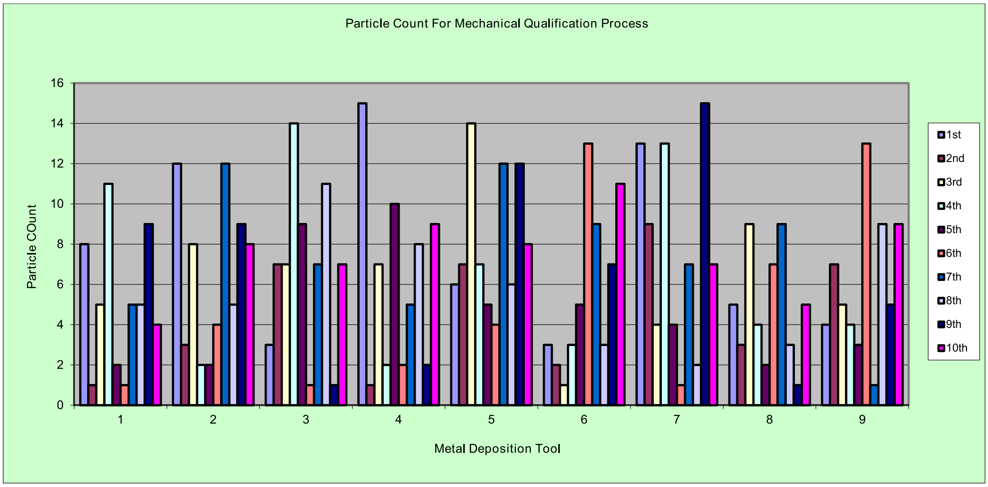


Fig. 19. Particle Count For Mechanical Qualification Process

4.4.3 Relationship between shutdown event and particle generation in gas line

Particle sensor is mounted at the gas line as shown in Figure 14 in section 3.7, to monitor particle existence in gas line. Since the gas lines are equipped with in-line gas filters, the particles are trapped in the filter. Zero reading was captured from the particle sensor. From this study, conclusion can be made that particle is not caused by the gas lines.

4.4.4 Relationship between shutdown event and sputter target life

Result for weekly shutdown event versus target life in Kilowatt per Hour (KW/H) is measured (refer Figure 20). By using Minitab, normality test was done for the target life as shown in Figure 21. The result shows the data is normal and valid to be studied.

Correlations test between shutdown and average sputter target life have been done. The Pearson correlation between shutdown and move is 0.981, which means strong correlations between both variables.

Regression analysis was made between shutdown event and sputter target life. Result for R^2 value of 96.3% indicates strong relationship between shutdown event and sputter target life. Conclusion can be made that shutdown is highly influenced by sputter target life.

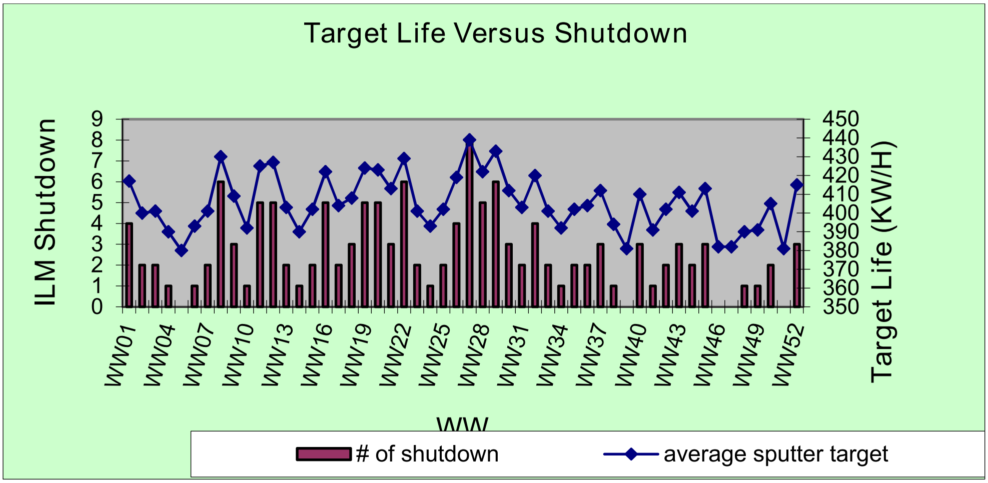


Fig. 20. Average Target Life (KW/H) versus ILM Shutdown (weekly)

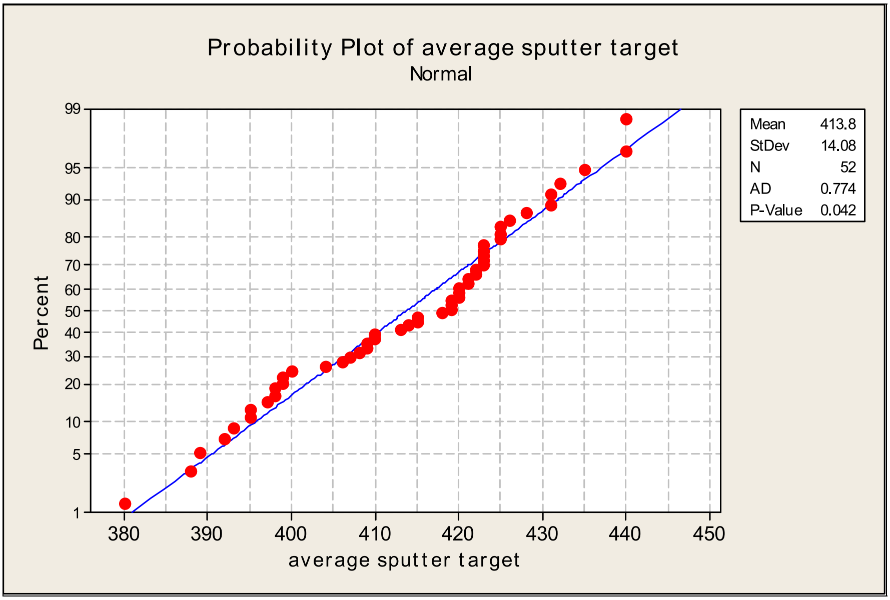


Fig. 21. Normality Test for Sputter Target Life

4.5 Environment

Result for the load lock environment is shown in Figure 22. Zero lots were captured with particle count more than three in the test. Since the specification is less than three counts, it can be concluded that load lock environment is free from particle.

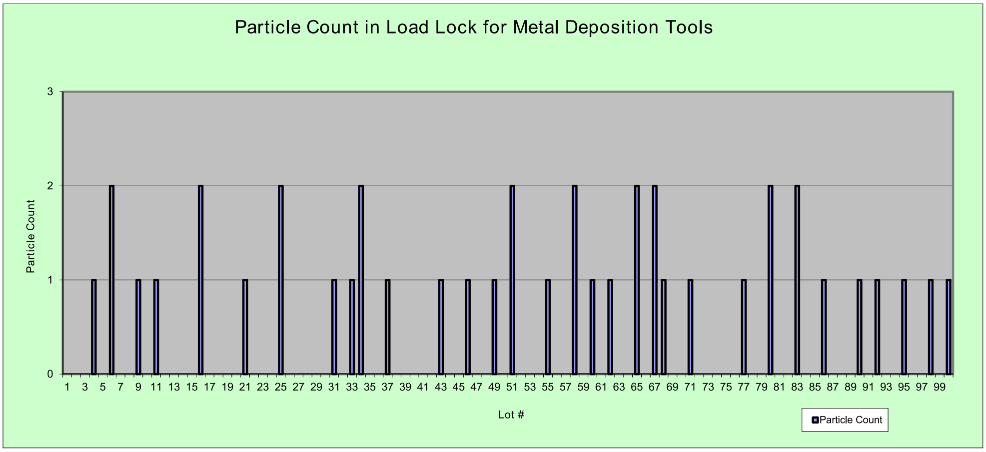


Fig. 22. Particle Count in Load Lock at Metal Deposition Machines

5. Analysis of true causes

From the verification process of potential true causes, wafer transfer in load lock and sputter target life has the most significant relationship to shutdown event. Since the most significant root cause is the sputter target life, this project will only focus on the improvement of sputter target life.

5.1 Sputter target life improvement

From the data of shutdown event versus sputter target life, observation can be made that the rate of machine shutdown is increasing by the increment of sputter target life. Even though the specification for the maximum sputter target life is 450KW/H, however in the study conducted, it shows that the chances of machine shutdown is higher if the sputter target life reach more than 410 KW/H. Zero shutdown per week was observed for sputter target life less than 390 KW/H, ten cases of one shutdown event per week were observed for sputter target life between 390 KW/H- 400 KW/H and 15 cases of two shutdown events per week were observed for sputter target life between 401 KW/H- 410KW/H.

In this study, three machines were selected to be improved by reducing the life of sputter target. Sputter target life is limited to 400 KW/H only, before replace with another new sputter target. Observation of the shutdown event versus the new sputter target life was monitored for three months, and the result is shown in Table 9.

From the data obtained, shutdown event was significantly improved. Three cases of shutdown event were observed within 12 weeks, however the cases were not related to metal deposition machine, but more to incoming particles from previous process.

Since the changes showed significant improvement in the shutdown event related to the particle issues, the new sputter target life reduction from 450 KW/H to 400 KW/H is introduced to the other six machines. Machine specification was updated with this new improvement and Preventive Maintenance job has also been revised to change the sputter target life when it reaches ~ 400 KW/H.

WW	Tool			Total	# of shutdown
	1	2	3		
WW01	2,875	2,790	2,698	8,363	0
WW02	2,650	2,923	2,596	8,169	0
WW03	2,748	2,866	2,777	8,391	1
WW04	2,931	3,257	2,956	9,144	0
WW05	3,096	2,385	2,343	7,824	0
WW06	2,386	3,397	2,248	8,031	0
WW07	2,847	2,711	3,464	9,022	1
WW08	3,186	3,290	3,355	9,831	0
WW09	3,727	3,513	3,628	10,868	0
WW10	3,726	3,053	3,972	10,751	0
WW11	3,527	3,331	3,319	10,177	0
WW12	3,459	3,175	3,736	10,370	1

Table 9. Shutdown versus Move base on new Sputter Target Life

6. Conclusion

Significant improvement can be seen in terms of In-Line Monitoring (ILM) shutdown event after the improvement of new sputter target life. Even though tool shutdown event still appear, it is mainly related to incoming process factors and not due to metal deposition machine. Therefore it is crucial to change the sputter target life when it reaches ~ 400 KW/H.

7. References

Alfred, M. 2001. Reduction of Particulate Contamination in Wafer Processing, pp. 220-225. Santa Clara, California: Applied Materials INC.

Bates, S.P., 2000. Silicon Wafer Processing, Applied Material Summer, US. http://www.mi.e-atech.edu/jonathan.colton/me4210/waferproc.pdf. Accessed on 24 March 2010.

Benjamin, Y.H., Ho-Ahn, L.K. 1987. Particle Technology Laboratory. Mechanical Engineering Department, University of Minnesota, Minneapolis, England.

Bordon, P. 1990. The Nature of Particle Generation in Vacuum Process Machines. IEEE Transactions on Semiconductor Manufacturing 3:189-194.

Chen, D., Seidel, T., Belinski, S. and Hackwood, S. 1989. Dynamic Particulate Characterization of a Vacuum Load- lock System. J. Vac, Science Technology 7:3105-3111.

Chowdhury, N.M. 1997. Designing a Bulk Specialty Gas System for High-Purity Applications. Proceedings of the Institute of Environmental Sciences, pp. 65-72.

Cooper, D.W. 1986. Aerosol Science Technology, pp. 25-34. New York: McGraw- Hill.

Faraday, M. 1857. The Bakerian lecture: experimental relations of gold (and other metals) to light, Philosophical Transactions of the Royal Society of London 147:145–181.

Grove, W.R. 1852. Electro-chemical polarity of gases. Philosophical Transactions of the Royal Society of London 142:87–101.

Hamberg, O.1985. Process Annual technical Meeting of Institute of Environment Sciences, Larrabee, G.B.

Hart, J. and Paterson, A. 1994. Evaluating the Particle and Outgassing Performance of High-Purity. Electronic-Grade Specialty Gas Cylinders Microcontamination 12:63-67.

Hart, J., McDermott, W., Holmer, A. and Natwora J. 1995. Particle Measurement in Specialty Gases, Solid State Technology 38:111-116.

Martin, R.W. 1989. Defect Density Measurement, in Proc. 9th International Symposium Contamination, Los Angeles.

McDermott, W.T. 1997. A Gas Diluter for Measuring Nanometer-Size Particles in Oxygen or Hydrogen. Proceedings of the Institute of Environmental Sciences and Annual Technical Meeting, pp. 26-33

Wu, J.J, Cooper, D.W. and Miller, R.J. 1989. An aerosol model of particle generation during pressure reduction, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films 8:1961-1968.

2

The Fundamentals of Global Outsourcing for Manufacturers

Aslı Aksoy and Nursel Öztürk
Uludag University Department of Industrial Engineering, Bursa
Turkey

1. Introduction

Today, international competition is growing rapidly, and enterprises must always remain ahead of the competition to ensure their survival. Therefore, firms must keep pace with dynamic conditions and rapid changes, be innovative, and adapt to new systems, techniques and technologies. In a competitive market environment, customers are becoming more conscious and tend to demand a particular number of customised products at a particular speed. Furthermore, fluctuations in national economies and in the global economy create significant risks. Because of all of these factors in today's competitive environment, firms have begun to make radical changes in their management and production structures. They must also reduce costs to maintain their current position in the market.

Manufacturers must be the forerunners in the competitive race in today's global markets. Today's enterprises are facing fierce competition, which is forcing them to seriously consider new applications that they can use to improve quality and to reduce cost and lead time. Manufacturers must keep pace with the dynamic requirements of the market and be receptive to reform.

Because of the intense global competition among manufacturers, the supply chain must be able to respond quickly to changes, and customer–supplier relationship management is becoming increasingly important. In recent years, very few manufacturers have owned all of the activities along the supply chain. The ability to make rapid and accurate decisions within the supplier network improves the competitive advantage of manufacturers.

Additionally, due to the intense global competition that exists today, firms should be reevaluate and redirect missing resources. Outsourcing plays a key role for enterprises because the cost of raw materials constitutes a significant part of the cost of the final product. Choosing the right supplier reduces purchasing costs and enhances the competitive advantage of firms. As organisations become more dependent on suppliers, the direct and indirect consequences of poor decision-making become more severe. Decisions about purchasing strategies and operations are the primary determinants of profitability. The globalisation of trade and the Internet have enlarged purchaser choice sets. Changing customer preferences require broader, more systematic and faster outsourcing decision making.

An enterprise may produce a specific product itself or may outsource that specific product to achieve a production cost advantage. Global outsourcing can be defined as the forwarding of specific business to a global supplier. Global outsourcing enhances the competencies of firms while also making firm structures more flexible.

In today's global markets, firms must use new methods to sustain their strength and compete. In recent years, under the influence of this intense competition, global outsourcing has become popular for firms. Firms are widely using global outsourcing to adapt to rapid changes, to reduce the effect of fluctuations, and to take advantage of know-how and current technologies. Global outsourcing allows firms to develop their core competencies and expand their flexibility.

This study reviews the literature on global outsourcing. There has been a great deal of research conducted on global outsourcing within information technology (IT) and service systems. To the best of our knowledge, there have not been many studies on the global outsourcing of manufacturing or production systems. Therefore, this study focused on the global outsourcing of these systems. This paper provides a basic definition of global outsourcing and analyses global outsourcing as either an opportunity or a threat. Furthermore, this study introduces the differences between local and global outsourcing. The methods used to make global outsourcing decisions and the decision criteria used in global outsourcing are also presented in the study.

2. Global outsourcing

Outsourcing is one of the responsibilities of purchasing departments and plays a critical role in an organisation's survival and growth. Materials sourced from outside rather than produced by in-house facilities will influence service quality and profitability (Zeng, 2000). Despite the ongoing debate over the benefits and risks of outsourcing for businesses, outsourcing has become a common approach that purchasing managers cannot ignore. Indeed, outsourcing has exhibited dramatic growth in recent years.

Since the 1980s, the opinions of purchasing managers and management scholars regarding optimal firm sourcing strategies have changed significantly in two respects. First, firms have replaced vertical integration with increased outsourcing based on the conviction that lean, flexible enterprises that focus on their core competencies perform better (Quinn & Hilmer, 1994). Second, in the era of globalisation of the 1990s, enterprises were advised to use the principles of "global outsourcing" to pick the best global suppliers and thereby to improve their competitiveness (Monczka & Trent, 1991; Quinn & Hilmer, 1994). Implementing both or either of these strategies has important consequences for the structure and performance of multinational corporations.

Given the rapidly shifting contours of the global economy, companies need to be able to anticipate changes in the economics and geography of outsourcing. Forward-thinking companies are making their value chains more elastic and their organisations more flexible. Furthermore, with the decline of the vertically integrated business model, outsourcing is evolving into a strategic process used to organize and fine-tune the value chain.

Supplier selection and evaluation play an important role in reducing the cost and time to market while improving product quality. Supplier selection can significantly affect manufacturing costs and production lead time. Although several techniques and models have been used to select and evaluate suppliers, each technique or model has its own strengths and limitations in different situations. Therefore, it is necessary to further improve the performance and effectiveness of supplier selection and evaluation in manufacturing in different contexts.

According to Boer et al. (2001), the purchasing function and purchasing decisions are becoming increasingly important. As organisations become more dependent on suppliers, the direct and indirect consequences of poor decision making are becoming more severe. In addition, several developments have further complicated purchasing decision making. The globalisation of trade and the Internet have expanded the choice sets of purchasers. Changing customer preferences requires broader and more rapid supplier selection.

Outsourcing can be defined as the provision of services by an outside company when those services were previously provided by the home company. In other words, outsourcing involves focusing on a firm's core competencies while allowing services that require other competencies to be provided by other expert enterprises. Outsourcing is a strategic decision in which the buying firm attempts to establish a long-term business relationship with its suppliers (Zeng, 2000).

It is not always easy to generate precise rules for the supplier selection process, but certain elements of the process remain constant. These elements may be identified based on intuition, experience, common sense, or inexplicable rules. Supplier ratings, for example, are usually generated via subjective criteria, based on personal experience and beliefs, based on the available information, and/or sometimes using techniques and algorithms intended to support the decision-making process (Albino & Garavelli, 1998). The key to enhancing the quality of decision making in the supplier selection process is to employ the powerful computer-related concepts, tools and techniques that have become available in recent years (Wei et al., 1997).

In today's competitive global markets, consumers look for the highest quality products at the lowest prices, regardless of where they are produced. This trend is continuously increasing the significance of global markets and forcing enterprises to enter global markets. Furthermore, increasing pressure from foreign competitors in domestic markets is forcing companies to analyse the available alternatives as they seek to remain competitive.

Monczka & Trent (1991) defined global outsourcing as the integration and coordination of procurement requirements across worldwide business units. As such, outsourcing might involve objects, processes, technologies and suppliers. Kotabe (1998) defined global outsourcing as the purchase of finished products or works-in-process from global suppliers. Under this definition, firms may purchase not only products themselves but also the services required to make these products marketable.

Narasimhan et al. (2006) reported that the strategic objectives of global outsourcing are different from those of traditional purchasing. Whereas traditional purchasing focuses on minimizing procurement costs, strategic global outsourcing considers quality, delivery, responsiveness and innovativeness in addition to costs. Sourcing strategies should be incorporated into the operating strategies of buying firms to support or even improve their competitive advantage (Tam et al., 2007). Internal or global outsourcing plays an important role in firm competitiveness and growth (Zeng, 2000).

Flexibility appears to be an important driver of global outsourcing strategy. Firms need to react more quickly to customer requirements, and global outsourcing is seen as a way to accomplish this. Global outsourcing may also be perceived as a way to reduce firm risk by sharing it with suppliers and simultaneously acquire the positive attributes of those suppliers (Kremic et al., 2006). The ultimate objective of global outsourcing strategy is for the firm to exploit both its own and suppliers' competitive advantage and to utilise the comparative location advantages of various countries in global competition (Kotabe & Murray, 2004).

The importance of global outsourcing has increased dramatically. Although firms may outsource for cost-related reasons, there are no guarantees that expected savings will be achieved (Kremic et al., 2006).

Global outsourcing strategy requires close coordination between the research and development, manufacturing, and marketing activities of a firm. Conflicts will most likely exist between the differing objectives of these divisions. For instance, excessive product modification and development intended to satisfy a set of ever-changing customer needs will negatively affect manufacturing efficiency and increase costs. Similarly, excessive product standardization intended to lower manufacturing costs will likely yield lower customer satisfaction levels (Kotabe & Murray, 2004). Therefore, effective global outsourcing requires firms to develop a balance between effective manufacturing and flexible marketing.

Global outsourcing is an expected response to competition. However, the choice of where to obtain goods and services is not an obvious decision. Rather, it is subject to continual reevaluation (Carter et al., 2008). Outsourcing strategy is an essential part of the value chain for corporate activities. Outsourcing strategy both affects and is affected by the other aspects of the supply chain (Kotabe et al., 2008).

The degree of internationalization of production and sourcing is negatively related to the size of the focal country. According to Mol et al. (2005) and Buckley & Pearce (1979) when working with a sample of 156 Japanese, French, Swiss, and "Benelux" companies, found the ratios of global outsourcing to final markets to be 2.4%, 8.0%, 91.6%, and 70.7%, respectively.

As Levy (2005) noted, global outsourcing is highly related to efforts to increase the organizational and technological capacity of firms. Mol et al. (2005) described global outsourcing as balancing international production cost advantage and domestic transaction cost advantage rather than characterising it as a performance-enhancing tool. The major operational problems in global outsourcing, as described by Kotabe et al. (2008), are logistics, inventory management, distance, nationalism, and a lack of working knowledge about foreign business practices.

Global outsourcing has become a popular subject of study in both managerial practice and the academic literature. Conflicting results have been presented in the relevant studies. The global outsourcing strategy literature offers arguments both for and against global outsourcing strategy (Kotabe et al., 2008).

According to Gottfredson et al. (2005), a recent survey of large and medium-sized companies indicates that 82% of large firms in Europe, Asia, and North America have outsourcing arrangements of some kind and that 51% use global outsourcers. However, nearly 50% say that the results achieved by their outsourcing programs have fallen short of expectations. What is more, only 10% are highly satisfied with the decreases in costs that they have achieved, and a mere 6% are highly satisfied with the results of their global outsourcing efforts overall. Mol et al. (2005) stated that global outsourcing can help a firm to enhance its competitive advantage in other markets or to improve its legitimacy. However, multinational supply chains are facing significant managerial problems related to international relations.

According to Kremic et al. (2006), the expected benefits of outsourcing may include providing the same or a better service at a lower overall cost, increased flexibility and/or quality, access to the latest technology and the best talent, and the ability to refocus scarce resources on core functions.

A lack of common methodology is believed to cause some outsourcing failures. Lonsdale (1999) also supported this thinking, suggesting that global outsourcing failures are not due to inherent problems with outsourcing but rather stem from a lack of guiding methodology for managers. Kremic et al. (2006) indicated that global outsourcing has potential pitfalls for strategic reasons. Gillett (1994) noted that enterprises may lose their core competences if they are not careful. If firms outsource the wrong functions, they may develop gaps in their learning or knowledge base that may hinder their ability to capitalise on future opportunities (Kremic et al., 2006). Literature also indicated that in industries with complex technologies and systems, internal synergy may decrease when some functions are outsourced. This could result in lower productivity or efficiency levels for the remaining functions (Quinn & Hilmer, 1994).

Kremic et al. (2006) discussed factors that may impact global outsourcing decisions. These factors are shown below:

- *Core competences*: "Core competences" can be described as a strategic factor that firms use to sustain competitive advantage. Quinn (1999) suggested that there are "core activities" that one firm will perform better than any other firm. In general, a function that is more core to an organization is less likely to be globally outsourced.
- *Critical knowledge:* Some data or knowledge must be under the control of the firm. In general, if a function provides critical knowledge, it is less likely to be globally outsourced.
- *Impact on quality:* The quality of the firm's services establishes its reputation and can create demand. If a firm is currently recognised in the industry for providing a high level of quality in a particular area, then global outsourcing in that area can harm quality. Quality is a relevant factor and can have either a positive or a negative influence on global outsourcing decisions. (Anderson, 1997).
- *Flexibility:* Flexibility includes demand flexibility, process flexibility and resource flexibility. Antonucci et al. (1998) noted that long contracts outsourced into a limited market have sometimes decreased flexibility. However, large enterprises may improve their flexibility via global outsourcing. In the literature, global outsourcing is used as a strategic driver to increase flexibility.
- *Cost:* In the literature, cost is the main reason for global outsourcing decisions. If the firm prefers to outsource a function for cost reasons, then it can be assumed that the current expenditures associated with that function are higher than the expected cost of purchasing the service. However, whether savings will actually accrue from global outsourcing is extremely uncertain. Sometimes, the reported cost savings may not be as high as was expected.

- *Characteristics of the functions outsourced or kept in-house:* In general, the more complex a function the less of a candidate it is for global outsourcing.
- *Integration*: Integration refers to the degree to which function is linked to other functions and systems within the enterprise. The more integrated the function, the more interactions and communication channels there are to maintain and monitor. Therefore, a function that is highly integrated is less of a candidate for global outsourcing.

Firms establish and execute global outsourcing plans in an effort to match competitors' attempts at outsourcing; improve non-competitive cost structures; focus on core competencies; reduce capital investment and overall fixed costs; achieve cost-competitive growth within their supply base for goods, services and technologies in the value chain; and establish future sales footprints in low-cost countries by outsourcing basic goods or business processes (Carter et al., 2008). An effective global sourcing strategy requires continual efforts to streamline manufacturing without sacrificing marketing flexibility (Kotabe & Murray, 2004).

According to the literature, firms prefer global outsourcing for the following reasons:

- *Strategic focus / reduction of assets:* Through global outsourcing activities, an enterprise can reduce its level of asset investment in manufacturing and related areas. Furthermore, global outsourcing can help management teams to redirect their attention to core competencies rather than focusing on maintaining a wide range of competencies (Kotabe et al., 2008).
- *Supplementary power / lower production costs:* Global suppliers are highly specialized in their own business, which lowers both their production costs for those of the firms that are outsourcing their business to them. Therefore, global outsourcing can decrease overall costs if firms globally outsource non-core activities (Quinn, 1999)
- *Strategic flexibility:* Global outsourcing can enhance a firm's strategic flexibility (Harris et al., 1998). If a firm is faced with a crisis in an external environment, it can simply change the volume of globally outsourced products it purchases. If the same product is outsourced to another firm within the home country, the firm will need to pay high reconstructing costs and may not respond quickly to the external environment.
- *Relationship:* Certain relationships with global suppliers can deliver competitive advantage for firms (Kotabe et al., 2008). Misunderstandings between buyer and suppliers may decrease a firm's level of performance (Carter et al., 2008).

Kotabe et al. (2008) suggested that an inverted-U shaped relationship exists between profitability and the degree of outsourcing. On the inverted-U shaped curve, there is an optimal degree of outsourcing for a firm. If the firm moves' away from this optimal point, profitability decreases dramatically.

In global outsourcing strategy, there are also some disadvantages of increasing total product cost. Unfortunately, through global outsourcing, the cost of transportation, communication and information-sharing may increase. Domestic purchasing strategies require only short lead times because they reduce communication and transportation time requirements. The literature suggests that this may be the key reason why some enterprises do not prefer global outsourcing (Dana et al., 2007).

The literature suggests some disadvantages of global outsourcing. These disadvantages can be seen below:

- *The scope of the functions:* If there are important interfaces between activities, decoupling them into separate activities performed by separate suppliers will generate less than optimal results and potential integration problems (Kotabe et al., 2008).
- *Competition loss:* Firms that engage in excessive outsourcing are essentially hollowing out their competitive base (Kotabe, 1998). Furthermore, an enterprise may lose negotiating power with its suppliers because the capabilities of the latter will increase relative to those of the former (Kotabe et al., 2008).
- *Opportunistic behaviour:* Global suppliers may behave opportunistically. Opportunistic behaviour allows a supplier to extract more rents from the relationship than it would normally do, for example, by supplying products of a lower quality than was previously agreed upon or by withholding information regarding changes in production costs (Kotabe et al., 2008).
- *Limited learning and innovation:* Suppliers may capture the critical knowledge by performing the activity. This situation is always a problem between buyer and supplier because both try to obtain all the individual benefits. Appropriation of innovations and rents is always a problem in such a complex buyer–supplier relationships (Nooteboom, 1999)
- *Negative impact of exchange rates:* Higher procurement costs can be seen by the negative impact of fluctuating exchange rates. During the Asian financial crisis, many foreign firms operating in Asian countries learned an invaluable lesson on the negative impact of fluctuating currency exchange rates on their procurement costs and profitability (Kotabe et al., 2008).

According to Kremic et al. (2006), the global outsourcing literature has referenced the following risks of global outsourcing: the potential for both unrealized savings and increased costs, employee morale problems, over-dependence on suppliers, lost corporate knowledge and future opportunities, and under-satisfied customers. Additionally, global outsourcing may fail because the requirements of the relationship are inadequately defined because of a poor contract, a lack of guidance regarding planning or managing outsourcing initiatives, or poor supplier relations. Dana et al. (2007) cited lower production costs as the key advantage of a global outsourcing strategy, with poor control of quality being the main disadvantage.

Lowe et al. (2002) addressed two risks of global outsourcing: fluctuations in exchange rates and relative rates of inflation in different countries. The impact of fluctuations in exchange rates can be analyzed in different ways, and these disparate analyses can yield different results. Brush et al. (1999) stated that many enterprises do not discuss exchange rates as a key factor in global outsourcing. Kouvelis (1999) stated that because of the high cost of switching global suppliers, purchasing managers do not switch suppliers until the effect of exchange rate fluctuation is extremely high. Vidal & Goetschalckx (2000) indicated that the impact of exchange rate fluctuation on overall cost is high.

Under competitive pressure, many U.S. multinational companies globally outsource components and finished products to countries such as China, South Korea, Taiwan, Singapore, Hong Kong, and Mexico. Those countries are also known as low-cost countries (Kotabe & Murray, 2004). Firms in the US and the EU make different choices when selecting global outsourcing locations. In the US, 23% of enterprises prefer China, 14% prefer India, 10% prefer Mexico, 9% prefer Argentina and 8% prefer Brazil. In the EU, 19%

prefer China, 14% prefer the Czech Republic, 12% prefer Poland and 10% prefer Hungary (Timmermans, 2005). The preferences of US and EU firms indicate what is known as "low-cost country sourcing" in the literature. Low-cost country sourcing entails the sourcing of services or functions from low-cost countries with lower labour and material costs. In recent years, low-cost country sourcing has created opportunities for purchasing managers (Carter et al., 2008).

Sourcing from global suppliers can be risky, especially when the projected quality of the outsourced products is unknown. Motwani et al. (1999) noted that as the low-cost countries develop, the quality of the products produced in those countries will likely increase. As a result, firms that choose to forge relationships in these low-cost countries now through sourcing and purchasing may have an edge in these markets in the future. Although they may encounter challenges at first, the advantages that they enjoy in the future could outweigh these problems. This may be especially true for firms that aim to be truly global. Although the main factor driving global outsourcing is lower costs, experienced purchasing managers consider many factors simultaneously in making the decision to outsource internationally. According to the relevant literature, lower labour cost is not the key factor for many US enterprises that engage in global/domestic outsourcing (Sarkis & Talluri, 2002).

3. Outsourcing methods in literature

Outsourcing has been widely discussed in the literature. There are several papers on supplier selection and global outsourcing in information technology (IT) and for service systems. To the best of our knowledge, few studies have been conducted on global outsourcing in manufacturing or production systems. This section is divided into two subsections: one that addresses general supplier selection methods in the literature and another that addresses global outsourcing methods in the literature.

3.1 General supplier selection methods in the literature

There are several methods of general supplier selection presented in the literature. Categorical methods are qualitative models. Based on the buyer's experience and historical data, suppliers are evaluated using a particular set of criteria. The evaluations involve categorizing the supplier's performance as 'positive', 'neutral' or 'negative' with reference to a series of criteria (Boer et al., 2001). After a supplier has been rated for all the criteria, the buyer provides an overall rating, allowing the suppliers to be sorted into three categories.

Data Envelopment Analysis (DEA) is concerned with the efficiency of decision making. The DEA method helps buyers to classify suppliers into two categories: efficient suppliers and inefficient suppliers. Liu et al. (2000) used DEA in the supplier selection process. They evaluated the overall performance of suppliers using DEA. Saen (2007) used IDEA (Imprecise Data Envelopment Analysis) to select the best suppliers based on both cardinal and ordinal data. Wu et al. (2009) proposed an augmented DEA approach to supplier selection. Songhori et al. (2011) presented a structured framework for helping decision makers to select the best suppliers for their firm using DEA.

Cluster Analysis (CA) is a class of statistical techniques that can be used with data that exhibit "natural" groupings (Boer et al., 2001).

Case-Based Reasoning systems (CBR) combine a cognitive model describing how people use and reason from past experience with a technology for finding and presenting experience (Choy et al., 2003-a). Choy et al. (2002-b) enhanced a CBR-based supplier selection tool by combining the Supplier Management Network (SMN) and the Supplier Selection Workflow (SSW). Choy et al. (2005) used CBR to select suppliers in a new product development process.

In linear weighting, the criteria are weighted, and the criterion with the largest weight has the greatest importance. The score for a particular supplier is based on the criteria and their different levels of importance, and some criteria have a high degree of precision. Ghodsypour & O'Brien (1998) integrated the Analytic Hierarchy Process (AHP) and linear programming to consider both tangible and intangible factors in choosing the best suppliers and the optimum order quantities. Lee et al. (2001) used only the AHP for supplier selection. They determined the supplier selection criteria based on purchasing strategy and criterion weights using the AHP. Liu & Hai (2005) used DEA to determine the supplier selection criteria. They then interviewed 60 administrators to determine the priority level of the criteria and used the AHP to select suppliers. Ting & Cho (2008) presented a two-step decision-making procedure. They used the AHP to select a set of candidate suppliers for a firm and then used a Multi-Objective Linear Programming (MOLP) model to determine the optimal allocation of order quantities to those suppliers. Boer et al. (1998) used the ELECTRE 1 technique to evaluate the five supplier candidates. Xia & Wu (2007) used an integrated approach to the AHP, which was improved using rough set theory and multi-objective mixed integer programming to simultaneously determine the number of suppliers to employ and the order quantities to be allocated to these suppliers in the case of multiple sourcing and multiple products. Multiple criteria and supplier capacity constraints were both taken into account. Wang et al. (2004) used an integrated AHP and preemptive goal programming (PGP)-based multi-criteria decision-making process to analyze both the qualitative and quantitative factors guiding supplier selection. Liu and Hai (2005) compared the use of the Voting Analytic Hierarchy Process (VAHP) and the use of the AHP for supplier selection. Chan & Kumar (2007) identified some of the important decision criteria, including risk factors in developing an efficient system of global supplier selection. They used the Fuzzy Extended Analytic Hierarchy Process (FEAHP) to select suppliers. Chan & Chan (2010) used an AHP-based model to solve the supplier evaluation and selection problem for the fashion industry. Kumar & Roy (2011) proposed the use of a rule-based model with the AHP to aid decision makers in supplier evaluation and selection.

Total Cost of Ownership models (TCO) include all costs related to the supplier selection process that are incurred during a purchased item's life cycle. Degraeve & Roodhooft (1999) evaluated suppliers based on quality, price and delivery performance using TCO. They emphasised that uncertainty related to demand, delivery, quality and price must be reflected in the decision problem. Ramanathan (2007) proposed the integrated DEA-TCO-AHP model for supplier selection.

According to Boer et al. (2001), Mathematical Programming models (MP) allows the decision maker to formulate the decision problem in terms of a mathematical objective function that must subsequently be maximized and minimized by varying the values of the variables in the objective function. MP models are more objective than rating models because they force the decision maker to explicitly state the objective function, but MP models often only consider more quantitative criteria. Karpak et al. (1999) developed a supplier selection tool that minimizing costs and maximizing quality reliability. Ghodsypour & O'Brien (1998) integrated the AHP and Linear Programming (LP) models. Their model presented a systematic approach that took into account both qualitative and quantitative criteria. They also developed sensitivity algorithms for different scenarios. Ghodsypour & O'Brien (2001) used mixed integer programming, taking into account the total cost of logistics. Degraeve & Roodhooft (2000) computed the purchasing cost associated with different purchasing strategies using MP. Barla (2003) reduced the number of suppliers from 58 to 10 using the multi-criteria selection method. Hong et al. (2005) decomposed the supplier selection process into two steps. They used cluster analysis to pre-select suppliers and then used MP to select the most appropriate supplier. Yang et al. (2007) studied a supplier selection problem in which a buyer facing random demand must decide the quantity of products it will order from a set of suppliers with different yields and prices. They provided the mathematical formulation for the buyer's profit maximization problem and proposed a solution method based on combining the active set method and the Newton search procedure. Kheljani et al. (2007) considered the issue of coordination between one buyer and multiple potential suppliers in the supplier selection process. In contrast, in the objective function in the model, the total cost of the supply chain is minimized in addition to the buyer's cost. The total cost of the supply chain includes both types of costs. The model was solved using mixed-integer nonlinear programming. Liao & Rittscher (2007) developed a multi-objective programming model, integrating supplier selection to procure lot sizing and carrier selection decisions for a single purchasing item over multiple planning periods during which the demand quantities are known but inconstant. Rajan et al. (2010) proposed a supplier selection model for use in a multiproduct, multi-vendor environment based on an integer linear programming model.

Artificial intelligence (AI)-based systems are computer-aided systems that can be trained using data on purchasing experience or historical data. The available types of AI-based supplier selection applications include Neural Networks (NN) and Expert Systems (ES). One of the important advantages of the NN method is that the method does not require the formulation of the decision-making process. As a result, NNs can cope better with complexity and uncertainty than traditional methods can; these systems are designed to be more similar to human judgment in their functioning. The system user must provide the NN with the properties of the current case. The NN provides information to the user based on what it has learned from the historical data. Albino & Garavelli (1998) further developed the neural network-based decision support system for subcontractor ratings in construction firms. The system includes a back-propagation algorithm. The constructed network is trained using examples so that the system does not require decision-making rules. Vokurka et al. (1996) and Wei et al. (1997) developed an expert system for supporting the supplier selection process. Chen et al. (2006) used linguistic values to assess the ratings and weights of various supplier selection factors. These linguistic ratings were expressed using trapezoidal or triangular fuzzy numbers. Then, they proposed the use of a hierarchy Multiple Criteria Decision-Making (MCDM) model based on fuzzy-sets theory to address supplier selection problems in the supply chain system.

Wang & Che (2007) presented an integrated assessment model for manufacturers to use to solve complex product configuration change problems efficiently and effectively. The model made it possible to determine what fundamental supplier combination would best minimize the cost–quality score if and when proposed by the customer and/or engineer. The researchers combined fuzzy theory, T transformation technology, and genetic algorithms. Liao & Rittscher (2007) studied the supplier selection problem under stochastic demand conditions. Stochastic supplier selection is determined by simultaneously considering the total cost, the quality rejection rate, the late delivery rate and the flexibility rate while also taking into account constraints on demand satisfaction and capacity. The researchers used GA to solve the problem. Wang (2008) developed a decision-making procedure that could be used for supplier selection when product part modifications were necessary. The aim of the research was to determine acceptable near-optimal solutions within a short period of time using a solution-finding model based on Genetic Algorithms (GA). Aksoy & Öztürk (2011) presented a neural network-based supplier selection and supplier performance evaluation system for use in a *just-in-time* (JIT) production environment. Chang et al. (2011) proposed the use of a fuzzy decision-making method to identify evaluation factors that could be used for supplier selection. Jiang & Chan (2011) proposed a method of using a fuzzy set theory with twenty criteria to evaluate and select suppliers.

3.2 Global outsourcing methods in the literature

Canel & Khumawala (1996) proposed a 0-1 mixed integer programming formulation model for international facilities location problems. They determined the location of the international facility and the capacity of that facility. The objective of the model is to maximize the after-tax profit. The proposed model includes different costs, including investment cost, fixed costs, transportation costs, shortage costs and holding costs. The researchers developed two different mathematical models: one for a capacitated case and the other for an uncapacitated case. They used demand and price as the deterministic parameters. Their research could be extended by relaxing the assumptions of deterministic demand, prices, costs, etc. within the problem and treating those factors as stochastic parameters. Huchzermeier & Cohen (1996) developed a stochastic dynamic programming formulation for evaluating global manufacturing strategy options while taking switching costs into account in a stochastic exchange-rate environment. The objective of the model is to maximize after-tax profits. The model includes taxes, fixed and variable costs, capacity and exchange rates. The decision variable in the model is production quantities. The researchers developed different scenarios for different exchange rates. Each model has its own solution. However, the model does not include qualitative parameters. Canel & Khumawala (1997) presented an efficient branch-and-bound procedure for solving uncapacitated, multi-period international facilities location problems. The branch-and-bound problems can be solved using LINDO. The parameters of the model are assumed to be deterministic. Dasu & Torre (1997) presented a model for planning a global supply network for a multinational yarn manufacturer. The objective of the model is to maximize the overall profits of the global supply chain network. The model includes tariffs, exchange rates and transportation costs. The proposed model is non-linear, but the authors make some assumptions in the model to give it a linear structure.

Kouvelis & Gutierrez (1997) solved the newsvendor problem in the textile industry for "style goods". The proposed model determines the production quantities while minimizing shortages and holding costs for a multiple-location manufacturer in a multiple-location market. Shortage costs in this context include the costs associated with lost sales, and holding costs includes the cost associated with excess inventory left over after the selling season. The model includes transportation costs, exchange rate uncertainty and stochastic demand uncertainty, but the model does not include global cost factors such as taxes and tariffs. The study evaluates alternative plans for supply chain design and centralized and decentralized production decision-making mechanisms. The researchers also stated that centralized production decision-making is superior to decentralized production decision-making but that application and control problems are associated with centralized coordination. The proposed model can be easily implemented by purchasing managers. The researchers noted that the production decision-making process can be affected by the uncertainty of global markets and that models with stochastic parameters (such as models for analysing political risk and exchange rate fluctuations) can be used in future research. Munson & Rosenblatt (1997) described local content rules and developed models for selecting global suppliers while satisfying local content provisions. The parameters of the model are deterministic, and the penalty for breaking local content rules is very high. The researchers used the mixed integer programming method to solve the model. The decision variables for the model are the selection of the global supplier and the allocation of orders among the selected suppliers. The objective of the model is to minimize purchasing, production, transportation and fixed costs. The model considers only costs and local content rules. The model does not take into account quantitative parameters or exchange rates.

Coman & Ronen (2000), formulated the global outsourcing problem as a linear programming (LP) problem, identified an analytical solution, and compared that solution with the solutions obtained using the standard cost accounting model and the theory of constraints. The decision variable for the model is the production quantity in terms of preference to manufacture versus preference to outsource. The solution attained indicated that linear programming yielded better results than the two other methods (standard cost accounting and the theory of constraints).

Canel & Khumawala (2001) solved an international facilities location problem using the heuristic method. They developed 12 heuristic methods, but their models do not include quantitative parameters or exchange rates. Vidal & Goetschalckx (2001) presented a model for optimizing global supply that maximizes after-tax profits for a multinational corporation. The model includes transfer prices and the allocation of transportation costs as explicit decision variables. Transfer prices and flows between multinational facilities are calculated in the model. The model does not address the supplier selection problem. The model entails a non-convex optimization problem with a linear objective function, a set of linear constraints, and a set of bilinear constraints. Because the resulting problem is NP hard, the researchers developed a heuristic successive linear programming solution procedure.

Canel & Das (2002) proposed the use of a 0-1 mixed integer programming model to determine for particular time periods which countries a firm should choose as the locations of its global manufacturing facilities. The model was also used to determine the quantity to be produced at each global manufacturing facility and the quantities to be shipped from the global facilities to customers. The study had two goals: to determine the location of the global manufacturing facility and to develop a mathematical model that included global marketing and manufacturing factors. The proposed model does not include quantitative parameters or take exchange rates into account. Hadjinicola & Kumar

(2002) presented a model that includes manufacturing factors together with factory location, inventory, economies of scale, product design, and postponement. In the study, different manufacturing and marketing strategies were evaluated for two different countries. The proposed model is a descriptive model; therefore, the decision variables are not entirely clear. The model can be used to evaluate different manufacturing and marketing strategies in terms of cost and profit functions. The model also includes the effect of exchange rates. Lowe et al. (2002), proposed a model that included exchange rates, using it to choose the location and capacity of global manufacturing facilities in the chemical industry. They analysed data from the year 1982, setting the production capacity, purchasing volume, capacity, fixed and variable costs, transportation costs and taxes for each production facility. To take into account the effect of exchange rates, they reviewed historical data from 22 years and constructed nine different scenarios, calculating the costs for each scenario. They developed a two-stage method of solving the problem. The first stage is a short planning period (as an example 1 year), and the second stage is an optional stage that can be used for long-term planning. However, for longer periods, the first stage can also be repeatedly used.

Teng & Jaramillo (2005) presented a model for global supplier selection in the textile industry. They weighted their criteria and sub-criteria based on expert opinions. The overall scores for the global supplier are calculated by multiplying the score and the weight of the criteria. Goh et al. (2007) presented a stochastic model of the multi-stage global supply chain network problem, incorporating a set of related risks: supply, demand, exchange, and disruption. The objective of the model is to maximize after-tax profits. The model includes demand uncertainty, exchange rates, taxes and tariffs. The researchers composed different scenarios for demand uncertainty and exchange rates and presented different clusters of stochastic parameters. Because the proposed model is a convex linear model, the authors relaxed certain parameters to make the problem linear, but theirs is a descriptive model that is not applied to a real-world scenario.

Lin et al. (2007) proposed the use of a decision model to support global decision making. The model uses two multiple-criteria decision aid techniques (the AHP and PROMETHEE II) and incorporates multiple dimensions (infrastructure, country risk, government policy, value of human capital and cost) into a sensitivity analysis. The authors used the AHP to determine the weight of the criteria and used PROMETHEE to select global suppliers based on weighted criteria. Kumar & Arbi (2008) used a simulation model to forecast lead time and total cost in a global supply environment. Based on their results, it seems that important cost savings can be attained through global outsourcing. However, lead time is an essential factor in real life. The authors stated that global outsourcing is not a viable way to meet short-term market demands but that for large seasonal orders, global outsourcing can be a significant cost-saver. Ray et al. (2008) described the cause of the outsourcing problem, formulated it as a linear programming problem, developed a corresponding function, and offered a simplified criterion for ordering products in terms of preference for manufacturing versus preference for global outsourcing. The authors used a hybrid approach, incorporating the Hurwicz criterion, the theory of constraints (TOC) and linear programming. Some weaknesses of the proposed method are that it is difficult to change the traditional cost accounting system, that it would take time to implement the approach, and that people may be reluctant to use the approach because it requires them to justify their preferences rather than simply saying yes or no. It also requires a new decision-making process.

Wang et al. (2008a) divided firm activities into core activities, core-close activities, core-distinct activities and disposable activities. They used the ELECTRE 1 method to determine which of those four activities can be globally outsourced. Wang et al. (2008b) presented a model for the global outsourcing of logistics activities. They determined the evaluation criteria, ranked the criteria using the AHP and constructed a method using PROMETHEE for global supplier selection. Feng & Wu (2009) presented several tax-saving approaches and developed a tax savings model for maximizing after-tax profit from logistics activities by global manufacturers. Using this model, logistics activities are evaluated in terms of tax savings. It has been observed that the tax saving model has dramatically increased manufacturer profits. The suppliers are selected based on tax savings, whereas any other criteria are disregarded.

Ren et al. (2009) treated global supply chains as agile supply chains and explained agility as the ability to change and adapt quickly to changing circumstances. The model facilitates supplier selection for agile supply chains. The authors determined 10 criteria and 32 sub-criteria for supplier selection. The weights of the criteria were determined based on expert opinions and used to rank the suppliers. Perron et al. (2010) presented a mathematical model for multinational enterprises to use to determine transfer prices and the flow of goods between global facilities. The model includes bilinear constraints; therefore, the authors relaxed the constraints to simplify the model. They developed a branch-and-cut algorithm and two different heuristics to solve the model. The heuristic methods can be summarised as follows:

1. *Variable Neighbour Search Method (VNS):* This method is based on the concept of systematic changes in neighbourhoods during the search. VNS explores nearby and then increasingly far neighbourhoods for the best-known solution in a probabilistic fashion. Therefore, often favourable characteristics of the best-known solution will be kept and used to obtain promising neighbouring solutions. VNS was repeatedly used with a local search routine to transition from these neighbouring solutions to local optima.
2. *Alternate Heuristic (ALT):* Given two subsets of variables, the ALT heuristic solves the problem by alternately fixing the variables of one of the subsets. The subsets of variables must be such that the model becomes linear when fixing the variables of one of the subsets. When one of these linear programs is solved, its solution becomes a set of parameters in the other one. ALT can be converging to local optima.

The objective of the model is to maximize after-tax profit given taxes, capacity, transfer prices and demand. Satisfactory results were reported when small problems were solved using heuristic methods.

4. Outsourcing decision criteria

To be competitive in the global market, firms must attain the knowledge necessary to systematically evaluate all potential suppliers and select the most suitable ones. The factors most often used in current supplier evaluation are quality, supplier certification, facilities, continuous improvement, physical distribution and channel relationships (Weber, 1991).

In the supplier selection process, it is not always easy to recognize precise rules, but there is, in general, a coherent way to solve the problem. This coherence can be rooted in intuition, experience, common sense, or inexplicable rules. Supplier rating is then a problem usually solved by subjective criteria, based on personal experiences and beliefs, on the available information and, sometimes, on techniques and algorithms supporting the decision process (Albino & Garavelli, 1998). The key to enhancing the quality of decision making in supplier selection include the powerful computer-related concepts, tools and techniques that have become available in recent years (Wei et al., 1997).

Chao et al. (1993) concluded that quality and on-time delivery are the most important attributes of purchasing performance. Ghodsypour & O'Brien (1998) agreed that cost, quality and service are the three main factors that should influence supplier selection. Brigs (1994) stated that joint development, culture, forward engineering, trust, supply chain management, quality and communication are the key requirements of supplier partnerships apart from optimum cost. Petroni & Braglia (2000) evaluated the relative performance of suppliers with multiple outputs and inputs, considering management, production facilities, technology, price, quality, and delivery compliance. Wei et al. (1997) examined factors such as supply history, product price, technological ability and transport cost.

Making sourcing decisions based on delivery speed and cost is the best way to improve performance (Tan 2001). Global outsourcing reduces the fixed investment costs of a firm in its own economic region. Today, in making global outsourcing decisions, many enterprises also consider quality, reliability, and technology when evaluating the components and products to be procured (Kotabe & Murray, 2004) rather than only considering price. In developing global outsourcing strategies, firms must consider not only manufacturing costs, the costs of various resources, and exchange rate fluctuations but also the availability of infrastructure (including transportation, communications, and energy), industrial and cultural environments, and ease of working with foreign host governments, among others (Kotabe et al., 2008).

Several factors influence global outsourcing decisions. Canel & Das (2002) outlined the factors that most commonly influence global manufacturing facility locations. Those factors are labour and other production inputs, political stability, the attitude of the host government towards foreign investment, host government tax and trade policies, proximity to major markets, access to transportation and the existence of other competitors. Choy et al. (2005) stated that good customer–supplier relationships are necessary for an organization to respond to dynamic and unpredictable changes. They considered price, delivery, quality, innovation, technology level, culture, commercial awareness, production flexibility, ease of communication and current reputation when selecting and evaluating suppliers. Teng & Jaramillo (2005) used five main criteria and 20 sub-criteria for global supplier selection in the textile-apparel industry. Those criteria are delivery (geographic location, freight terms, trade restrictions and total order lead time), flexibility (capacity, inventory availability, information sharing, negotiability and customization), cost (supplier selling price, internal costs, ordering and invoicing), quality (continuous improvement programs, certification, customer service and the percentage of on-time shipments), reliability (feelings of trust, the national political situation, the status of the currency exchange and warranty policies).

Narasimhan et al. (2006) composed a model, for global supplier selection and order allocation and considered criteria such as direct product cost, the indirect cost of coordination, quality, delivery reliability and complexity of the supply base. Lin et al. (2007) used infrastructure, country risk, government policy, human capital and cost when evaluating global suppliers. Carter et al. (2008) analysed the low-cost countries and their capabilities. They evaluated the factors such as labour cost, work ethic, intellectual property, market attraction, delivery reliability, reliable transportation, transportation costs, government support for business, political stability, flexibility, predictable border crossing and corruption in 12 different low-cost countries using perceptual mapping. They stated that experienced purchasing managers not only consider cost but also conduct a multi-criteria evaluation of global outsourcing decisions.

Au & Wong (2008) identified four main categories of factors from the literature: cost (labour costs, material costs and transportation costs), product quality (technological capabilities, reliability and trust), time to market (geographical proximity and transportation time) and country factors, including both internal country factors (such as infrastructure and ethical issues) and external country factors (such as the political and economical situation and social, linguistic and cultural differences).

Chan et al. (2008) examined the decision variables influencing global supplier selection and identified five main criteria and 19 sub-criteria: total cost of ownership (product cost, total logistics management cost, tariffs and taxes), product quality (conformance with specifications, product reliability, quality assessment techniques and process capabilities), service performance (delivery reliability, information sharing, flexibility, responsiveness and customer responses), supplier background (technological capabilities, financial status, facilities, infrastructure and market reputation), and risk factors (geographical location, political stability and foreign policies, exchange rates and economic position, terrorism and the crime rate).

Ku et al. (2010) identified the following criteria as important to global supplier selection: cost (product price, freight costs and custom duties), quality (rejection rate, process capabilities and quality assessments), service (on-time delivery, technological support, responses to changes and ease of communication), risk (geographical location, political stability and the status of the economy). Ku et al. (2010) suggested that qualitative criteria (e.g., the characteristics of the purchased items) be considered in future research.

5. Conclusion

Manufacturers must be the forerunners to be competitive in today's global markets. That is why manufacturers must keep in touch with the dynamic requirements of the market and be receptive to reforms. An increasing proportion of raw materials and work-in-process (WIP) for manufactured products is sourced globally by multinational manufacturers in today's industries.

To become a world-class manufacturer, a firm must not only compete globally in the marketplace but also be competitive and consistent in terms of costs, technological leadership, and quality. High-quality inputs are becoming the focus of many purchasing departments.

The design of global supply chains has been a challenging optimization problem for many years. In a continuing effort to remain competitive, many firms are considering new sources for their raw materials and components, new locations for their production and distribution facilities, and new markets in which to sell their products without regard for national boundaries. The design of global supply chains has been a challenging optimization problem for many years. The current globalization of the economy is forcing firms to design and manage their supply chains efficiently on a worldwide basis.

It is well known that large enterprises no longer operate in a single market. In seeking to penetrate global markets and obtain their benefits, firms are under excess pressure to reduce the price of their products and thus their production and material costs.

According to the literature, global outsourcing is mainly analyzed in terms of cost. During the evaluation of global outsourcing, neither multi-criteria evaluations nor qualitative assessments are always made, but many decision criteria must be considered when configuring a global supply chain system. Historically, labour cost has been one of the most decisive factors in global outsourcing decision making. Recently, the rapidly changing business environment has increased pressure on decision makers to properly analyze the relevant decision criteria. Strategic decision making requires the use of tangible, intangible, strategic and operational decision criteria. There are many factors that need to be considered simultaneously, and purchasing managers need a structured method of using these criteria in their decision making.

The AHP is the widely used a multi-criteria, decision-making method used by academicians and practitioners for supplier selection and global outsourcing decision making. The AHP is based on dual comparisons of decision criteria. As the number of decision criteria increases, the complexity of the system also increases. Mathematical models are not sufficient for evaluating qualitative criteria.

One of the major factors complicating the modelling of global outsourcing decision making is "uncertainty". Exchange rate fluctuations, variable transportation times, demand uncertainty, the variability of market prices, and political instability are among the most important sources of uncertainty. An effective decision-making methodology for global outsourcing must address those uncertainties. In recent years, artificial intelligence tools such as neural networks, fuzzy logic and genetic algorithms have been increasingly used for outsourcing decision making. Those methods are more appropriate under uncertainty and can better address qualitative criteria.

6. References

Aksoy, A., & Öztürk, N. (2011). Supplier Selection And Performance Evaluation In Just-In-Time Production Environments. *Expert Systems with Applications*, Vol. 38, No. 5, pp. 6351-6359, ISSN: 0957-4174

Albino, V., & Garavelli, A.C. (1998). A Neural Network Application to Subcontractor Rating in Construction Firms. *International Journal of Project Management*, Vol. 16, No. 1, pp. 9-14, ISSN: 0263-7863

Anderson, M.C., (1997). A Primer in Measuring Outsourcing Results. *National Productivity Review,* Vol. 17, No.1, pp. 33-41, ISSN 0277-8556

Antonucci, Y.L., Lordi, F.C., & Tucker J.J. (1998). The Pros and Cons of IT Outsourcing. *Journal of Accountancy*, Vol.185, No.6, ISSN: 1945-0729

Au, K.F., & Wong, M. C. (2008). Decision Factors In Global Textile And Apparel Sourcing After Quota Elimination. *The Business Review*, Vol.9, No.2, pp. 153-157, ISSN 1540 - 1200

Barla, S.B. (2003). A Case Study Of Supplier Selection For Lean Supply By Using A Mathematical Model. *Logistics Information Management*, Vol.16, No.6, pp. 451-459, ISSN 0957-6053

Boer, L., Labro, E., & Morlacchi, P. (2001). A Review of Methods Supporting Supplier Selection. *European Journal of Purchasing & Supply Management*, Vol.7, pp. 75-89, ISSN: 1478-4092.

Boer, L., Wegen, L., & Telgen, J. (1998). Outranking Methods In Support Of Supplier Selection. *European Journal of Purchasing & Supply Management*, Vol.4, pp. 109-118, ISSN: 1478-4092

Briggs, P. (1994). Case Study: Vendor Assessment For Partners In Supply. *European Journal of Purchasing and Supply Management*, Vol. 1, No. 1, pp. 49–59, ISSN 0969-7012

Brush, T.H., Martin, C.A., & Karnani, A. (1999). The Plant Location Decision in Multinational Manufacturing Firms: An Empirical Analysis of International Business and Manufacturing Strategy Perspectives. *Production and Operations Management*, Vol.8, pp.109-132, ISSN: 1937-5956

Canel, C., & Das, S.R. (2002). Modeling Global Facility Location Decisions: Integrating Marketing And Manufacturing Decisions. *Industrial Management and Data Systems*, Vol.102, No.2, pp. 110-118, ISSN 0263-5577

Canel, C., & Khumawala, B.M. (2001). International Facilities Location: A Heuristic Procedure For The Dynamic Uncapacitated Problem. *International Journal of Production Research*, Vol.39, No.17, pp. 3975-4000, ISSN 0020-7543

Canel, C., & Khumawala, B.M. (1997). Multi-Period International Facilities Location: An Algorithm And Application. *International Journal of Production Research*, Vol.35, No.7, pp. 1891- 1910, ISSN 0020-7543

Canel, C. & Khumawala, B.M. (1996). A Mixed-Integer Programming Approach For The International Facilities Location Problem. *International Journal of Operations & Production Management*, Vol.16, No.4, pp. 49-68, ISSN 0144-3577

Carter, J.R., Maltz, A., Yan, T., & Maltz, E. (2008). How Procurement Managers View Low Cost Countries and Geographies. *International Journal of Physical Distribution & Logistics Management*, Vol.38, No.3, pp. 224-243, ISSN: 0960-0035

Chan F.T.S., & Chan H.K. (2010). An AHP Model For Selection Of Suppliers In The Fast Changing Fashion Market. *International Journal of Advanced Manufacturing Technology*, Vol. 51, pp. 1195–1207, ISSN 0268-3768

Chan, F.T.S., & Kumar, N. (2007). Global Supplier Development Considering Risk Factors Using Fuzzy Extended AHP-Based Approach. *The International Journal of Management Science,* Vol.35, pp. 417-431, ISSN 0305-0483

Chan, F.T.S., Kumar, N., Tiwari, M. K., Lau, H. C. W., & Choy, K. L. (2008). Global Supplier Selection: A Fuzzy-AHP Approach. *International Journal of Production Research*, Vol.46, No.14, pp. 3825–3857, ISSN: 0020-7543

Chang B., Chang C., & Wu C. (2011). Fuzzy DEMATEL Method For Developing Supplier Selection Criteria. *Expert Systems with Applications*, Vol. 38,pp. 1850–1858, ISSN 0957-4174

Chao, C., Scheuing, E.E., & Ruch, W.A. (1993). Purchasing performance evaluation: an investigation of different perspectives. *International Journal of Purchasing and Materials Management,* Vol.29, No.3, pp. 33–39, ISSN 1055-6001

Chen, C.T., Lin, C.T., & Huang, S.F. (2006). A Fuzzy Approach For Supplier Evaluation And Selection In Supply Chain Management. *International Journal of Production Economics*, Vol.102, pp.289–301, ISSN: 0925-5273

Choy, K.L., Lee, W.B., Lau, H.C.W., & Choy, L.C. (2005). A Knowledge Based Supplier Intelligence Retrieval System For Outsource Manufacturing. *Knowledge-Based System*, Vol.18, pp. 1-17, ISSN 0950-7051

Choy, K.L., Lee, W.B., & Lo, V. (2003). Design Of A Case Based Intelligent Supplier Relationship Management System- The Integration Of Supplier Rating System And Product Coding System. *Expert Systems with Applications*, Vol. 25, pp. 87-100, ISSN: 0957-4174

Choy, K.L., Lee, W.B., & Lo, V. (2002). On The Development Of A Case Based Supplier Management Tool For Multi-National Manufacturers. *Measuring Business Excellence*, Vol.6, No.1,pp. 15-22, ISSN : 1368-3047

Coman, A., & Ronen, B. (2000). Production Outsourcing: A Linear Programming Model For The Theory-Of-Constraints. *International Journal of Production Research*, Vol. 38, No.7, pp. 1631-1639, ISSN: 0020-7543

Dana, L.P., Hamilton, R. T., & Pauwels, B. (2007). Evaluating Offshore and Domestic Production in the Apparel Industry: The Small Firm's Perspective. *Journal of International Entrepreneurship*, Vol.5, pp. 47–63, ISSN1570-7385

Dasu, S., & Torre, J. (1997). Optimizing an International Network of Partially Owned Plants Under Conditions of Trade Liberalization. *Management Science*, Vol.43, No.3, pp. 313-333, ISSN: 0025-1909

Degraeve, Z., & Roodhooft, F. (2000). A Mathematical Programming Approach For Procurement Using Activity Based Costing. *Journal of Business Finance and Accounting*, Vol. 27, No.1&2, pp. 69-98, ISSN 1468-5957

Degraeve, Z., & Roodhooft, F. (1999). Improving The Efficiency Of The Purchasing Process Using Total Cost Of Ownership Information: The Case Of Heating Electrodes At Cockerill Sambre S.A.. *European Journal of Operational Research*, Vol.112, pp. 42-53, ISSN 0377-2217

Feng, C.M., & Wu, P.J. (2009). A Tax Savings Model For The Emerging Global Manufacturing Network. *International Journal of Production Economics*, Vol. 122, No.2, pp. 534-546, ISSN 0925-5273

Ghodsypour, S.H., & O'Brien, C. (1998). A Decision Support System For Supplier Selection Using An Integrated Analytic Hierarchy Process And Linear Programming. *International Journal of Production Economics*, Vol. 56-57, pp. 199-212, ISSN 0925-5273

Ghodsypour, S.H., & O'Brien, C. (2001). The Total Cost Of Logistics In Supplier Selection, Under Conditions Of Multiple Sourcing, Multiple Criteria And Capacity Constraints. *International Journal of Production Economics*, Vol. 73, pp. 15-27, ISSN 0925-5273

Gillett, J. (1994). Viewpoint. The Cost-Benefit of Outsourcing: Assessing The True Cost of Your Outsourcing Strategy. *European Journal of Purchasing & Supply Management*, Vol. 1, No. 1, pp. 45-7, ISSN: 1478-4092

Goh, M., Lim, J.Y.S., & Meng, F. (2007). A Stochastic Model For Risk Management In Global Supply Chain Networks. *European Journal of Operational Research*, Vol.182, pp. 164–173, ISSN 0377-2217

Gottfredson, M., Puryear, R., & Phillips, S. (2005). Strategic Sourcing From Periphery to the Core. *Harvard Business Review*, February, ISSN: 0017-8012

Hadjinicola, G.C., & Kumar, K.R. (2002). Modeling Manufacturing And Marketing Options In International Operations. *International Journal of Production Economics*, Vol.75, pp. 287–304, ISSN: 0925-5273

Harris, A., Giunipero, L.C. , Tomas, G., & Hult, M. (1998). Impact of Organizational and Contract Flexibility on Outsourcing Contracts. *Industrial Marketing Management*, Vol.27, pp. 373–384, ISSN: 0019-8501

Hong, G.H., Park, S.C., Jang, D.S., & Rho, H.M. (2005). An Effective Supplier Selection Method For Constructing A Competitive Supply-Relationship. *Expert Systems with Applications*, Vol. 28, pp. 629-639, ISSN: 0957-4174

Huchzermeier, A., & Cohen, M.A. (1996). Valuing Operational Flexibility under Exchange Rate Risk. *Operations Research, Special Issue on New Directions in Operations Management, Vol.*44, No.1, pp 100-113, ISSN 0030-364X

Jiang W., & Chan F.T.S. (2011). A New Fuzzy Dempster MCDM Method And Its Application In Supplier Selection. *Expert Systems with Applications*, Vol. 38, No.8, pp. 9854-9861, ISSN 0957-4174

Karpak, B., Kasuganti, R.R., & Kumcu, E. (1999). Multi-Objective Decision-Making In Supplier Selection: An Application Of Visual Interactive Goal Programming. *Journal of Applied Business Research*, Vol.15, No.2, pp. 57-71, ISSN 0892-7626

Kheljani, J.G., Ghodsypour, S.H., & O'Brien, C. (2007). Optimizing Whole Supply Chain Benefit Versus Buyer's Benefit Through Supplier Selection. *International Journal of Production Economics*, Vol. 121, No.2, pp. 482-493, ISSN: 0925-5273

Kotabe, M. (1998). Efficiency vs. Effectiveness Orientation of Global Sourcing Strategy: A Comparison of U.S. and Japanese Multinational Companies. *Academy of Management Executive*, Vol. 12, No. 4, pp. 107-119, ISSN: 1079-5545

Kotabe, M., Mol, M.J., & Murray, J.Y. (2008). Outsourcing, Performance, and The Role of E-Commerce: A Dynamic Perspective. *Industrial Marketing Management*, Vol.37, pp. 37–45, ISSN: 0019-8501

Kotabe, M., & Murray, J.Y. (2004). Global Sourcing Strategy and Sustainable Competitive Advantage. *Industrial Marketing Management*, Vol. 33, pp. 7– 14, ISSN: 0019-8501

Kouvelis, P., & Gutierrez, G.J. (1997). The Newsvendor Problem in a Global Market: Optimal Centralized and Decentralized Control Policies For A Two Market Stochastic Inventory System. *Management Science*, Vol.43, No.5, pp.571-585, ISSN: 0025-1909

Kremic, T., Icmeli Tukel, O., & Rom, W. O. (2006). Outsourcing Decision Support: A Survey Of Benefits, Risks, And Decision Factors. *Supply Chain Management: An International Journal*, Vol. 11, No.6, pp. 467–482, ISSN: 1359-8546

Ku, C.Y., Chang, C.T., & Ho, H.P. (2010). Global Supplier Selection Using Fuzzy Analytic Hierarchy Process And Fuzzy Goal Programming. *Quality & Quantity*, Vol. 44, No.4, pp. 623-640, ISSN: 0033-5177

Kumar, S., & Arbi, A.S. (2008). Outsourcing Strategies For Apparel Manufacture: A Case Study. *Journal of Manufacturing Technology Management*, Vol.19, No.1, pp. 73-91, ISSN 1741-038x

Kumar J., & Roy N. (2011). Analytic Hierarchy Process (AHP) For A Power Transmission Industry To Vendor Selection Decisions. *International Journal of Computer Applications*, Vol.12, No.11, pp. 26-30, ISSN 0975 - 8887

Lee, E.K., Sungdo, H., & Kim, S.K. (2001). Supplier Selection And Management System Considering Relationships In Supply Chain Management. *IEEE Transactions on Engineering Management,* Vol.48, No.3, pp. 307-318, ISSN: 0018-9391

Levy, D.L. (2005). Offshoring in the New Global Political Economy. *Journal of Management Studies*, Vol.42, No.3, pp. 685-693, ISSN 1467-6486

Liao, Z., & Rittscher, J. (2007). Integration Of Supplier Selection, Procurement Lot Sizing And Carrier Selection Under Dynamic Demand Conditions. *International Journal of Production Economics,* Vol.107, pp. 502–510, ISSN: 0925-5273

Lin, Z.K., Wang, J.J., & Qin, Y.Y. (2007). A Decision Model For Selecting An Offshore Outsourcing Location: Using A Multicriteria Method. *Proceedings of the 2007 IEEE International Conference on Service Operations and Logistics, and Informatics*, pp. 1-5, ISBN 978-1-4244-1118-4, Philadelphia, PA, USA, August 27-29, 2007

Liu, J., Ding, F.Y., & Lall, V. (2000). Using Data Envelopment Analysis To Compare Suppliers For Supplier Selection And Performance Improvement. *Supply Chain Management: An International Journal*, Vol. 5, No.3, 143-150, ISSN: 1359-8546

Liu, F.F.H., & Hai, H.L. (2005). The Voting Analytic Hierarchy Process Method For Selecting Supplier. *International Journal of Production Economics*, Vol.97, No.3, pp. 308-317, ISSN 0925-5273

Lonsdale, C. (1999). Effectively Managing Vertical Supply Relationships: A Risk Management Model for Outsourcing. *Supply Chain Management: An International Journal*, Vol. 4 No. 4, pp. 176-83, ISSN 1359-8546

Lowe, T.J., Wendell, R.E., & Hu, G. (2002). Screening Location Strategies to Reduce Exchange Rate Risk. *European Journal of Operational Research*, Vol.136, pp. 573-590, ISSN 0377-2217

Mol, M.J., Tulder, R.J.M., & Beije, P.R. (2005). Antecedents and Performance Consequences of International Outsourcing. *International Business Review*, Vol.14, pp. 599–617, ISSN 0969-5931

Monczka, R.M., & Trent, R.J. (1991). Global Sourcing: A Development Approach, *International Journal of Purchasing and Materials Management*, No. Fall, pp. 2-6.

Motwani, J., Youssef, M., Kathawala, Y., & Futch, E. (1999). Supplier Selection in Developing Countries: A Model Development. *Integrated Manufacturing Systems*, Vol.10, No.3, pp.154-161, ISSN 0957-6061

Munson, C.L., & Rosenblatt, M.J. (1997). The Impact Of Local Content Rules On Global Sourcing Decisions. *Production and Operations Management*, Vol.6, No.3, pp. 277-290, ISSN 1937-5956

Narasimhan, R., Talluri, S., & Mahapatra, S. K. (2006). Multiproduct, Multicriteria Model For Supplier Selection With Product Life-cycle Considerations. *Decision Sciences*, Vol. 37, No. 4, pp. 577-603, ISSN 1540-5915

Nooteboom,B. (1999). *Inter-Firm Alliances : Analysis and Design*. Routledge. ISBN 0-415-18154-2, USA

Perron, S., Hansen, P. , Digabel, S., & Mladenovic, N. (2010). Exact And Heuristic Solutions Of The Global Supply Chain Problem With Transfer Pricing. *European Journal of Operational Research*, Vol. 202, No.3, pp. 864-879, ISSN 0377-2217

Petroni, A., & Braglia, M. (2000). Vendor Selection Using Principal Component Analysis. *Journal of Supply Chain Management,* Vol.36, No.2, pp. 63–69

Quinn, J.B. (1999). Strategic Outsourcing: Leveraging Knowledge Capabilities. *Sloan Management Review*, Vol. 40,No. 4, pp. 9-21

Quinn, J.B., & Hilmer, F.G. (1994). Strategic Outsourcing. *Sloan Management Review*, Vol. 35, No. 4, pp. 43-55

Rajan, A.J., Ganesh, K., & Narayanan, K.V. (2010). Application of Integer Linear Programming Model for Vendor Selection in a Two Stage Supply Chain. *International Conference on Industrial Engineering and Operations Management,* Dhaka, Bangladesh, January 9-10, 2010.

Ramanathan, R. (2007). Supplier Selection Problem: Integrating DEA With The Approaches Of Total Cost Of Ownership And AHP. *Supply Chain Management: An International Journal,* Vol. 12, No.4, pp. 258-261, ISSN: 1359-8546

Ray, A., Sarkar, B., & Sanyal, S. (2008). A Holistic Approach For Production Outsourcing. *Strategic Outsourcing: An International Journal,* Vol.1, No.2, pp. 142-153, ISSN 1753-8297

Ren, J., Yusuf, Y. Y., & Burns, N. D. (2009). A Decision-Support Framework For Agile Enterprise Partnering. *The International Journal of Advanced Manufacturing Technologies*, Vol. 41, pp.180–192, ISSN 0268-3768

Saen, R.F. (2007). Suppliers Selection In The Presence Of Both Cardinal And Ordinal Data. *European Journal of Operational Research,* Vol. 183,pp. 741–747, ISSN: 0377-2217

Sarkis, J., & Talluri, S. (2002). A Model For Strategic Supplier Selection. *Journal of Supply Chain Management*. Vol.38, No.1, pp.18-28

Songhori, M.J., Tavana, M., Azadeh, A., & Khakbaz, M.H. (2011). A Supplier Selection And Order Allocation Model With Multiple Transportation Alternatives. *International Journal of Advanced Manufacturing Technology*, Vol.52, No.1-4, pp. 365–376, ISSN 0268-3768

Tam, F.Y., Moon, K.L., Ng, S.F., & Hui, C.L. (2007). Production Sourcing Strategies and Buyer-Supplier Relationships: A Study of the Differences Between Small and Large Enterprises in the Hong Kong Clothing Industry. *Journal of Fashion Marketing and Management*, Vol. 11, No.2, pp. 297-306, ISSN: 1361-2026

Tan, B. (2001). On Capacity Options In Lean Retailing. *Harvard University, Center for Textile and Apparel Research, Research Paper Series*. February.

Teng, S.G., & Jaramillo, H. (2005). A Model For Evaluation And Selection Of Suppliers In Global Textile And Apparel Supply Chains. *International Journal of Physical Distribution & Logistics Management*, Vol.35, No.7/8, pp. 503-523, ISSN 0960-0035.

Timmermans, K. (2005). The Secrets of Successful Low-Cost Country Sourcing. *Accenture,* Vol.2, pp. 62-72

Ting, S.C., & Cho, D.I. (2008). An Integrated Approach For Supplier Selection And Purchasing Decisions. *Supply Chain Management: An International Journal,* Vol. 13, No.2, pp. 116-127, ISSN : 1359-8546

Vidal, C.J., & Goetschalckx, M. (2001). A Global Supply Chain Model With Transfer Pricing And Transportation Cost Allocation. *European Journal of Operational Research*, Vol.129, pp. 134-158, ISSN 0377-2217

Vidal, C.J., & Goetschalckx, M. (2000). Modeling the Effect of Uncertainties on Global Logistics Systems. *Journal of Business Logistics*. Vol.21, No.1, pp. 95-120, ISSN 0197-6729

Vokurka, R., Choobineh, J., & Vadi, L. (1996). A Prototype Expert System For The Evaluation And Selection Of Potential Suppliers. *International Journal of Operations & Production Management*, Vol. 16, No.12, pp. 106-127, ISSN: 0144-3577

Wang, G., Huang, S.H., & Dismukes, J.P. (2004). Product-Driven Supply Chain Selection Using Integrated Multi-Criteria Decision-Making Methodology. *International Journal of Production Economics,* Vol. 91, pp. 1-15, ISSN: 0925-5273

Wang, H.S., & Che, Z.H. (2007). An Integrated Model For Supplier Selection Decisions In Configuration Changes. *Expert Systems with Applications,* Vol.32, pp. 1132–1140, ISSN 0957-4174

Wang, H.S. (2008). Configuration Change Assessment: Genetic Optimization Approach With Fuzzy Multiple Criteria For Part Supplier Selection Decisions. *Expert Systems with Applications,* Vol. 34, pp. 1541-1555, ISSN: 0957-4174

Wang, J.J, Hu, R.B., & Diao, X.J. (2008-a). Developing an Outsourcing Decision Model Based on ELECTREI Method. *4th International Conference on Wireless communications, Networking and Mobile Computing*, pp. 1-4, ISBN 978-1-4244-2107-7, Dalian, October 12-14, 2008.

Wang, J.J., Gu, R., & Diao, X.J. (2008-b). Using A Hybrid Multi-Criteria Decision Aid Method For Outsourcing Vendor Selection. *4th International Conference on Wireless communications, Networking and Mobile Computing*, pp. 1-4, ISBN 978-1-4244-2107-7, Dalian, October 12-14, 2008.

Weber, C.A., Current, J.R., & Benton, W.C. (1991). Vendor Selection Criteria And Methods. *European Journal of Operational Research*, Vol. 50, pp. 2-18,I SSN 0377-2217

Wei, S., Zhang, J., & Li, Z. (1997). A Supplier Selecting System Using A Neural Network, *IEEE International Conference on Intelligent Processing Systems*, pp. 468-471, ISBN: 0-7803-4253-4, Beijing China, October 28-31, 1997.

Wu, D. (2009). Supplier selection: A Hybrid Model Using DEA, Decision Tree And Neural Network. *Expert Systems with Applications,* Vol.36, pp. 9105-9112, ISSN: 0957-4174

Xia, W., & Wu, Z. (2007). Supplier selection with multiple criteria in volume discount Environments. *The International Journal of Management Science*, Vol. 35, pp. 494-504, ISSN: 0305-0483

Yang, S., Yang, J., & Abdel-Malek, L. (2007). Sourcing With Random Yields And Stochastic Demand: A Newsvendor Approach. *Computers & Operations Research,* Vol.34, pp.3682 – 3690, ISSN: 0305-0548

Zeng, A.Z. (2000). A Synthetic Study of Sourcing Strategies. *Industrial Management & Data Systems*, Vol.100, No.5, pp.219-226, ISSN: 0263-5577

An Approach-Based on Response Surfaces Method and Ant Colony System for Multi-Objective Optimization: A Case Study

Barthélemy H. Ateme-Nguema[1], Félicia Etsinda-Mpiga[2], Thiên-My Dao[2] and Victor Songmene[2]
[1]Université du Québec en Abitibi-Témiscamingue
[2]École de Technologie Supérieure, Montreal, Canada

1. Introduction

Numerical optimization is the search approach we adopted in order to determine the best mechanical design. Several algorithms related to the problems at hand were developed, most of them single-objective. However, given the complexity of the products involved and the multiple objectives of the design considered, the researchers focused on the optimization algorithms for such problems. In short, the optimization problems have multiple objectives, and in many cases, there are multiple constraints. Design processes often require expensive evaluations of objective functions. That is particularly the case when such performance indexes and their constraints are obtained through intermediate simulations by finite elements involving fine meshes, many freedom degrees and nonlinear geometrical behaviours. To overcome these difficulties, the response surface method (RSM) is employed (Myers & Montgomery, 2002; Roux et al., 1998; Stander, 2001; Zhang et al., 2002) to replace a complex model by an approximation based on results calculated on certain points in search space.

When an adequate model is obtained with the RSM approach, it then becomes necessary to consider the optimization step. The method used to find the best solution assesses several objectives simultaneously; since some such objectives are fundamentally conflicting vis-à-vis of another, we therefore need to establish a compromise. Existing literature shows that desirability or metaheuristic functions are normally used, the most common being the genetic algorithm (GA). Sun & Lee (Sun & Lee, 2005), present an approach which associates the RSM and GA with the optimal aerodynamic design of a helicopter rotor blade. The ACO is a metaheuristic, which has been successfully used to solve several combinatorial optimization problems. We however see that very little exists in terms of documentation for optimization using ACO, as far as multiobjective problems are concerned. Some works lead us to believe that ant colonies can produce an optimum situation faster than the GA (Nagesh, 2006; Liang, 2004). In the literature, ACOs are used almost exclusively for "*Travelling Salesman Problem*" (TSP), quadratic assignment problem allocation (QAS), constraint satisfaction problems (CSP), design manufacturing systems (DMS), and for discrete and combinatorial optimization problems. Our contribution consists in an extension of the ACO in the multiobjective optimization of mechanical system design in a continuous field. This paper starts with the modelling process with RSM, and then goes on to describe the ACO and the Hybrid method developed for a problem regarding multiobjective optimization with constraints. An application of the suggested method for optimizing the mechanical process design is presented.

2. Modeling with RSM

RSM is a collection of statistical and mathematical techniques used to develop, improve and optimize processes (Myers & Montgomery, 2002). Furthermore, it has important applications in the design and formulation of new products, as well as in the improvement of existing products.

The objective of RSM is to evaluate a response, i.e., the objective physical quantities, which are influenced by several design variables. When we use RSM, we seek to connect a continuous answer Y with continuous and controlled factors X_1, X_2... X_k, using a linear regression model which can be written (Myers & Montgomery, 2002) as:

$$y = f_{\beta}(X_1, X_2, ..., X_p) + \varepsilon \tag{1}$$

Since the response surface is described by a polynomial representation, it is possible to reduce the optimization resolution process time by assessing the objectives with their models rather than using more complex empirical models such as those obtained through the FEM analysis. Although the specific form of response factor f_β is unknown, experience shows that it can be significantly approximated using a polynomial.

In the case of two factors, the linear regression model is one of the simplest available, and corresponds to a first-degree model with interaction, and which has the following form:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \varepsilon \tag{2}$$

Whenever this model is unable to describe the experimental reality effectively, it is common practice to use a second-degree model, which includes the quadratic effects of the factors involved:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \varepsilon \tag{3}$$

Where y is the response (study objective, for example, the total manufacturing cost); ε is the estimate of the error; X_1 and X_2 are influential factors of the coded response (e.g., design variables).

The unknown parameters of this mathematical model, βi values, are estimated through the least-squares technique, and the adjustment quality of the model is assessed using traditional multiple linear regression tools.

Ideally, the number of experiments carried out, either with the finite element model (FEM) or using other approaches, during the application of RSM, should be as small as possible, in order to reduce data-processing requirements. Properly selecting the points to be used for the simulation will allow a reduction of the variance of the coefficients of the mathematical model, which will in turn ensure that the response surfaces obtained are more reliable. To that end, we need to determine the experimental design to be adopted in order to obtain the most interesting simulation for this problem. The central composite design (CCD) was employed in the case of the second-order response surface, but other types of plans, such as the complete factorial design and the fractional factorial design, are also available for use.

Once the mathematical models are obtained, we need to verify that they produce an adequate approximation of the actual study system. The statistic selection criterion is the coefficient of determination R^2, which must be as close as possible to 1 ($0<R^2<1$).

Once this stage is completed, we will have all the equations which make up our multiobjective optimization problem. Generally, such problems are as the following form:

$$
\begin{array}{ll}
\text{Find} & x = [x_1, x_2, x_3, \ldots, x_n]^T \\
\text{Which minimize} & f(x) = \{f_1(x), f_2(x), f_3(x), \ldots, f_n(x)\} \\
\text{Subject to} & g_j(x) \le 0 \quad \text{for } j = 1, m \\
 & x_i^L \le x_i \le x_i^U \quad \text{for } i = 1, n
\end{array}
\tag{4}
$$

To optimize this problem, we explored the ant colony algorithm (ACO). Some options are offered for this kind of problem, such as the desirability function and the genetic GA used by some authors, such as Sun & Lee (Sun & Lee, 2005) or Abdul-Wahad & Abdo (Abdul-Wahad & Abdo, 2007). The literature shows that for many problems, the ant colony approach produces better results in terms of quality solutions and resolution speed, as compared to the GA. This allowed us to begin this research with the resolution of the multiobjective continuous optimization problem in mechanical design.

3. Ant colony algorithm approach

The ACO metaheuristic, called the ant system (Dorigo, 1992), was inspired by studies of the behaviour of ants (Deneubourg et al., 1983; Deneubourg & Goss, 1989; Goss et al., 1990), as a multi-agent approach for resolving combinative optimization problems such as the TSP.

Ants communicate among themselves through the "*pheromone*", a substance they deposit on the ground in variable amounts as they move about. It has been observed that the more ants use a particular path, the more pheromone is deposited on that path, and the more attractive it becomes to other ants seeking food. If an obstacle is suddenly placed on an established path leading to a food source, ants will initially go randomly right or left, but those choosing the side that is in fact shorter will reach the food more quickly, and will make the return journey more often. The pheromone concentration on the shorter path will therefore be more strongly reinforced, and it will eventually become the new preferred route for the stream of ants; however, it must also be borne in mind that the pheromone deposited along the way does evaporate. Works by Colorni et al. (Colorni et al., 1992), Dorigo et al. (Dorigo et al., 1996; Dorigo et al., 1999; Dorigo et al., 2000) provide detailed information on the operation of the algorithm and on the determination of the values of the various parameters (see Fig. 1).

In our field, the ACO has been used very sparingly, and has been focused primarily on single-objective problems (Chegury, 2006). For multiobjective problems, the ACO has hardly been used at all (Zhao, 2007), and when used, has been mainly on combinatorial optimization problems. The importance of this work therefore lies in its attempt to adapt continuous ant colonies to multiobjective problems.

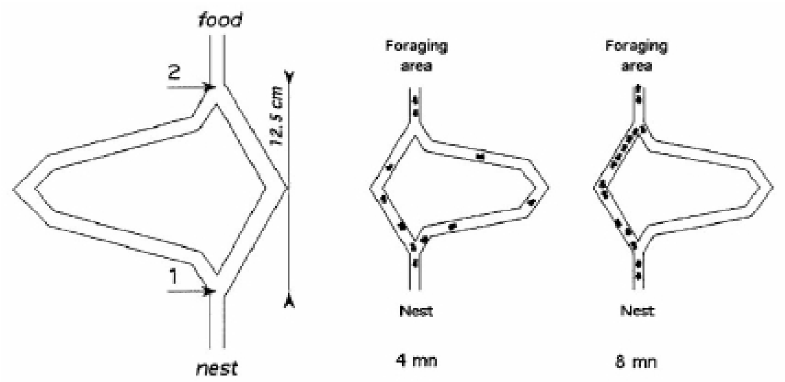


Fig. 1. Experimental setup and drawings of the selection of short branches by a colony of "*Linephitema humile*", 4 and 8 min after the bridge was placed (Dorigo et al., 2000)

4. Proposed design optimization approach

The objective of this chapter is to determine the best design for a mechanical system such as a plane wing, an engine, etc., or for an unspecified mechanical process that sometimes simultaneously optimizes several conflicting objectives. The ACO, like the GA, requires an objective function which can be quickly assessed. We use RSM modeling to determine such objective functions, and the ACO as the research method. Reducing the resolution time in the optimization process requires a reduction of the preciseness of the assessment of objective functions, since we use an approximate modeling of our objectives instead of their exact representations.

Each objective f_i is expressed according to the variables of real design x_i which influence its value. The multiobjective optimization model obtained with RSM is:

$$\begin{aligned} &\text{Minimize} \quad f(x) = \{f_1(x), f_2(x), f_3(x), \ldots, f_n(x)\} \\ &\text{subject to} \quad x_i^L \le x_i \le x_i^U \quad \text{for } i = 1, n \end{aligned} \tag{5}$$

After obtaining a mathematical model for that problem, the optimization phase must be able to determine the best compromise solution for the various objectives.

The steps for the general ACO metaheuristic for compromise solutions for combinatorial problems presented by Gagné et al. (Gagné et al., 2004) constitute an interesting approach to be considered in our resolution process for developing the fitness function.

4.1 Continuous ant colonies

There are several ant colony algorithms available for continuous optimization, the first of which was developed by Bilchev & Parmee (Bilchev & Parmee, 1995), and named CACO

(Continuous Ant Colony Optimization), using ant colonies for local searches and calling upon revolutionary algorithms for global searches. Ling et al. (Ling et al., 2002) present an unspecified hybrid algorithm whose main premise is to consider the differences between two individuals on each dimension as many parts of a path on which the pheromones are deposited. The evolution of the individuals dealt with mutation and crossing-over operators. This method thus tries to reproduce the construction mechanism of the solution component by component.

Monmarché et al. (Monmarché et al., 2000) developed the API algorithm which takes the primitive ant behaviour of the species *Pachycondyla Apicalis*, and which does not use indirect communication by tracks of pheromone. In this method, it is necessary to start by positioning a nest randomly on the research space, after which ants are distributed randomly over it. These ants explore their "*hunting site*" locally by evaluating several points within a given perimeter. Socha (Socha, 2004) presents the ACO algorithm for continuous optimization which tries to maintain the iterative construction solutions for continuous variables. He considers that the components of all solutions are formed by the various optimized variables. Moreover, before considering the algorithm from the ant's point of view, he opts to operate at the colony level, with the ants being simply points to be evaluated. Pourtakdoust & Nobahari (Pourtakdoust & Nobahari, 2004) developed the CACS (Continuous Ant Colony System) algorithm, which is very similar to that of Socha. Indeed, in CACS, as is the case with ACO, for continuous optimization, the core of the algorithm consists in evolving a probability distribution which for CACS is normal.

4.2 Proposed algorithm

Once the steps used in making a choice regarding the elements to be included in the resolution process are explained. We present the new proposed algorithm for our approach (see Fig. 2a and Fig. 2b).

Step 1: System configuration

Determine the objectives of this study, the constraints and the variables which can influence these objectives. Evaluate the field of application of these variables.

Step 2: RSM

Set up an experimental design, carry out tests, and model the various objectives according to influential parameters.

Step 3: Seek ideal point

Using RSM, determine distinct optimum for each study objective.

Step 4: Optimization function formulation

a. State user preferences (weighting of the objectives).
b. The various objectives are expressed in a single function: the fitness function. It acts as an equation which for each objective, expresses the standard and balanced distance at the ideal point F^* of an unspecified solution k, whose various objectives are given by F^k. This function makes it possible to standardize objectives in order to reduce the adverse effects obtained from the various measuring units, as well as the extent of the field of the variables, in order to not skew the fitness function (Gagné et al., 2004):

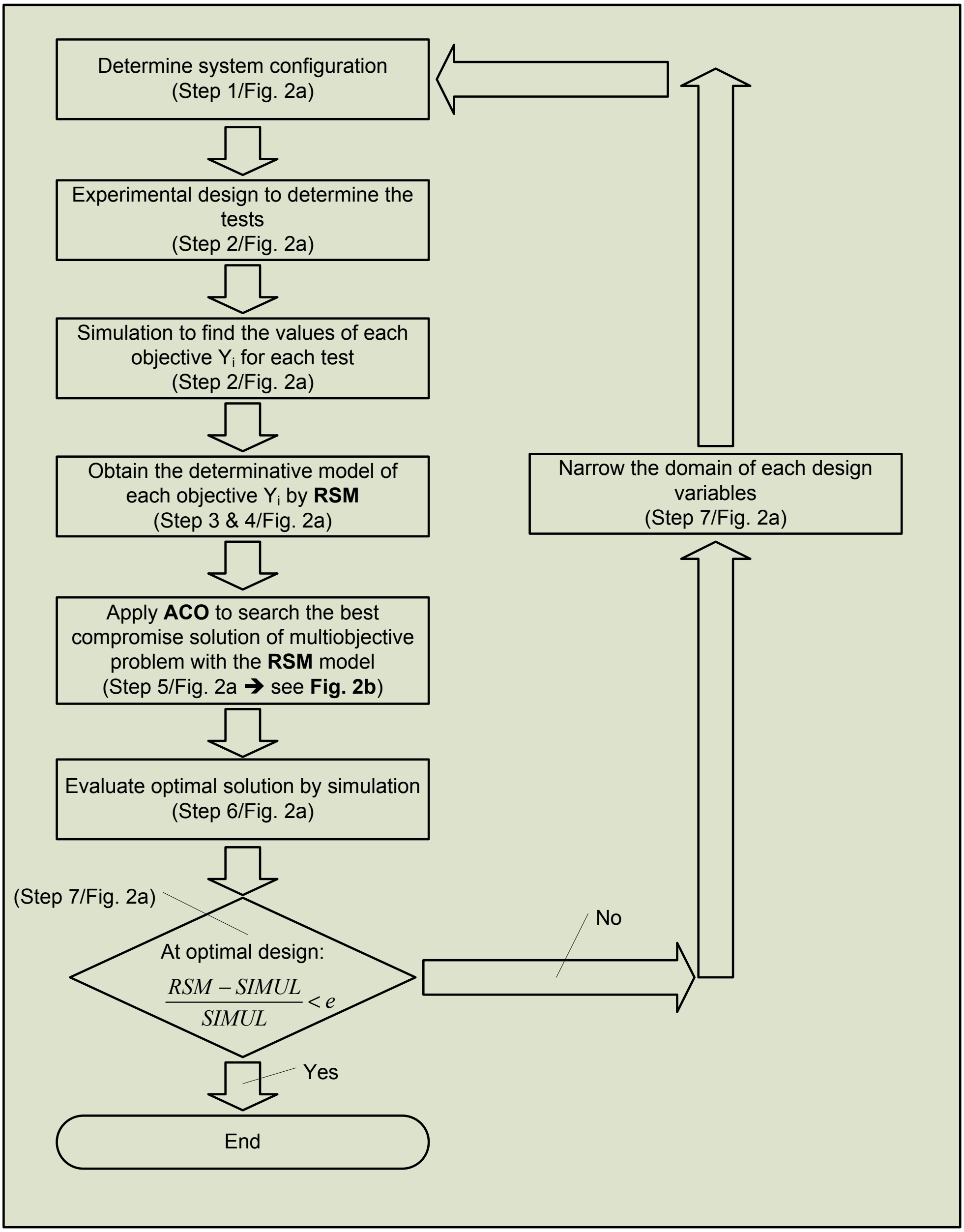


Figure 2a. Flow chart of the optimization approach

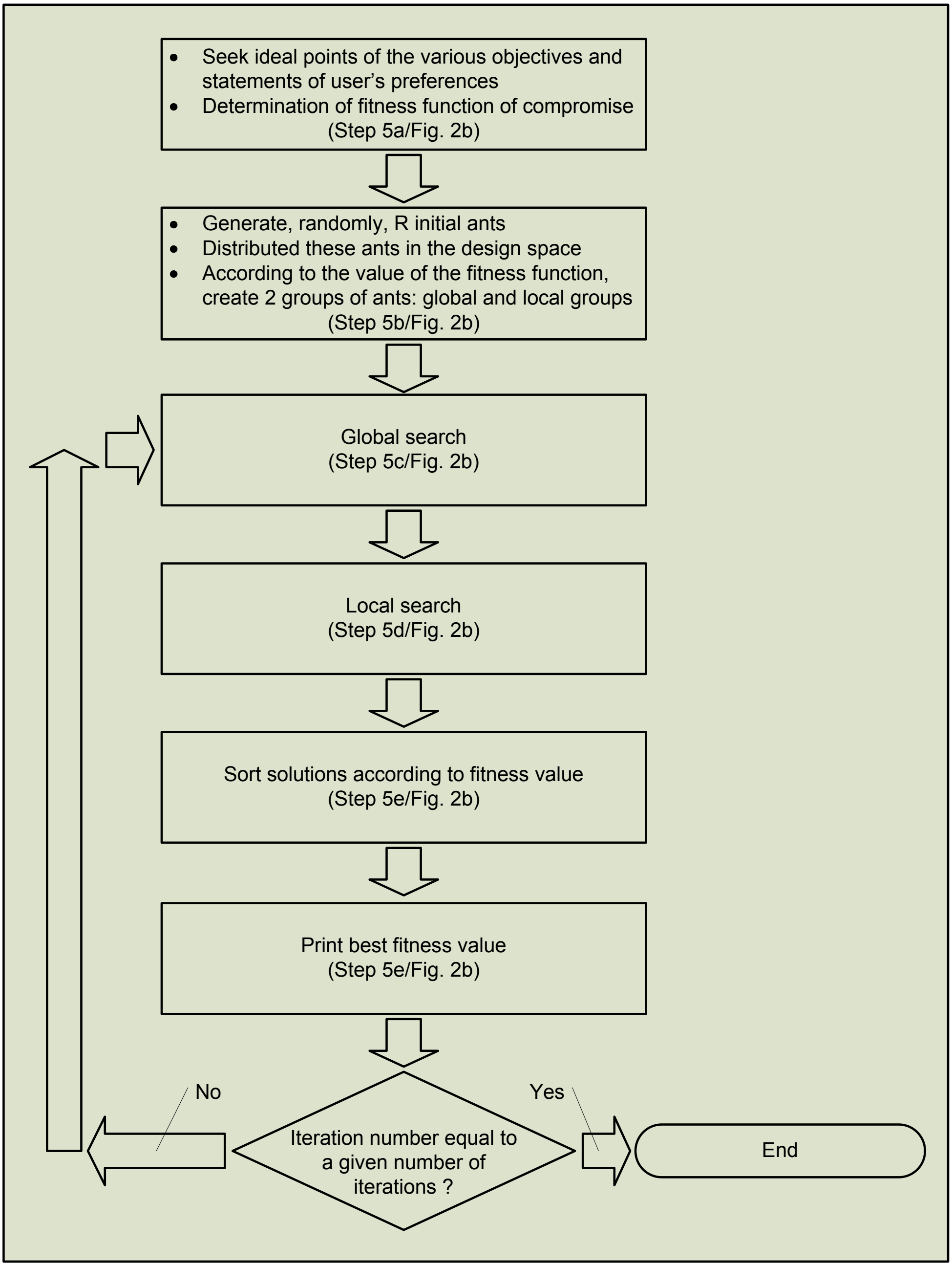


Figure 2b. Flow chart of the ACO process

$$fitness = \left(\sum_{i=1}^{Z} p_i \left(\frac{F_i^k - F_i^*}{F_i^{nad} - F_i^*} \right)^2 \right)^{\frac{1}{2}} \tag{6}$$

where F* is a solution vector corresponding to the ideal point of each separate objective, and probably expressed by:

$$F^* = \left\{ F_1^*, F_2^*, \ldots, F_Z^* \right\} \tag{7}$$

where $F_i^* = \min_{x \in S} f_i(x)$.

F* generally corresponds to an unrealizable solution. S is the space of acceptable search, and F^{nad} is the Nadir point, which represents the maximum values for each objective in the set of optimal Pareto solutions:

$$F^{nad} = \left\{ F_1^{nad}, F_2^{nad}, \ldots, F_z^{nad} \right\} \tag{8}$$

where $F_i^{nad} = \max_{x \in S^*} f_i(x)$.

Step 5: Determine compromise solution

a. Randomly generate R initial ants corresponding to feasible solutions.

To apply the ACO methodology for continuous optimization function problems, the field must be subdivided into a specific area, R, distributed by chance. Next, we need to generate feasible solutions representing the initial ants, each forming a part of the research area to be explored.

b. The fitness function of these solutions is assessed, and the values obtained are lines in descending order.

We obtain our initial ants "R" and the proportion of the higher values of R will be taken to constitute the global ants "G".

c. Apply a global search to a percentage of the initial ants, with "G" constituting the "*worst*" solutions available.

The percentage of global ants is an important parameter of CACO, which can be changed depending on the problem at hand. A global search creates new solutions for "G" by replacing the weaker parts of the existing field. This process is composed primarily of two genetic operators. In the terminology of CACO, these are called *random walk and trail diffusion*. In the random search process, the ants move in new directions in search of more recent and richer sources of food.

In the CACO simulation, a global search is conducted in all fields through a process that is equivalent to a GA crossover and mutation.

- *Crossover or random walk*: The crossover operation is conducted to replace inferior solutions with superior ones, with the crossover probability (CP).

- *Mutation*: The replaced solutions are further improved by mutation. The mutation step is completed in CACO by making an addition or proportional subtraction to the mutation probability. The mutation step size is reduced or increased as per Eq. 9. (Mathur et al., 2000)

$$\Delta = R(1 - r^{(1-T)^b}) \tag{9}$$

where r is a random number from 0 to 1, R is the maximum step size, T is the ratio of the current iteration number and that of the total number of iterations, and b is a positive parameter controlling the degree of nonlinearity.

- *Trail diffusion*: In this step, the field of the global search is gradually reduced, as the search progresses. This reduction makes it possible to increase the probability of locating the optimum through more concentric search procedures. Trail diffusion is similar to the arithmetic GA crossover. In this step, two parents are randomly selected from the parent population space. The elements of the child's vector can be any one of the following:

1. The child corresponds to an element from the first parent
2. The child corresponds to an element from the second parent
3. The child is a combination of the parents (Eq.10) (Mathur et al., 2000)

$$X(child) = (\alpha) X_{i(parent1)} + (1-\alpha) X_{i(parent2)} \tag{10}$$

where α is a uniform random number ranging from [0 to 1]

The probability of selecting one of the three options depends on the mutation probability. Thus, if the mutation probability is 0.5, option 3 can be selected with a probability of 50%, whereas the probability of selecting option 1 or 2 is 25%.

d. Send local ants *L* in the various *R* areas

Once the global search is completed, the zones to which you send the local ants are defined and a local search can begin.

In a local search, the local ants choose the area to be explored among the areas of the matrix R, according to the current quantity of pheromones in the areas. The probability of choosing a solution "i" is given by: (Mathur et al., 2000)

$$P_i(t) = \tau_i(i) \Big/ \sum_k \tau_k(t) \tag{11}$$

where "i" is the solution index and $\tau_i(t)$ is the pheromone trail on the solution "i" at time "t".

After choosing its destination, the ant proceeds across a short distance. The search direction remains the same from one local solution to the next as long as there is improvement in the fitness function. If there is no improvement, the ant reorients itself randomly to another direction. If an improvement in the fitness function is obtained in the preceding procedure, the position vector of the area is updated. The quantity of pheromone deposited is proportional to the improvement of the fitness function. If, in the search process, a higher fitness function value is obtained, the age of the area is increased. This age of the area is another major parameter in the CACO algorithm. The size of the ant displacement in a local search depends on the current age. The search ray is maximum for age zero, and minimal for the maximum age, with a linear variation.

e. Evaluate the fitness function for each ant obtained, and continue the iterative process, beginning with a global search until stop conditions are observed.

Step 6: Evaluate the best solutions (quasi optimal) by simulation or experimentation for the experimental design (Example, FEM)

Step 7: Evaluate the stop criterion $\frac{RSM - SIMUL}{SIMUL} < e$ with SIMUL being the simulation result.

In the optimization design problem for a mechanical system, the number of design variables is very often equal to or higher than 3, and each one of them has a broad field of variation. Consequently, in our resolution process, it is possible for the search field for each design variable to be gradually narrowed for as long as the stop criterion has not been encountered. The search process ends when $\frac{RSM - SIMUL}{SIMUL} < e$ with e being a margin of error defined beforehand.

5. Application: Numerical example of two-objective problem

In order to illustrate the performances of the recommended resolution approach used in this paper, we carried out the optimization of a multistage flash desalination process. The problem was taken from Abdul-Wahab & Abdo (Abdul-Wahab & Abdo, 2007), and was solved using the experimental designs, and optimized using desirability functions.

5.1 Problem definition

Multistage flash (MSF) desalination is an evaporation and condensation process, which involves boiling seawater and condensing the vapour to produce distilled water. A more extensive description of the multistage flash desalination MSF considered in this work can be found in Hamed et al. (Hamed et al., 2001).

In this study, two performance objectives are considered: the maximization of the distillate produced rate (DF) and the minimization of the blow down flow rate (BDF). The operation variables which influence these objectives are presented in Table 1 (**Step 1**). They include:

Parameter name	Nomenclature	Low level	High level
Seawater inlet temperature (ºC)	SWIT (A)	24	35
Temperature difference (ºC)	TD (B)	5.2	8.0
Last-stage brine level (mm)	LSBL (C)	50	850
First-stage brine level (mm)	FSBL (D)	40	320
Brine recycle pump flow (m3/h)	BRPF (E)	8200	11 500

Table 1. Design parameters

5.2 Modeling with RSM

To express our objectives according to decision variables, we need to use modeling with RSM (**steps 2 and step 3**). We considered the experiments carried out by Abdul-Wahab & Abdo (Abdul-Wahab & Abdo, 2007), which helped us to design our model.

Abdul-Wahab & Abdo resorted to a two-level factorial design, carried out 64 experiments and five central-point tests with design variables coded on two levels: low (-1) and high (+1). The experimental design provides us with a linear regression model coded for each response in this study (see Fig. 3 & Fig. 4).

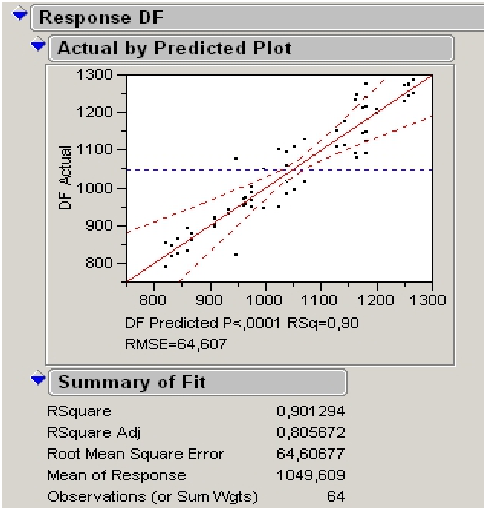


Fig. 3. Modeling with RSM – part I

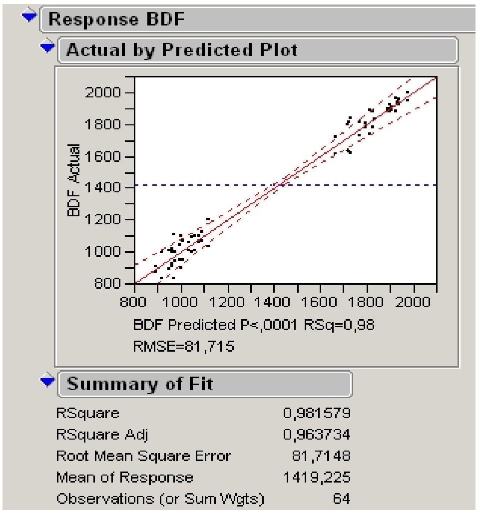


Fig. 4. Modeling with RSM – part II

Finally, the equations representing our objectives are:

$$\begin{cases} DF &= 1041.61 + 20.45(B) + 18.65(C) + 120.29(E) \\ & \quad +26.46(AC) + 30.08(CD) - 25.06(ABE) \\ BDF &= 1419.22 + 414.54(C) + 34.77(D) \\ & \quad +28.71(AC) - 25.63(ABD) \end{cases} \quad (12)$$

We also note that in our experimental design, the variables C and E are the most influential on our objectives (see Fig. 5).

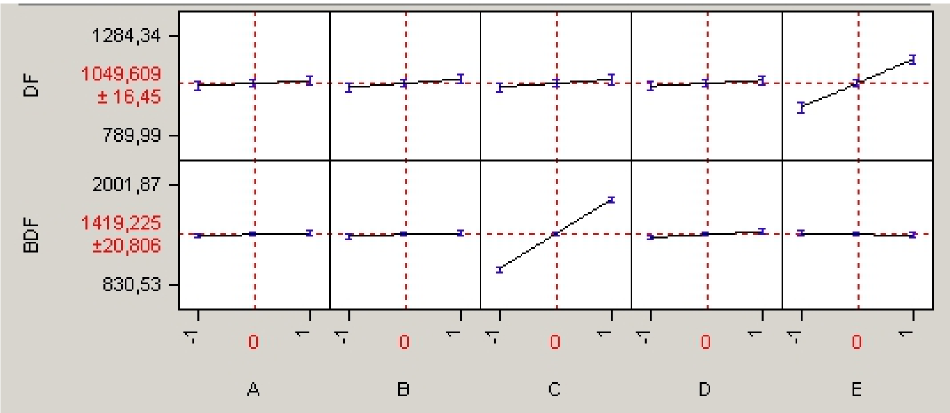


Fig. 5. Impact of the decision variables on study objectives

Additional information on optimization, as well as the goals of the study, is summarized in the following table:

Objectives	Goal	Lower limit	Upper limit	Weighting
BDF	To minimize	830.53	2001.87	3
DF	To maximize	789.99	1284.34	5

Table 2. Constraints on objectives of study

5.3 Multiobjective optimization

Let us optimize the following problem with our CACO multiobjective approach:

$$\begin{aligned} \text{Find} \quad & x = [A, B, C, D, E]^T \\ \text{which minimize} \quad & f(x) = \{-DF(x), BDF(x)\} \\ \text{subject to} \quad & DF(x) \le 1284.34 \\ & DF(x) \ge 789.99 \\ & BDF(x) \le 2001.87 \end{aligned} \quad (13)$$

The fitness function used, obtained by the CP (compromise programming) (Gagné et al., 2004) method, allows the search for solutions approaching the ideal point for each objective (**Step 4**):

$$fitness = \left(\frac{5}{8} * \left(\frac{DF_{\max} - DF_i}{DF_{\max} - DF_i^{Nad}} \right)^2 + \frac{3}{8} * \left(\frac{BDF_i - BDF_{\min}}{BDF_i^{Nad} - BDF_{\min}} \right)^2 \right)^{\frac{1}{2}} \tag{14}$$

The minimization of the fitness function enables us to reach our "*BDF*" minimization and "*DF*" maximization goals.

The result is a set of optimal Pareto solutions. We present more than one solution to the user in order to provide him with a margin of makeover. Abdul-Wahab & Abdo (Abdul-Wahab & Abdo, 2007), in their paper, present their 10 best solutions. We will do the same in order to make some comparisons.

Using the MatLab software (**step 5**), Figure 6 allows us to say that we get the best solution after 310 iterations. The staircase shape of the curve (Fig. 6) is explained by the memory effect that we used in the program code. Thus, when iteration produces a worse solution than the last one, this last solution (previous iteration) is retained.

The Table 3 presents the results obtained by the optimization process. The best solution is the number 1, while the 9 other solutions offer alternatives to user. These solutions meet problem constraints and, gives results which minimize "*BDF*" and maximize "*DF*" while remaining in the field of each decisional variable.

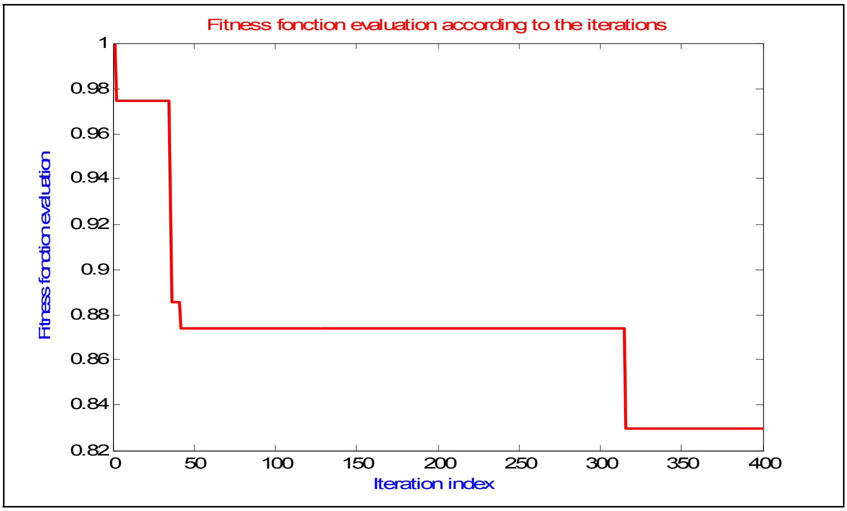


Fig. 6. Values of the fitness function according to the iteration count

# solutions	SWIT	TD	LSBL	FSBL	BRPF	DF	BDF	FITNESS
1	24.05	7.96	53.38	46.97	11499	1249.698	979.903	0.8295
2	24.26	7.99	66.23	44.787	11472	1245.603	990.740	0.8604
3	24.26	7.997	62.69	46.9	11492	1247.492	988.099	0.8738
4	24.37	7.94	54.20	48.49	11497	1245.577	981.335	0.8741
5	24.26	7.99	66.23	44.79	11472	1245.603	990.740	0.8854
6	24.61	7.99	53.99	43.67	11443	1245.552	985.796	0.8863
7	24.56	7.96	53.38	46.97	11499	1244.994	979.456	0.8904
8	24.33	7.94	50.86	71.3	11492	1240.785	987.596	0.8909
9	24.17	7.76	70.76	40.45	11439	1236.892	997.412	0.8963
10	24.72	7.97	60.13	40.12	11436	1239.208	982.990	0.8986

Table 3. Optimal solutions

The above Table (see Table 3) present the 10 best results of our study. These solutions meet the constraints of the problem and give excellent results which minimize "*BDF*" and maximize "*DF*" while remaining within the confines of each decision variable. It's interesting to observe the values of the decision variables in their respective fields. We can see that these best solutions are obtained under the following conditions (see Fig.7):

- Low temperature for seawater (SWIT) entering into the system
- The temperature difference (TD), which is high and similar for each of the solutions
- A final level of low salinity (LSBL)
- A first level of low salinity (FSBL)
- A high flow rate of the pump recycling salt (BRPF)

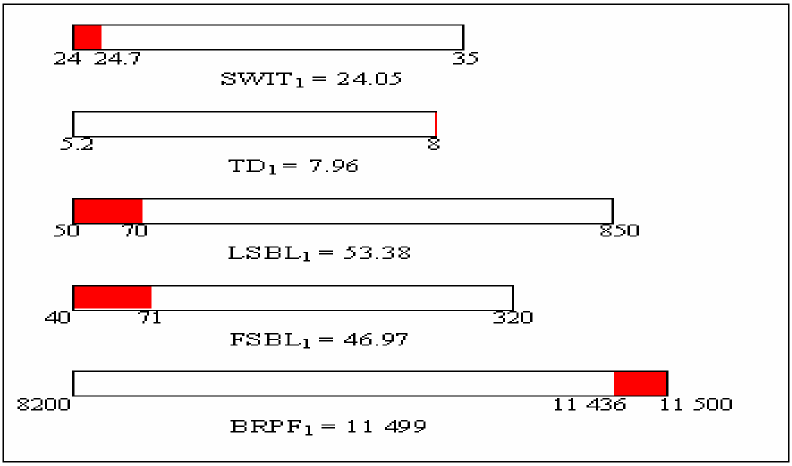


Fig. 7. Value margins of variables for optimal solutions

5.4 Comparison with authors' results

A comparison between the results obtained with the desirability function ("*DF*") and the hybrid approach developed ("*DF/CACO multiobjective*") shows that the second gives better quality results. Recall that this comparison is made between the results obtained by the proposed approach and those of Abdul-Wahab & Abdo.

# of solution	BDF desirability	BDF-CACO multiobjective	% BDF improvement	DF desirability	DF CACO multiobjective	% DF improvement
1	991.725	979.903	1.19%	1224.52	1249.698	2.01%
2	1038.83	990.740	4.63%	1222.91	1245.603	1.82%
3	1033.34	988.099	4.386%	1213.11	1247.492	2.76%
4	1035.78	981.335	5.26%	1210.98	1245.577	2.78%
5	975.718	990.740	-1.54%	1181.75	1245.603	5.13%
6	965.137	985.796	-2.14%	1173.29	1245.552	5.80%
7	996.446	979.456	1.71%	1176.02	1244.994	5.54%
8	1035.8	987.596	4.65%	1166.87	1240.785	5.96%
9	973.657	997.412	-2.44%	1151.36	1236.892	6.92%
10	1005.05	982.990	2.19%	1141.94	1239.208	7.85%

Table 4. Results of multiobjective CACO versus the desirability function

Firstly, by observing the change in the response values we obtain for the various solutions (see Table 4), we can see that the solutions achieved with the hybrid approach vary much less than those obtained with the desirability function of Abdul-Wahab & Abdo (Abdul-Wahab & Abdo, 2007). It seems that our solutions are closer to each other. The reason is that the hybrid approach causes small displacements during the ant's research process. Thus when the fitness function decreases, the ants move over a short distance before re-test the function, if and only if, the obtained value is better than the previous one. Otherwise, the process reorients itself in case of declining performance.

Secondly and always in Table 4, by comparing our results with those of Abdul-Wahab & Abdo (Abdul-Wahab & Abdo, 2007) for the desirability function, the CACO-multiobjective approach shows that the second objective gives better quality results, with a 4.66% average improvement for the main goal (BDF & CACO-multiobjective), and 1.79% for the secondary one (DF & CACO-multiobjective).

Moreover, the observed variations in the answers values of the various solutions are visualized on figures 8a and 8b. (see below). These variations from the point of view of the BDF desirability function are shown on Fig 8a while those related to DF function are illustrated on Fig 8b. By observing these Figures, we observe that the solutions obtained with the CACO-multiobjective approach are smaller than those of classical approaches. As previously stated, these variations are explained by a small displacement of local ants, and when the fitness function decreases, ants move on a short distance before re-test the fitness function to obtain a new solution. These mechanisms and process and mechanisms are the same for the second desirability function visualized on Fig. 8b.

Following this application, and having obtained appreciable results, we can conclude that our algorithm functions correctly, while leading to coherent solutions, and that it has proven its effectiveness by obtaining better solutions than those of the authors, Abdul-Wahab & Abdo (Abdul-Wahab & Abdo, 2007).

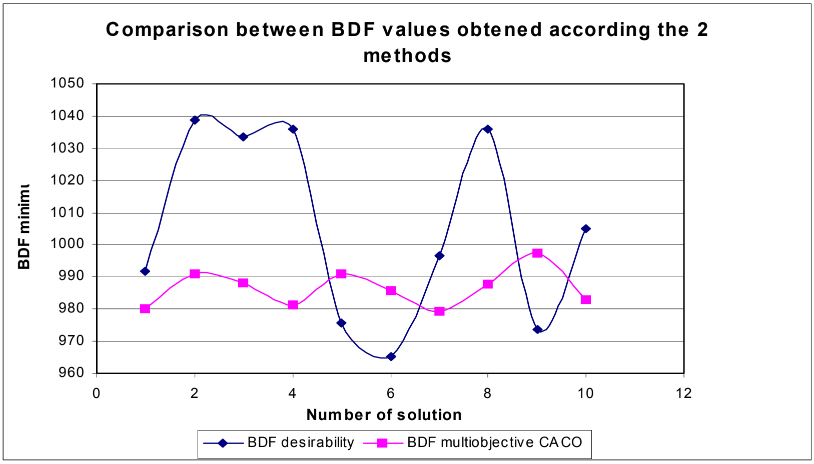


Fig. 8a. Chart of BDF values for optimal solutions (Part I)

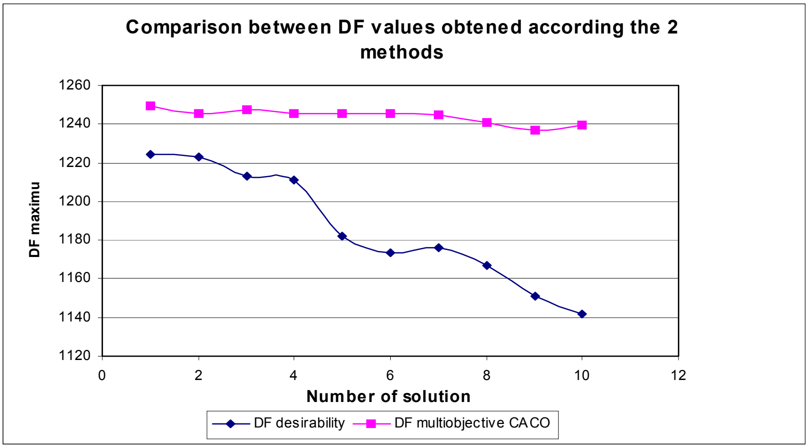


Fig. 8b. Chart of DF values for optimal solutions (Part II)

6. Conclusion

This book chapter presents a new multiobjective optimization approach for mechanical system design. Various techniques have traditionally been employed to resolve this kind of problem, including an approach combining RSM, GA and a simulation tool such as FEM. We have the ACO, which allows the exploration of a combination which includes another optimization algorithm. The ACO captured our interest because we were able to note in various works that in multiobjective optimization, it does produce better results than the quadratic programming technique and the GA. The ACO thus appears to be an innovative and leading solution for design optimization, because it is completely generalized and independent of problem type, which allows it to be modified in order to optimize the design of a complex mechanical system, subject to various economical and mechanical criteria, and respecting many constraints. However, it must be recalled that the ACO was developed to resolve discrete problems, and that its use on continuous problems is constantly under development; our study contributes to the development of the continuous ACO for multiobjective problems.

The approach we present makes it possible to effectively optimize a mechanical design problem. The approach performs much better when compared to using the desirability function. The results of the application allow it to validate the suggested design optimization method.

7. Acknowledgment

The author wishes to thank the co-authors of this chapter. Specifically, Miss Estinda-Mpiga for his major contribution and the quality of work. Thanks to all colleagues who have contributed from time to time, partially or more to obtain this manuscript.

Thank you to the Mechanical Engineering Department members at the "*École de Technologie Supérieure*" in Montreal and those of the Management Sciences Department at the University of Quebec in Abitibi-Témiscamingue (UQAT), Canada.

8. References

Abdul-Wahab, S.A. & Abdo, J. (2007). Optimization of multistage flash desalination process by using a two-level factorial design, *Applied Thermal Engineering*, Vol. 27, pp. 413–421

Bilchey, G. & Parmee, I.C. (1995). Ant colony metaphor for searching continuous design spaces, in: *Lecture Notes in Computer Science*, No. 993, pp. 25–39

Chegury Viana, F.A.; Iamin Kotinda, G.; Alves Rade, D. & Steffen, V. (2006). Can Ants Mechanical Engineering Systems?, *Proceeding of IEEE Congress on Evolutionary computation*, Vancouver, Canada

Chen, L.; Chen, J.; Qin, L. & Fan, J. (2004). A method for solving optimization problem in continuous space using improved ant colony algorithm, *Data Mining and Knowledge Management: Chinese Academy of Sciences*, Vol. 3327, pp. 61-70

Colorni, A.; Dorigo, M. & Maniezzo, V. (1992). Distributed optimization by ant colonies, *Proceedings in the First European Conference on Artificial Life, Elsevier Science Publisher*, pp. 134 -142

Deneubourg, J.L. & Goss, S. (1989). Collective patterns and decision-making. *Ethology, Ecology & Evolution*, No. 1, pp. 295-311

Deneubourg, J.L.; Pasteels, J.M. & Verhaeghe, J.C. (1983). Probabilistic behaviour in ants: a strategy of errors?, *Journal of Theoretical Biology*, No. 105, pp. 259-271

Dorigo, M. (1992). *Optimization, learning and natural algorithms*. PhD thesis, Dipartimento di Elettronica, Politecnico di Milano, Italia

Dorigo, M.; Bonabeau, E. & Theraulaz, G. (2000). Ant algorithms and stigmergy, *Proceeding in Future Generation Computer Systems*, Vol. 16, No. 8, pp. 851-871

Dorigo, M.; Di Caro, G. & Gambardella, L.M. (1999). Ant algorithms for discrete optimization, *Artificial Life*, Vol. 5, No. 2, pp. 137-172

Dorigo, M.; Maniezzo, V. & Colorni, A. (1996). Ant System: Optimization by a Colony of Cooperating Agents, *IEEE Transactions On Systems, Man., and Cybernetics – part B*, Vol. 26, No. 1, pp. 29-41

Gagné, C.; Gravel, M. & Price, W.L. (2002). Scheduling continuous casting of aluminium using a multiple objective ant colony optimization metaheuristic, *European journal of operational Research*, Vol. 143, pp. 218-229

Gagné, C.; Gravel, M. & Price, W.L. (2004). Optimisation multiobjectfs à l'aide d'un algorithme de colonie de fourmis, *INFOR*, Vol. 42, No.1, (February 2004), pp. 23–42

Goss, S.; Beckers, R.; Deneubourg, J.L.; Aron, S. & Pasteels, J.M. (1990). How trail laying and trail following can solve foraging problems for ant colonies, in: *Behavioural Mechanisms of Food Selection*, (Ed. R.N. Hughes), NATO ASI Series, G 20, Springer-Verlag, Berlin, pp. 661-678

Hamed, O.A.; Al-sofi, M.A.K.; Imam, M.; Mustafa, G.M.; Bamardouf, K. & Al-Washmi, H. (2001). Simulation of multistage flash desalination process, *Desalination*, Vol. 134, No. 1-3, pp. 195-203

Liang, Y.C. & Smith, A.E. (2004). An Ant Colony Optimization Algorithm for the Redundancy Allocation Problem (RAP), *IEEE transaction on reliability*, Vol. 53, No. 3, pp. 417-423

Ling, C.; Jie, S. & Ling, Q. (2002). A method for solving optimization problem in continuous space by using ant colony algorithm, *Journal of Software*, Vol. 13, No. 12, pp. 2317-2323

Mathur, M.; Karale Sachin, B.; Prive, S.; Jayaraman, V.K. & Kulkarni, B.D. (2000). Ant colony approach to continuous optimization, *Industrial and Engineering Chemistry Research*, Vol. 39, No. 10, pp. 3814–3822

Monmarché, N.; Venturini, G. & Slimane, M. (2000). On how Pachycondyla apicalis ants suggest a new search algorithm, *Future Generation Computer Systems*, Vol. 16, No. 8, pp. 937-946

Myers, R.H. & Montgomery, D.C. (2002). *Response surface methodology process and product optimization using designed experiments*, John Wiley & Sons, Inc., 2nd ed., New-York, USA

Nagesh Kumar, D. & Janga Raddy, M. (2006). Ant Colony Optimization for Multi-Purpose Reservoir Operation, *Water Resources Management*, Vol. 20, pp. 879–898

Pourtakdoust, S.H. & Nobahari, H. (2004). An extension of ant colony system to continuous optimization problems. *Proceedings in Ant Colony Optimization and Swarm Intelligence (ANTS) 4th International Workshop*, pp. 294–301

Roux, W.J.; Stander, N. & Haftka, R.T. (1998). Response surface approximations for structural optimization. *International Journal for Numerical Methods in Engineering*, Vol. 42, pp. 517–534

Socha, K. (2004). ACO for continuous and mixed-variable optimization. *Proceeding in Ant Colony Optimization and Swarm Intelligence (ANTS) 4th International Workshop*, in: *Lecture Notes in Computer Science*, Vol. 3172, pp. 25-36

Socha, K. & Dorigo, M. (2006). Ant colony optimization for continuous domains, *European journal of operational research*, Vol. 185, No. 3, pp. 1155-1173

Stander, N. (2001). The successive response surface method applied to sheet-metal forming, *Proceedings of the first MIT conference on computational fluid and solid mechanics*, pp. 481-485, June 12–15, 2001

Sun, H. & Lee, S. (2005). Response surface approach to aerodynamic optimization design of helicopter rotor blade, *International journal for numerical methods in engineering*, pp. 125–142

Zhang, T.; Rahman, S. & Choi, K.K. (2002). A hybrid method using response surface and pattern search for derivative-free design optimization, in: *Mang HA, Rammerstorfer FG, Eberhardsteiner J.*, (Ed.), *Proceedings of the fifth world congress on computational mechanics*, Vienna, Austria, July 7–12, 2002

Zhao, J.H.; Zhaoheng, L. & Dao, T.M. (2007). Reliability optimization using multiobjective ant colony system approaches, *Reliability Engineering and System Safety*, Vol. 92, No. 1, pp. 109-120

4

Reconfigurable Tooling by Using a Reconfigurable Material

Jorge Cortés, Ignacio Varela-Jiménez and Miguel Bueno-Vives
Tecnológico de Monterrey, Campus Monterrey
México

1. Introduction

Changes in manufacturing environment are characterized by aggressive competition on a global scale and rapid changes in process technology; these require creation of production systems easily upgradable by themselves and into which new technologies and new functions can be readily integrated (Mehrabi et al, 2000). In USA; industry, government and other institutions have identified materials and manufacturing trends for 2020 (Vision 2020 Chemical Industry of The Future, 2003 & National Research Council, 1998). The materials Technology Vision committee, in the publication of "Technology Vision 2020- The U.S. Chemical Industry" has identified a number of broad goals, which are enclosed in five main areas:

- New materials
- Materials characterization
- Materials modeling and prediction
- Additives
- Recycling

An important point in this vision is the development of smart materials, which have properties of self-repair, actuate and transduce. Polymers, metals, ceramics and fluids with these special characteristics belong to this class of materials and are already used in a great diversity of applications (Vision 2020 Chemical Industry of The Future, 2003).

In the other hand, the Visionary Manufacturing Challenges (National Research Council, 1998), published by the US National Academy of Sciences, presented six Grand Challenges:

- Integration of Human and Technical Resources
- Concurrent Manufacturing
- Innovative Processes
- Conversion of Information to Knowledge
- Environmental Compatibility
- Reconfigurable Enterprise

To reach these challenges, innovative processes to design and to manufacture new materials and components along with adaptable, integrated equipment, processes, and systems that can be readily reconfigured for a wide range of customer requirements or products, features, and services are needed (National Research Council, 1998).

The field of smart materials and structures is emerging rapidly with technological innovations appearing in engineering materials, sensors, actuators and image processing (Kallio et al, 2003). One of the smart materials is Nickel – Titanium alloy (NiTi) that possess an interesting property by which the metal 'remembers' its original size or shape and reverts to it at a characteristic transformation temperature (Srinivasan & McFarland, 2001).

Next manufacturing system generation requires of reconfigurable systems which go beyond the objective of mass, lean and flexible manufacturing systems. Because of the manufacturing trends towards a customer focused production. A reconfigurable manufacturing system is designed in order to rapid adjustment of production capacity and functionality, in response to new circumstances, by rearrangement or change of its components (Mehrabi et al, 2000). As can be seen in Fig. 1, there are many aspects of reconfiguration, such as, configuration of the product system, reconfiguration of the factory communication software, configuration of new machine controllers, building blocks and configuration of modular machines, modular processes, and modular tooling. So that, the development and implementation of key interrelated technologies to achieve the goals of reconfigurable manufacturing systems are needed.

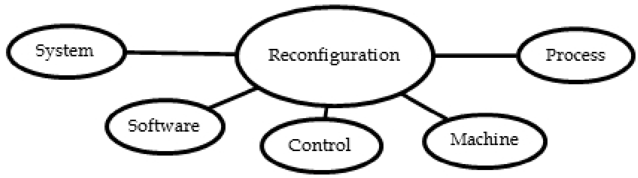


Fig. 1. Aspects of reconfiguration (Mehrabi et al, 2000).

Of relevant importance are the control, monitoring and sensing of reconfigurable manufacturing systems. By noting that the system configuration changes, the parameters of the production machines and some other physical parameters will change accordingly. The controller and process monitoring systems should have the ability to reconfigure and adapt themselves to these new conditions (Mehrabi et al, 2000).

1.1 Research justification

The use of NiTi requires proper characterization according to the environment surrounding the material when it is applied in some device; due to this requirement, a constitutive model is needed in order to relate the microstructure and thermo-mechanical behavior of the material.

In the manufacture industry, a variable shape die has always been an attractive idea to reduce design time and costs, since it allows as many designs to be rapidly manufactured at nearly free cost, using one tool for several shapes (Li et al, 2008).

The use of NiTi as an actuator in manufacturing systems is an opportunity area, allowing that several products can be formed by the same tool; this way, NiTi will help to evolve the traditional manufacturing industry.

1.2 Research aim

To develop a reconfigurable manufacture system for sheet metal/plastic forming controlled by NiTi actuators and to formulate a constitutive model of its thermo- mechanical behavior.

2. Constitutive model

NiTi is a smart material with properties such as shape memory effect (SME) and superelasticity (Chang & Wu, 2007). SME involves the recovery of residual inelastic deformation by raising the temperature of the material above a transition temperature, whereas in superelasticity, large amounts of deformation (up to 10%) can be recovered by removing the applied loads (Azadi et al, 2007).

The microscopic mechanisms involved in SME are strongly correlated to the transformation between the austenite parent phase at high temperatures and the martensite at low temperatures (Lahoz & Puértolas, 2004). It is a reversible, displacive, diffusionless, solid–solid phase transformation from a highly ordered austenite to a less ordered martensite structure (McNaney et al, 2003). Austenite has a body centered cubic lattice while martensite is monoclinic. When NiTi with martensitic structure is heated, it begins to change into the austenitic phase. This phenomenon starts at a temperature denoted by A_s, and is complete at a temperature denoted by A_f. When austenitic NiTi is cooled, it begins to return to its martensitic structure at a temperature denoted by M_s, and the process is complete at a temperature denoted by M_f (Nemat-Nasser et al, 2006). Because austenite is usually higher in strength than martensite, a large amount of useful work accompanies the shape change. Austenite exhibits higher stiffness than martensite (De Castro et al, 2007).

When NiTi is stressed at a temperature close to A_f, it can display superelastic behavior. This stems from the stress-induced martensite formation, since stress can produce the martensitic phase at a temperature higher than M_s, where macroscopic deformation is accommodated by the formation of martensite. When the applied stress is released, the martensitic phase transforms back into the austenitic phase and the specimen returns back to its original shape (Nemat-Nasser et al, 2006). The stress-induced austenite-martensite transformation is effected by the formation of martensitic structures which correspond to system energy minimizers (McNaney et al, 2003) as result of the need of the crystal lattice structure to accommodate to the minimum energy state for a given temperature (Ryhänen, 1999).

Shaw explains in more detail martensite behavior, affirming that due to its low degree of symmetry, the martensite exists either as a randomly twinned structure (low temperature, low stress state) or a stress-induced detwinned structure that can accommodate relatively large, reversible strains. Fig. 2 shows the thermomechanical response of a wire specimen. The specimen is first subjected to a load/unload cycle at low temperature, leaving an apparent permanent strain. The material starts in a twinned martensite (TM) state and becomes detwinned (DM) upon loading. The specimen is then subjected to a temperature increase while holding the load. The SME is seen as the strain is recovered and the material transforms to austenite (A). The temperature is then held at high value and the specimen is again subjected to a load/unload cycle. In this case the material shows superelasticity and transforms from austenite to detwinned martensite during loading and then back to austenite during unloading (Shaw, 2002).

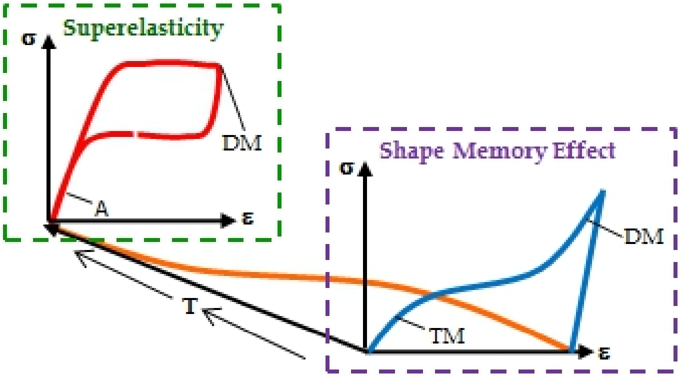


Fig. 2. Thermomechanical cycle of NiTi (Shaw, 2002)

It is considered that composition (Nemat-Nasser et al, 2006) and heat treatments have effect on the temperature at which material exhibits SME, called transformation temperatures (TTR) which are the prerequisite for the material to exhibit the SME and are one of the key parameters for SME based actuation, they also define the proper application for a certain NiTi composition alloy (Malukhin & Ehmann, 2006).

Establishment of a constitutive equation for phase transformation in NiTi requires considering the Stress-Strain-Temperature behavior and the phase transformations shown in Fig. 2, from which is observed that volume fraction of each microstructure depends of the strain and temperature conditions; it also has influence on the mechanical behavior of the material.

The possible phase transformations that can occur on NiTi are shown in Fig. 3, strain induces detwinned martensite while temperature increase induces austenite. As shown in Fig. 2, at low temperature and low stress, transformation of twinned martensite into detwinned martensite is started, and continues its plastic strain and then stress is released. When temperature is increased detwinned martensite transforms into austenite. If high temperature is kept and strain is applied superelasticity occurs and transformation of austenite into detwinned martensite is started, it finishes when stress is released and then microstructure transforms back into austenite.

According to Fig. 3, phase transformations on NiTi are:

1. Twinned martensite to Detwinned martensite (Strain induced)
2. Detwinned martensite to Austenite (Temperature increase induced)
3. Austenite to Detwinned Martensite (Strain induced)
4. Twinned martensite to Austenite (Temperature increase induced)
5. Austenite to Twinned Martensite (Temperature decrease induced)

Up to day, several studies have been made about NiTi, but there is a lack in the development of numerical analysis of the phenomenology of the material since its application requires a proper characterization; a constitutive model is needed in order to relate microstructure and thermo-mechanical behavior of NiTi. A similar model was developed by Cortes (Cortes et al, 1992) for determining the flow stress of aggregates with phase transformation induced by strain, stress or temperature and demonstrated its use for stainless steels. This model has also been applied on shape memory polymers (Varela et al, 2010) and it has been extended for modeling the displacement on electroactive polymers (Guzman et al, 2009) through a phase transformation approach, induced by some external stimulus. The model is based on an energy criterion which defines the energy consumed to deform the phases in the system as being equivalent to energy consumed to deform the aggregate and it is able to predict the flow stress behavior of the material. In order to apply Cortes' constitutive model on SMAs, experiments have to be carried out; the austenite, twinned martensite and detwinned martensite are considered as the aggregates and the microstructural transition between them becomes the basis of the constitutive model; this way, the constitutive expression will result in terms of the mechanical properties of each phase and its volume fraction.

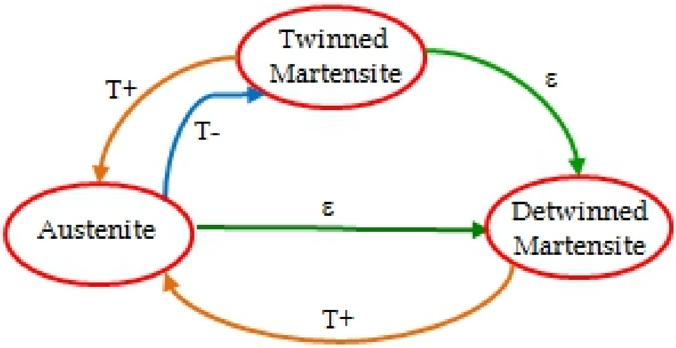


Fig. 3. Phase Transformations on NiTi and their induction stimulus.

2.1 Constitutive model of flow stress

In the case of the present aggregate composed of austenite and martensite, based on Cortes model (Cortes et al, 1992), V_f of each structure or aggregate are defined as:

$$V_{fa=}\frac{V_a}{V_t} \qquad V_{ftm=}\frac{V_{tm}}{V_t} \quad V_{fdm=}\frac{V_{dm}}{V_t} \tag{1}$$

where subscripts *a, tm and dm* indicate austenite, twinned martensite and detwinned martensite, respectively . Cortes constitutive model of flow stress for multi phases aggregate (Cortes et al, 1992) applied on NiTi is

$$\sigma_{NiTi} = V_{fa} \cdot \sigma_a + V_{ftm} \cdot \sigma_{tm} + V_{fdm} \cdot \sigma_{dm} \tag{2}$$

where σ_{NiTi} is stress of NiTi and σ_a, σ_{tm} and σ_{dm} are stress of each structure.

2.2 Kinetics of strain/temperature induced twinned martensite-detwinned martensite-austenite phase transformation

Based on the thermomechanical behavior of Fig. 2 and the phase transformations shown in Fig. 3, volume fraction of the microstructures varies by:

$$V_{fa} + V_{ftm} + V_{fdm} = 1 \tag{3}$$

$$V_{fa} = \left(1 - V_{fa-dm}\right) \cdot V_{fa_0} + V_{ftm-a} \cdot V_{ftm_0} + V_{fdm-a} \cdot V_{fdm_0} \tag{4}$$

$$V_{fdm} = \left(1 - V_{fdm-a}\right) \cdot V_{fdm_0} + V_{ftm-dm} \cdot V_{ftm_0} + V_{fa-dm} \cdot V_{fa_0} \tag{5}$$

where subscripts 0 indicate the initial valume of each volume fraction.

For strain induced detwinned martensite phase transformation:

$$V_{fdm} = \left[1 + \left(\frac{\varepsilon}{\varepsilon_c}\right)^{-B}\right]^{-1} \tag{6}$$

For temperature induced austenite phase transformation:

$$V_{fa} = \left[1 + \left(\frac{T}{T_c}\right)^{-B}\right]^{-1} \tag{7}$$

where B is a fitting constant; while ε_c and T_c represent the values of strain and temperature, respectively at which 50% of the phase transformation is occurred. Experimental work is required for determining these values for each phase transformation. Substituting (6) and (7) in (4) and (5):

$$V_{f_a} = \left\{1 - \left[1 + \left(\frac{\varepsilon}{\varepsilon_{C3}}\right)^{-B_3}\right]^{-1}\right\} \cdot V_{f_{a0}} + \left[1 + \left(\frac{T}{T_{C4}}\right)^{-B_4}\right]^{-1} \cdot V_{f_{tm0}} + \left[1 + \left(\frac{T}{T_{C2}}\right)^{-B_2}\right]^{-1} \cdot V_{fdm_0} \tag{8}$$

$$V_{f_{dm}} = \left\{1 - \left[1 + \left(\frac{T}{T_{C2}}\right)^{-B_2}\right]^{-1}\right\} \cdot V_{f_{tm0}} + \left[1 + \left(\frac{\varepsilon}{\varepsilon_{C1}}\right)^{-B_1}\right]^{-1} \cdot V_{f_{tm0}} + \left[1 + \left(\frac{\varepsilon}{\varepsilon_{C3}}\right)^{-B_3}\right]^{-1} \cdot V_{fa_0} \tag{9}$$

where subscripts 1, 2, 3 and 4 represent the B, T_c or ε_c value corresponding to that phase transformation.

2.3 Stress of microstructures

Since NiTi contains a heterogeneous microstructure under given conditions, an incremental change test to determine hardening parameters in a given structure has to be carried out. Based in Cortes work (Cortes et al, 1992) flow stress of austenite, twinned martensite and detwinned martensite is determined under isothermal conditions, by pre-straining NiTi wires at a temperature at which only one microstructure exists, and then the specimens were individually deformed at a predefined temperature. The yielding point in reloading is registered as the flow stress at that temperature and those strain conditions. From this experiment equations for estimating σ_a, σ_{tm} and σ_{dm} are determined. These should be of the form:

$$\sigma_a = K_a \cdot \varepsilon^{N_a} \tag{10}$$

$$\sigma_{tm} = K_{tm} \cdot \varepsilon^{N_{tm}} \tag{11}$$

$$\sigma_{dm} = K_{dm} \cdot \varepsilon^{N_{dm}} \tag{12}$$

where *K* and *N* represent material constants which are determined experimentally.

By substituting equations (3) and (8)-(12) into equation (1) stress on NiTi can be described relating thermomechanical behavior with microstructure.

3. Reconfigurable die

Conventional type of mold fabrication involves time and money investment to achieve design of a die; the concept about forming a die of variable shape has always been attractive as a means of rapid iterations and almost cost free (Li, 2008). Thus, multi forming methods have been developed in order to achieve reconfigurability of the process.

3.1 Multi point forming (MPF)

This method has been used to replace solid dies with three-dimensional surfaces. The main key of the MPF is the two matrices of punches allowing that create a three-dimensional surface which forms according to the shape of the design; this way the surface can be approximated to a continuous die, as shown in Fig. 4 (Zhong-Yi, 2002).

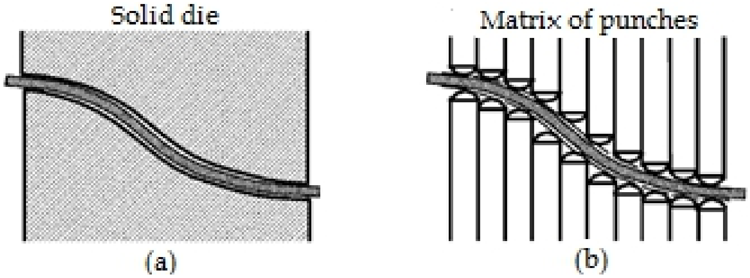


Fig. 4. (a) Conventional die forming; (b) Multi-point forming (Zhong, 2002).

MPF is based on controlling the elements and punches, hence, a matrix on punches can be shaped as required (Li, 2002). Fig. 5. illustrates the parameters related to the reconfigurable process: design and manufacture of the pin heads; since its shape, size and length play an important role in the arrangement of the closed matrix (Walczyk, 1998).

Design of a tool based on multi-point technique has many considerations since it involves several variables and many issues and problems use to occur such as dimpling, buckling and non linear deformation of the material; due to this issues four main designs have been researched with different punches types (Li, 2002).

- Multi-point full die
- Multi-point half die
- Multi-point full press
- Multi-point half press

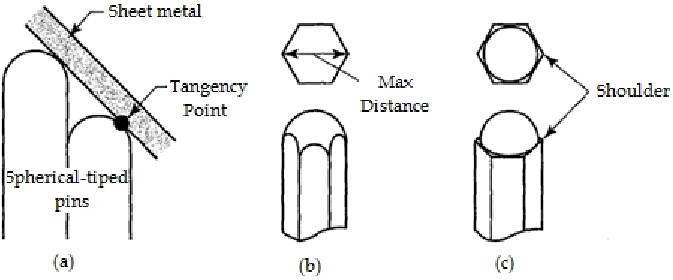


Fig. 5. Spherical and Hexagonal pin head designs (Li, 2002).

The arrangement of each design is described and shown in Table 1 and Fig. 6, respectively.

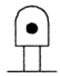
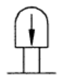
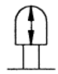
Type of Punch	Adjustment	Required Force	Drawing / Mark
Fixed	Before Forming	Small	
Passive	While Forming	None	
Active	Free Movement	Large	

Table 1. Different types of MPF.

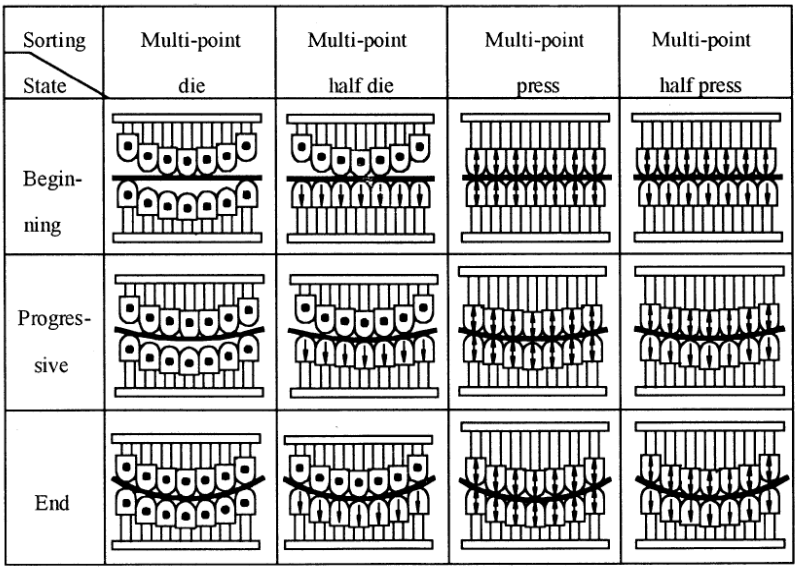


Fig. 6. Different types of Multi-point forming and the interaction with the process (Walczyk, 1998).

3.1.1 Multi-point sandwich forming (MPSF)

MPSF is an accessible method to manufacture components in small batches. Fig. 7 represents the MPSF method which uses an interpolator material to assure the surface quality of the metal sheet, this last also depends of the tool elements and the position between the pins (Zhang, 2006).

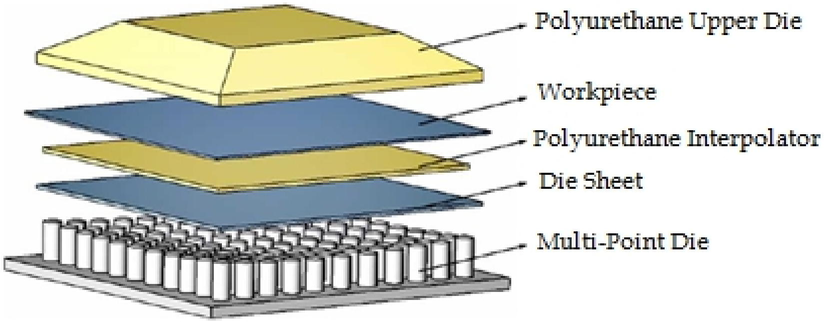


Fig. 7. Schematic Components for MPSF (Zhang, 2006).

3.1.2 Digitized die forming (DDF)

With DDF, forming procedures and integration between parameters such as deformation path, sectional forming; punches and control loop are being developed in order to avoid forming defects. The process is shown in Fig. 8 (Li, 2007).

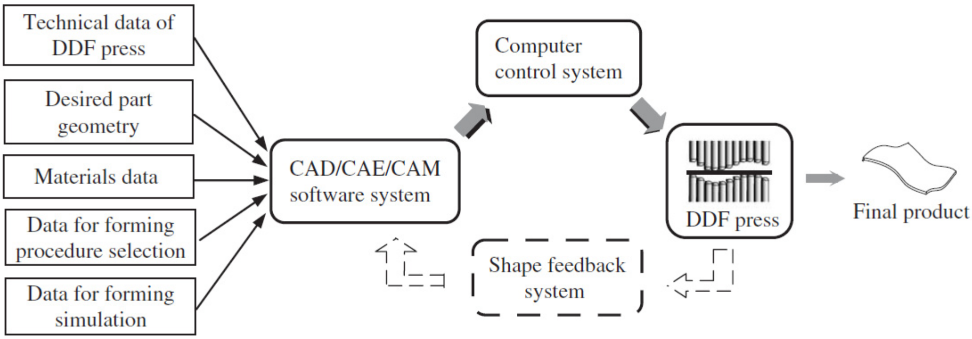


Fig. 8. Schematic of DDF integration system (Li, 2007).

3.2 Active multi point forming parameters

3.2.1 Pin head

A pin head must be strong enough to support the mechanical and thermal loads of the material to manufacture avoiding the problems in the final piece such as bending, buckling and dimpling (Walczyk, 1998). In addition to normal or vertical load, the pin will be subjected to lateral load depending the height and of the adjacent pins on the shape; this variable is a key factor on the pin head design (Schwarz, 2002), as shown in Fig. 9.

Uniformity in the pin heads and elements make the design easy to fabricate and assembly in the arrangement of the matrix, a comparison of different geometries evaluating the cross-sectional area shape is shown in Table 2 (Schwarz, 2002).

Pin Cross-Section Shape	Equilateral Triangle	Square	Hexagonal	Circle
Number of Sides	3	4	6	6
No. Isolated straight load paths	0	2	3	0

Table 2. Comparison of Cross-Sectional Geometry.

Also the structure and the size of the pins affect the quality of the piece and it is recommended the use of square shape elements, as shown in Fig. 10, in dense packed arrangements of matrices (Schwarz, 2002).

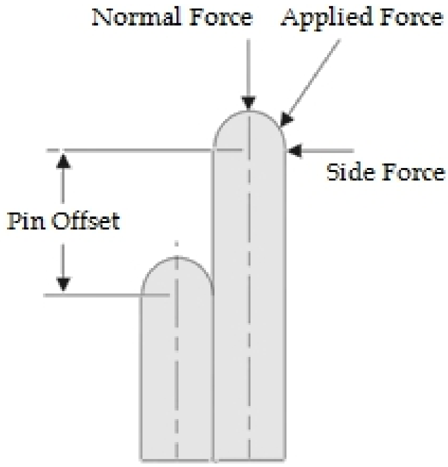


Fig. 9. Pin head forces interaction and offset.

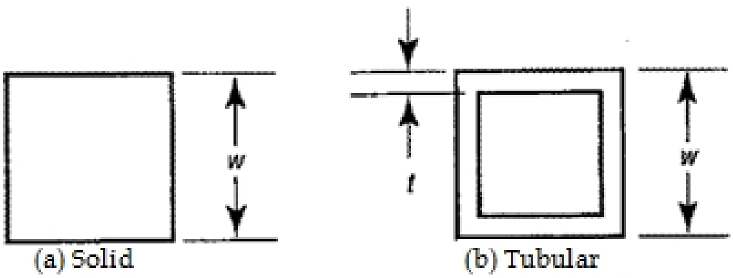


Fig. 10. Square design a) Solid and b) Tubular (Schwarz, 2002).

The use use of tubular elements is considered if the weight of the die has to be reduced, however it is not always is the best approach considering the scale and size of pins and the forces of the process (Schwarz, 2002).

3.2.2 Actuators

Each design of a shape has different means of independently moving pins in a large matrix arrangement (Walczyk, 2000). Automation of pin setting was not realized until in 1969 Nakajima positioned a matrix of pins controlled by a vibration mechanism mounted to a three-axis servomechanism (Nakajima, 1986). It is shown in Fig. 11.

Researches from different groups such as, the Massachusetts Institute of Technology (MIT) developed a Sequential Set-up Concept, the Rensselaer Polytechnic Institute build up a Hydraulic Actuation Concept and the Northrop Grumman Group Corporation created a Shaft-driven Lead screw Concept (Walczyk, 2000)

Sequential Set-up (SSU) by MIT

Each hollow pin has a threaded nut passed from its base. The pin moves up and down as the lead screw rotates, and the next pin prevent the rotation movement; the design eliminates the matrix external clamping force to position the pin heads.

Hydraulically Actuation (HA) by Renselaer Polytechnic

Individual elements are essentially hydraulic actuators, controlled by an in-line servo valve. The hydraulic pressure makes the element rise from the initial position maintaining the height until the pressure is released.

Shaft Driven Lead screw (SDL) by Northrop Group

This method depending on the need of the process needs a single or dual electric motor (one each on opposing sides of the die) mounted externally to drive worms mounted on cross shafts; the worm gear is connected to each pin´s lead screw.

Table 3 shows a comparison between these designs.

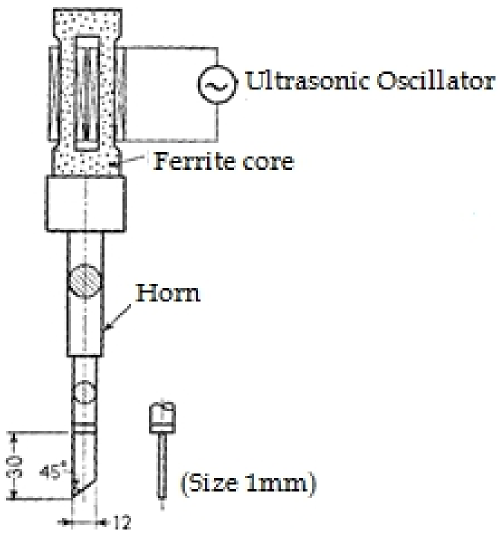


Fig. 11. Nakajima servomechanism (Nakajima, 1986).

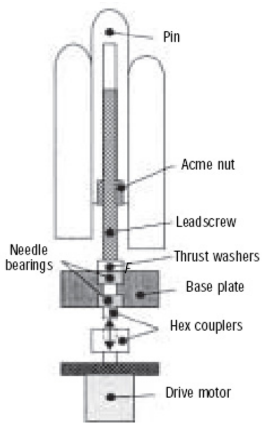
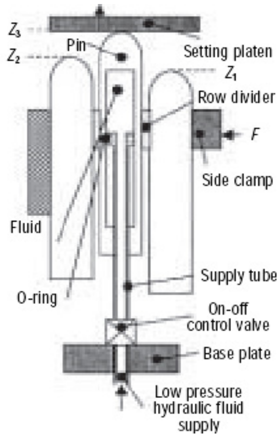
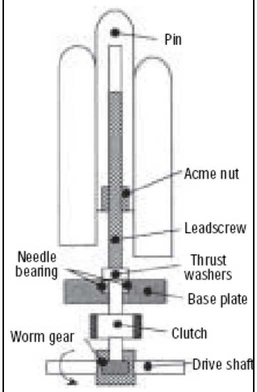
Characteristic	SSU	HA	SDL
Matrix of Pins	42x64 (28.6mm pin)	48x72 (25.4mm pin)	42x64 (28.6mm pin)
Number of Actuators	19 (16 drive motors, X, Y and Z axes)	1 (Hydraulic pump)	42 (drive motor per row)
Number of position control devices	0	3456 (servo valve per pin)	2688 (clutch per pin)
Number of sensors	19	1	42 (encoders per motor)
Potential mayor of positioning error	Backlash in lead screw	Insufficient platen stiffness	Rotational compliance
Setting mode	Serial	Parallel	Parallel
Potential control mode error	Lead screw is continuously engaged	Moving pins are in contact with platen	Clutch does not slip
Concept Design			

Table 3. Comparison of Actuation schemes for Dies (Walczyk, 2000).

As shown in table 3; the use of common actuators have problems to form a continuous even surface, due to the size of the actuators the element diameter has to be at least 25.4mm, also the actuators and the relationship in a potential control mode error by mechanical characteristics such as fatigue and cycles limit the use of the die in order to make different shapes (Walczyk, 2000).

3.2.3 Matrix arrangement

Design and manufacturability of the pins impact the position and design of the matrix due to the number of sides on the cross-sectional shape geometry. Fig. 12 shows the key factor of densely packed pin heads in a matrix; allowing management of the load and maintaining a smoother surface when subjected to loads. The contact elements have to be maximized while the gaps between elements need to be minimized (Schwarz, 2002).

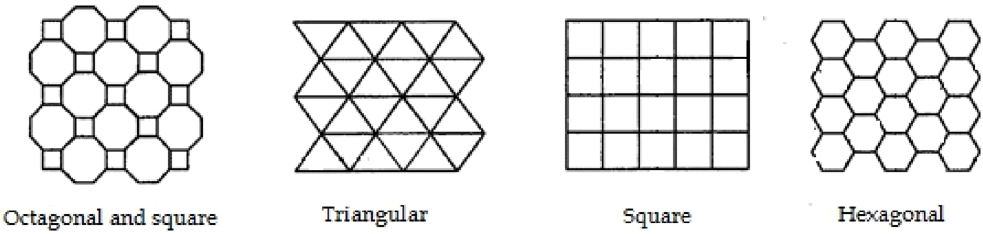


Fig. 12. Various Cross-Sectional shapes for die pins (Schwarz, 2002).

Shown in Fig. 12 is a discrete digitalized arrangement to a continuous surface with multi point forming die technology. It is important to have as many pins as possible since a poor transition surface may result in delicate wrinkling defect and possible cracking of the work piece (Peng, 2006).

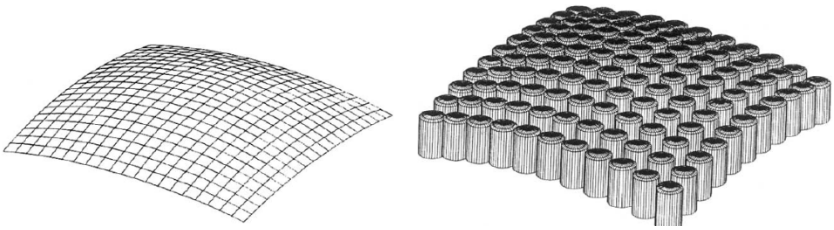


Fig. 13. Discrete approximation to a continuous surface square position (Rao, 2002).

3.2.4 Control and software

The desired part to be manufactured generating a controlling data in height of the piece is sent to a control system to perform the DDF (Li, 2007). Fig. 15. shows the different technologies merging to make the software design and control available to a reconfigurable design. An open loop control system and electronic on a single pin element, are used to evaluate a simpler circuit and timed software to excite the actuator (Walczyk & Hardt, 1998), as shown in Fig. 14.

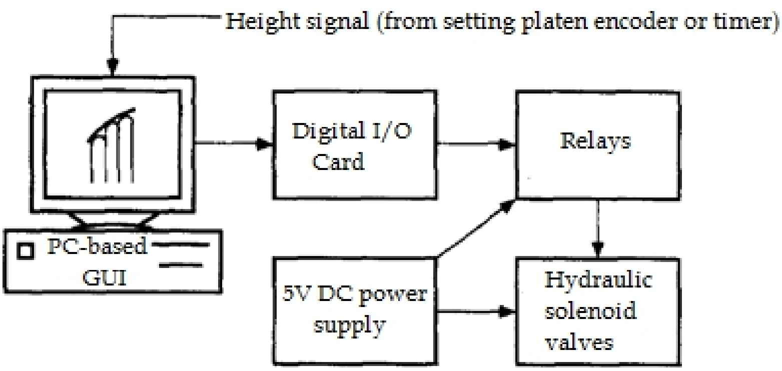


Fig. 14. Test Schematic of Die Control System (Walczyk & Hardt, 1998).

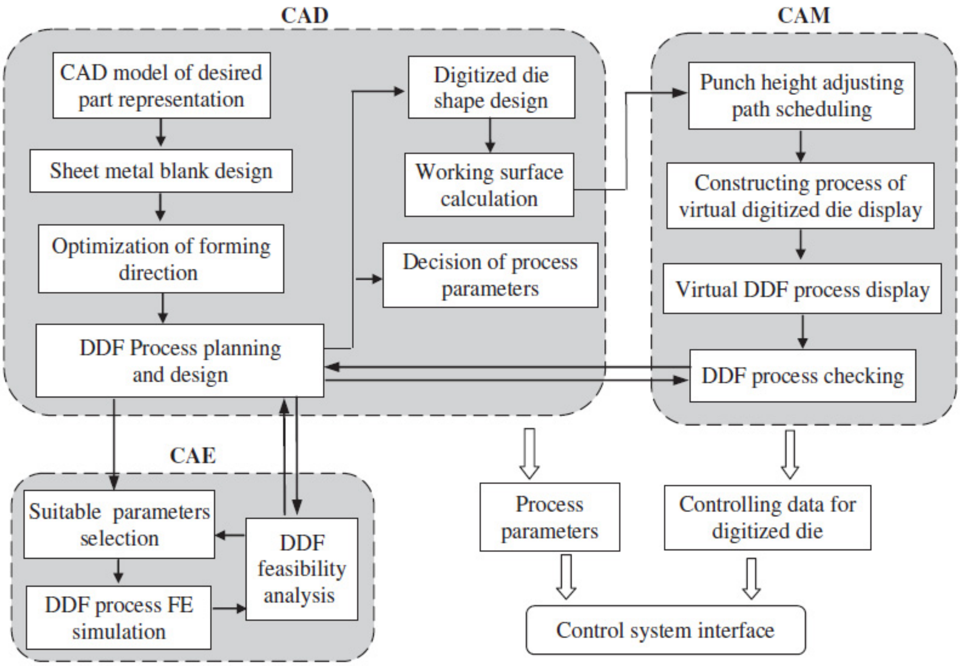


Fig. 15. CAD/CAE/CAM software control for DFF (Li, 2007).

3.3 Shape memory alloy actuators

The uses of SMA as actuators nowadays are not only medical devices; there are also novelty actuators such as (Humbeeck, 1999):

- Fashion and gadgets: single products were created from cell phone antennas, eye glasses frames and in the clothing industries frames for brassieres and wedding dresses pericoats (Duering, 1990).
- Couples: Heat-recoverable couplings of an F-14 hydraulic turbine were the first large scale produce actuator (Duering, 1990).
- Micro-actuators: The central Research Institute of Electric Power Industry in Japan built a piston-driver from a 2mm diameter wire based on 26 bars, having a life cycle over 500 000 cycles minimum.
- Adaptive materials: A vibrator frequency control of a polymer beam has been used to increase the natural frequency of the composite beam.
- Other applications: Wear cavitations defects where hydraulic machinery are used like water turbines, ship propellers and sluice channels (Jardine et al, 1994).

3.4 Development of reconfigurable die based on NiTi actuator

An active multi point forming tool, based on NiTi wires as main actuators is developed. The devices is known as ‘reconfigurable die’.

3.4.1 Hypothesis

The main issue on the development of multi point surfaces is that a high density matrix of pins is required, as smaller are the pins smoother is the surface that can be formed. This issue can be solved by using a small actuator that allows formation of a dense pins matrix. SME of NiTi can be applied to achieve the movement of a mechanism that controlls the movement of each pin

3.4.2 Methodology

Development of a reconfigurable die follows the concept of DDF. Hence, it is required to design a mechanism, a controller and a graphical user interface (GUI), the full system is shown in Fig. 16.

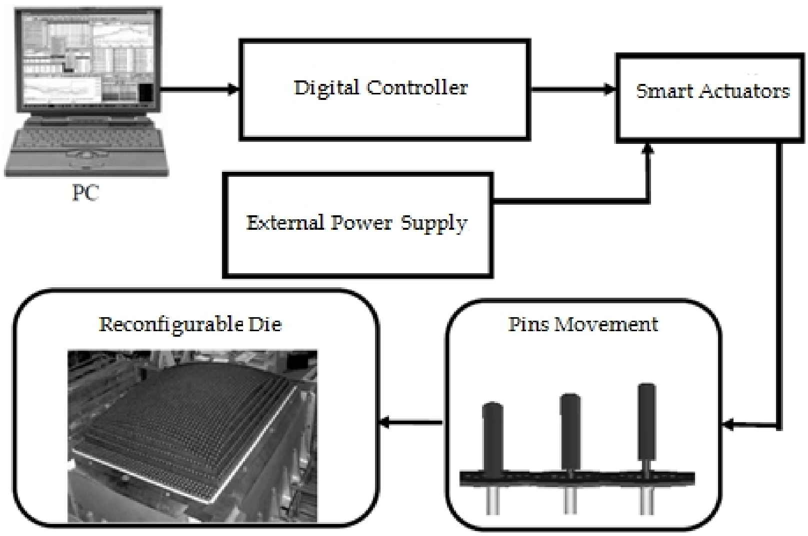


Fig. 16. Overall Process Variables.

3.4.3 Mechanism design

A design of a reconfigurable die proposed has been reviewed in order to identify its components and characterize them, sucha as the length of the shafts, springs parameters and SMA properties (length, diameter, electric current).

3.4.4 Functional prototype description

Fig 18 shows the mechanism that controls the vertical movement of each square pin. Each pin has a SMA wire subjected to a spring that deforms it, thus, NiTi has a martensitic structure. When a electric pulse activates the electric current the wire will be heated reaching the austenitic structure returning the wire to the non deformed position pushing the springs and the shaft, that causes that the pin rotates and move up; when the pulse is inactive the springs will deforms the SMA again and the cylce is restarted.

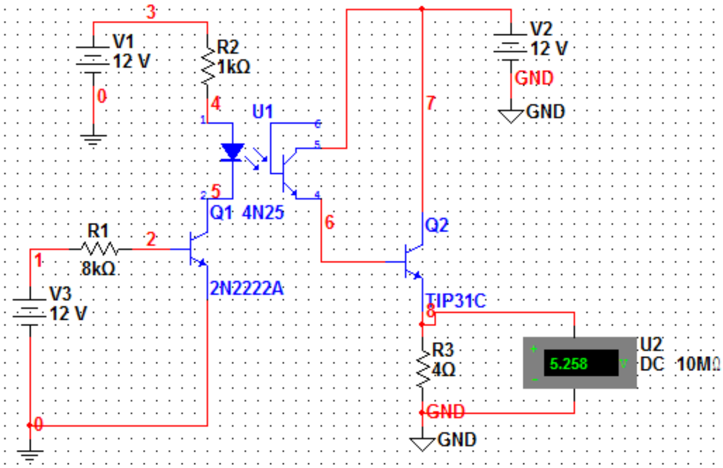


Fig. 17. Electronic design. Power and control circuits, R3 represents NiTi wire.

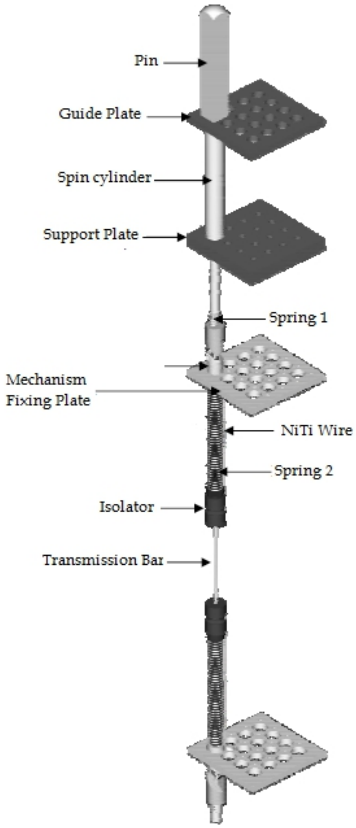


Fig. 18. General Design of the reconfigurable die.

3.4.5 Circuitry

The design consist in a basic control and power electronic circuitry which supplies the current needed to the SMA. Circuit is illustrated in Fig. 17.

3.4.6 Software

It was programmed with *Labview* by *National Instrument* and consists of three set of parts. The first one is the image codification from a solid figure to a virtual 3D figure. The second part is the pin head elements array and virtual configuration in the software; finally the last part is the digital output of electric pulses needed per pin head to change its height.

Image codification is performed by pictures taken from the object of interest, in order to make a virtual solid image in a 40 by 40 matrix. The number of views of the object depends of the geometry, a maximum of three pictures can be uploaded (sideway, front, and top views) in order to arrange the pins matrix.

The file from the generated matrix is then placed in the next file, resulting of a specific height for each element; this is then saved in a file as an array in order to visualize the pin design and making a blank. Each element has a resolution of a 28.5 pixel per pin. Fig 19 shows the Human Machine Interface (HMI) software and Fig. 20 shows the matrix of the pins height loaded on the software.

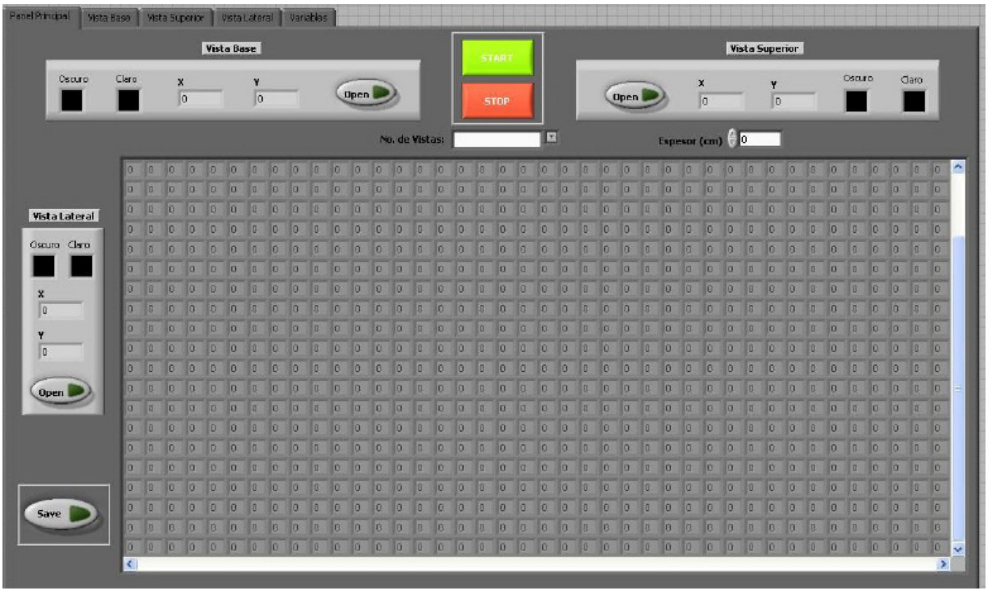


Fig. 19. Software HMI.

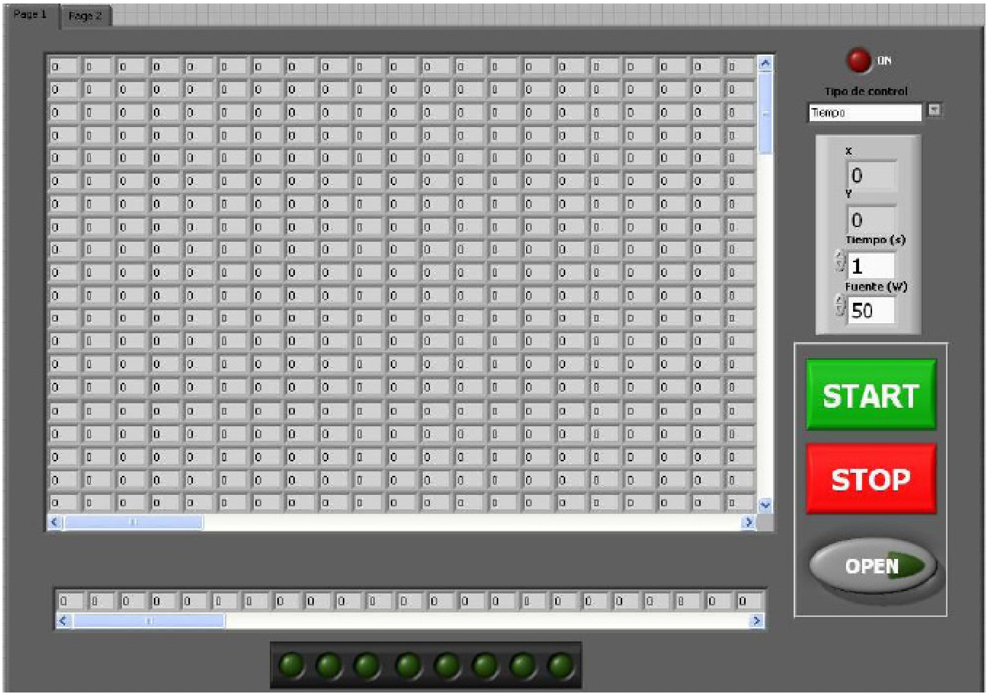


Fig. 20. Matrix pin height.

3.4.7 Control

Digital pulses: A digital pulse, as shown in Fig. 21, consists in a square wave of direct current output, on which the duty cycle is fixed, resulting a 50% high and 50% low. In this case the "high time" the actuator will be activated with current and the "low time" the signal will be deactivated in order to cool down the actuator.

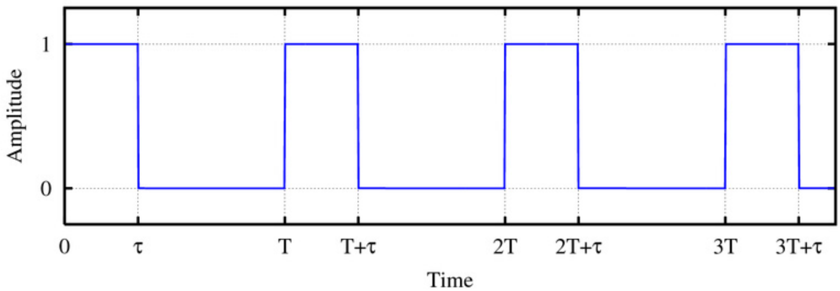


Fig. 21. Digital pulse.

The Duty Cycle (DC) represents the pulse duration divided by the pulse period, where τ is the duration that the function is actve hight (normally when the voltage is greater then zero) and T is the period of the function.

Pulse-width Modulation (PWM): is one of the most efficient ways to provide electrical power between the ranges fully on and fully off. It is a great electric tool to supply voltage/current in ht power electronics field to devices such as electric stoves, robot sensors, dimmers. The main characteristics are the variation and switching between the high and low levels in high frequencies ranges.

PWM, as shown in Fig. 22, uses a rectangular pulse wave whose pulse width is modulated resulting in the variation of the average value of the waveform. The wave has a changing duty cycle, making many pulses in a period of desired time.

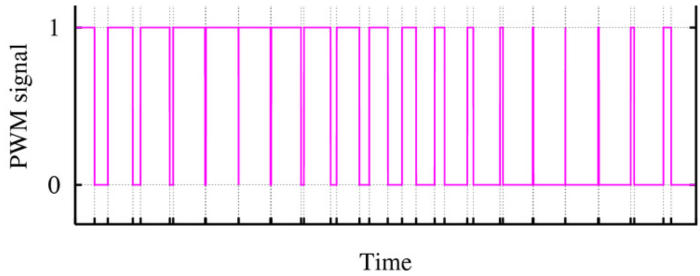


Fig. 22. PWM signal.

In order to select the most appropriate output design in the system, a Design of Experiments (DOE) it is implemented, it is shown in Table 4.

DOE

Factor	High level	Low level
Time of pulse	4 second	2 second
Wave output	Single Pulse	PWM Pulse
Power Type	5.5 Watts	13 Watts

Table 4. DOE for evaluating the electric pulse output of each pulse.

For testing purposes, the PWM has a high frequency of 10 KHz, for the single pulse the duty cycle is a 50%. The time of pulse is the total amount of the pulse.

According to the tests, the results indicates that the factor of type of power and time of pulse have an impact and the best results on the actuator to return to the original position and having a bigger recovery force is the single pulse in a lower time with a medium range power supply.

The control design to activate the movement of the elements, there are two different methods that can be considered, as described in Table 5.

Method	Definition	Elapsed time
Serial	Elevates the height of each pin from each row is elevated one at a time.	Longer
Parallel	Elevates the height of all pins from each row is elevated.	Shorter

Table 5. Pin actuation scheme

The difference in time setting can be calculated simply by adding the quantity of total cycles from all the pins in the matrix and multiplying by the total time per one pulse in seconds.

3.4.8 Final prototype

The final alpha prototype is shown in Figs 23 and 24.

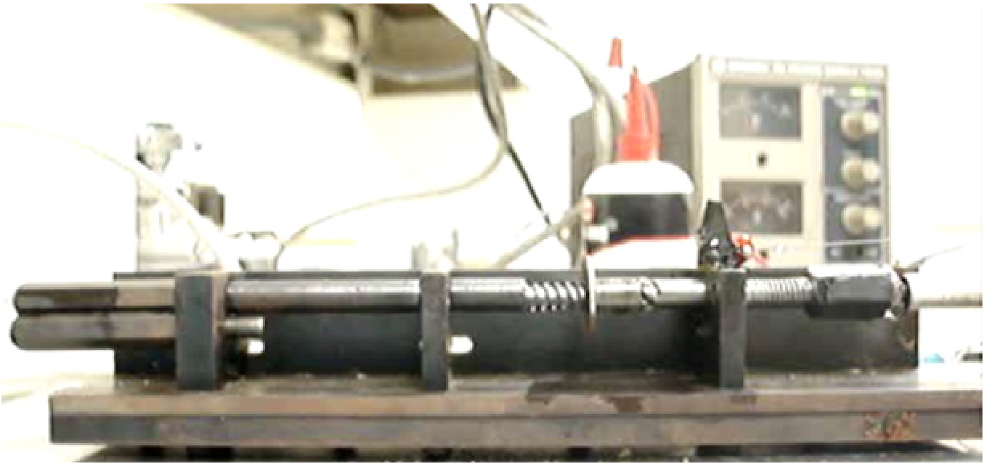


Fig. 23. Machined prototype.

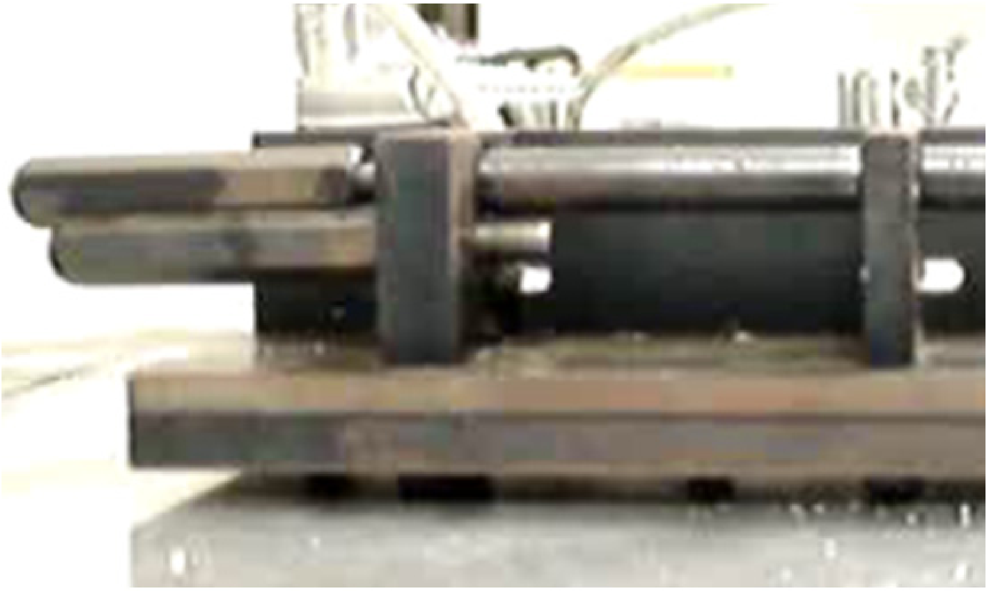


Fig. 24. Pin element changes height.

3.4.9 Operation process

The manufacture process and the use of this tool, allows a technological advantage in the design of a multi-point shape with a reconfigurable die. Its use involves CAD, CAM and CAE technologies as shown in Fig. 25.

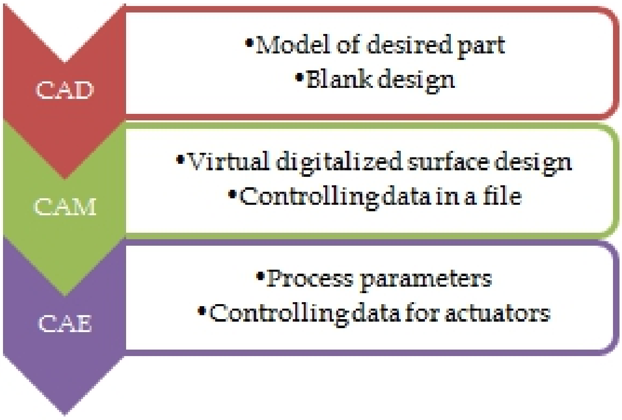


Fig. 25. Technologies used.

Image: the image depends on the views taken by the camera and the position of the object, the files are imported to the matrix arrangement software file to create a virtual object.

Virtual Matrix: when selecting the geometry of the matrix, the columns and rows are selected to visualize the object. The object then will be saved in a matrix file in the hard drive, containing the position of the elements needed from all the matrix, this allow the user to save as many designs and objects without repeating the image step every time.

Physical Matrix: Importing the matrix file, then the software will calculate the number of total cycles per element needed to reach the certain height. In this file the type of control (Parallel or Serial) is defined with the Data Adquisition Card (DAQ) digital outputs.

These proceeses are resumed in Fig. 26.

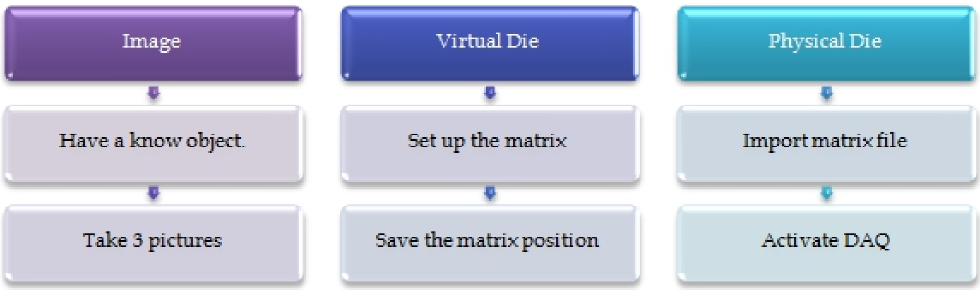


Fig. 26. Processes description.

4. Conclusions

The proposed cconstitutive model for stress on NiTi relates microstructure with thermo-mechanical behavior of NiTi. A single expression considers the 3 possible existent microstructures and their strain/temperature induced phase transformation.

The size and design of a reconfigurable tool has a strong relationship with size and design of the actuator on which the elements will be positioned in a matrix array.

The use of a reconfigurable actuator such as SMA, makes a more detail design of the pieces and decrements the size and shape of the overall elements, as seen previously in this chapter, making a highly dense pin head per matrix area resulting on a more continuous shape for the discrete shape.

The use of a step by step method such as the one proposed, makes the process an adaptable and enhanced the Vision of Manufacturing Challenges 2020 a reachable goal.

The establishment of the mechanical and electronic parameters of the proposed in order to make a functional prototype, demonstrates that the use of shape memory alloy as actuator can be possible.

The methodology recommended complies the flexibility of a reconfigurable tool according to the Manufacture Vision Challenges 2020, making a modular and adaptable process. (From the object processing image to a virtual matrix array visualizing the final surface created by the matrix arrangement of the pin elements.)

5. References

Azadi, B., Rajapakse, R., & Maijer D. Multi-dimensional constitutive modeling of SMA during unstable pseudoelastic behavior. *International journal of solids and structure*s,Vol. 44, No, 20, (2007), pp. 6473-6490, ISSN 0020-7683

Chang, S., & Wu S. Internal friction of R-phase and B19′ martensite in equiatomic TiNi shape memory alloy under isothermal conditions. *Journal of Alloys and Compounds*, Vol. 437, (2007), pp. 120-126, ISSN 0925-8388

Cortes, J., Tsuta, T., Mitani, Y., & Osakada, K. Flow stress and phase transformation analyses in the austenitic stainless steels under cold working. *Japan Society of Mechanical Engineers International Journal, Vol*. I35 No. 2, (1992), pp. 201-209.

De Castro J A., Melcher K., Noebe R., & Gaydosh D. Development of a numerical model for high-temperature shape memory alloys. *Smart Materials and Structures*. Vol 16. (2007). pp. 2080-2090.

Duering, TW., Melton, K., & Stockel D.Engineering Aspects of Shape Memory Alloys. *Materials and Manufacturing Processes,* (1990).

Guzman, J., Cortes, J., Fuentes, A., Kobayashi, T., & Hoshina, Y. Modeling the displacement in three-layer electroactive polymers using different counter-ions by a phase transformation approach. *Journal of Applied Polymer Science*, Vol. 112, No.6,(June 2009), pp. 3284–3293. ISSN 1097-4628

Humbeeck, J. Non-medical applications of shape memory alloys. *Materials Science and Engineering*, (1999).

Kallio, M., Lahtinen, R., & Koskinen, J. (2003). Smart materials. *Smart materials and structures*. VTT Research Program 2000-2002 Seminar. Retrieved from <http://www.vtt.fi/inf/pdf/symposiums/2003/S225.pdf>

Lahoz, R., & Puértolas, J. Training and two-way shape memory in NiTi alloys: influence on thermal parameters. *Journal of Alloys and Compounds*, Vol. 381, (2004), pp. 130–136, ISSN 0925-8388

Li, M. Multi-point forming technology for sheet metal. *Journal of Materials Processing Technology*, (2002), pp. 333-338.

Li, M. Manufacturing of sheet metal parts based on digitized-die.*Robotics and Computer-Integrated Manufacturing*, (2007), pp. 107-115.

Li, M., Cai, Z., Sui, Z., & Li, X. Principle and applications of multi-point matched-die forming for sheet metal. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture,* Vol. 222, (May, 2008), pp. 581-589.

Malukhin K & Ehmann K. Material Characterization of NiTi Based Memory Alloys Fabricated by the Laser Direct Metal Deposition Process. *Journal of Manufacturing Science and Engineering*. Vol. 128, (2006), pp. 691-696, ISSN 1087-1357

McNaney J., Imbeni V., Jung Y., Papadopoulos P., & Ritchie R. An experimental study of the superelastic effect in a shape-memory Nitinol alloy under biaxial loading. *Mechanics of Materials*, Vol. 35 (2003), pp. 969–986, ISSN 0167-6636

Mehrabi, M., Ulsoy, G., & Koren Y. Reconfigurable Manufacturing Systems: Key to Future Manufacturing. Journal of Intelligent Manufacturing, Vol.11, No.4, (August 2000), pp. 403-419, ISSN 0956-5515

Nakajima, N. A Newly Develop Technique to Fabricate Complicated Dies and Electrodes with Wires. *Japan Society of Mechanical Engineers International Journal*, (1986).

National Research Council. *Visionary Manufacturing Challenges For 2020*, (1998). National Academies Press. Retrieved from <www.nap.edu/readingroom/books/visionary>

Nemat-Nasser S., & Guo W. Superelastic and cyclic response of NiTi SMA at various strain rates and temperatures *Mechanics of Materials.* Vol 38, (2006), pp. 463–474, ISSN 0167-6636

Peng, L. Transition surface design for blank holder in multi-point forming. *International Journal of Machine Tools & Manufacture*, (2006), pp.1336-1342.

Rao, P. A flexible surface tooling for sheet-forming processes: conceptual studies and numerical simulation. *Journal of Materials Processing Technology*, (2002).

Ryhänen J. Biocompatibility Evaluation Of Nickel-Titanium Shape Memory Metal Alloy. *University of Oulu*. (1999). Retrieved from <http://herkules.oulu.fi/isbn9514252217/html>

Schwarz, R. Design and Test of a Reconfigurable Forming Die. *Journal of Manufacturing Processes,* (2002), pp. 77-85.

Shaw, J. A thermomechanical model for a 1-D shape memory alloy wire with propagating instabilities. International Journal of Solids and Structures, Vol. 39, (2002), pp. 1275–1305, ISSN 0020-7683

Srinivasan, A., & McFarland D. *Smart Structures Analysis and Design* (2001), Cambridge University Press. ISBN 0-521-65026-7, Cambridge UK

Varela, M., Cortes, J., & Chen Y. Constitutive Model for Glassy – Active Phase Transformation on Shape Memory Polymers considering Small Deformations.

Journal of Materials Science and Engineering, Vol. 4, No. 5 (May 2010), pp 14-22, ISSN 1934-8959

Vision 2020 Chemical Industry of The Future. *Roadmap for Process Equipment Materials Technology*, (2003). Retrieved from <www.chemicalvision2020.org/techroadmaps.html>

Walczyk, D., & Hardt, D. Design and Analysis of Reconfigurable Discrete Dies for Sheet Metal Forming. *Journal of Manufacturing Systems*, Vol.17, No.6, (1998)

Walczyk, D. A Comparison of Pin Actuation Schemes for Large-Scale Discrete Dies. *Journal of Manufacturing Processes*, (2000), pp. 247-257.

Zhang, Q. Numerical Simulation of deformation in multi-point sandwich forming. *Machine Tools & Manufacturing*. (2006): pp. 699-707.

Zhong-Yi, C. Multi-point forming of three-dimensional sheet metal and the control of the forming process. *International Journal of Pressure Vessels and Piping*. (2002): pp. 289-296.

5

Adaptation of Manufacturing Systems in Dynamic Environment Based on Capability Description Method

Eeva Järvenpää, Pasi Luostarinen, Minna Lanz and Reijo Tuokko
Tampere University of Technology, Department of Production Engineering
Finland

1. Introduction

Nowadays manufacturing systems are characterized by constantly changing requirements caused by short lifecycle times of products, small batch sizes, increasing number of product variants and fast emergence of new technical solutions. Today's turbulent production environment calls for adaptive and rapidly responding production systems that can adjust to required changes both in production capacity and processing functions. The European level strategic goal towards Competitive Sustainable Manufacturing (CSM) asks for the re-use and adaptation of production systems (Jovane et al., 2009). Adaptation allows users to utilize the full lifetime and potential of the systems and equipment and in this way supports the sustainability, both from economic and ecologic perspectives. However, in the previous projects (e.g. (Harms et al., 2008)), it has been recognized that, because of often expensive and inefficient adaptation process, companies rarely decide to adapt their production systems. This is mainly due lacking or insufficient information and documentation about the capabilities of the current system and its lifecycle, as well as lack of extensive methods to plan the adaptation. Today the adaptation of production systems is practically a human driven process, which relies strongly on the expertise of the system integrators or the end user of the system.

Different manufacturing paradigms have been initiated in recent years to overcome the challenges relating to adaptivity requirements. Reconfigurable manufacturing systems (RMS) aim to meet these requirements by offering rapid adjustment of production capacity and functionality, in response to new circumstances, by rearrangement or change in their structure as well as in hardware and software components (ElMaraghy, 2006; ElMaraghy, 2009; Koren et al., 1999; Mehrabi et al., 2000). Agent-based and holonic systems take more dynamic approach to cope with the changeability requirements. Distributed Manufacturing System (DiMS) concept developed by Nylund et al. (2008) and Salminen et al. (2009) is based on holonic architecture, where the holons are autonomous entities able to communicate with other holons and form set of holons, holarchies, through common, well-defined, interfaces and negotiation process. In DiMS the production environment is seen as dynamic and evolving open complex system, where the decision making is based on negotiation process between these entities. Holonic manufacturing systems aim to offer a solution for changeability requirements by providing self-organizing capabilities. Whereas the reconfigurable system research focuses mainly on physical adaptation, in the latter approach the adaptation is performed also on logical and parametric levels. However, no matter if the adaptation is happening on physical, logical or parametric level, intelligent methods and tools are needed to support efficient planning of adaptation.

A critical factor for computer-aided production system design and adaptation planning is efficient resource models, which provide the needed information for equipment selection and system integration. This chapter will introduce a novel method to formally describe the capabilities of resources to support the manufacturing system adaptation. First, in section 2, the term adaptation is defined. Section 3 reviews some existing approaches for describing resource capabilities. Section 4 will then describe the developed capability description method and in section 5 the mapping between product requirements and system capabilities will be covered. In section 6 case application of the capability descriptions in a holonic manufacturing framework, build into TUT heavy machining laboratory, will be discussed. Also the modular software system architecture of the holonic system will be introduced. Finally, the future work is discussed in section 7, followed by the conclusions in section 8.

2. Adaptation of manufacturing systems

This section aims to clarify the concept of adaptation in the context of this research. First, the traditional classification of different adaptation types is introduced. After that the Systems approach to adaptation is discussed. Finally, the relation between capabilities and adaptation is highlighted.

2.1 Classification of adaptation types

Wiendahl and Heger (2004) identified five types of changeability of manufacturing systems: reconfigurability, changeoverability, flexibility, transformability and agility. Later on Wiendahl et al. (2007) used changeability as a general term as a characteristic of a system to accomplish early and foresighted adjustment of the factory's structures and processes on all levels to change impulses economically. Based on the literature around flexible, reconfigurable and adaptive manufacturing (ElMaraghy, 2009; ElMaraghy, 2006; Koren, 2006; Mehrabi et al,. 2000; Tolio & Valente, 2006), it is difficult to completely differentiate these concepts. Flexibility is often referred to the ability to adapt to different requirements without physical changes to the system, whereas reconfigurability refers to the ability to change system components when new requirements arise (ElMaraghy, 2006). However, these definitions can be used only if the boundary of the system is clearly defined. Tolio and Valente (2006) stated that depending on the border, the type of changeability can be interpreted as reconfigurability or as flexibility and therefore, it is not possible to define general statements for these characteristics. ElMaraghy (2006) divided the manufacturing system reconfiguration into both physical and logical reconfiguration touching those both definitions of flexibility and reconfigurability.

In order to avoid misunderstandings caused by the old definitions of reconfigurability, flexibility and other related terms, the term 'adaptivity' is used in this work, including both physical adaptation (reconfiguration) and logical adaptation. Besides those two types, also parametric adaptation is included into the definition. Fig. 1 represents and explains these three types of production system adaptation.

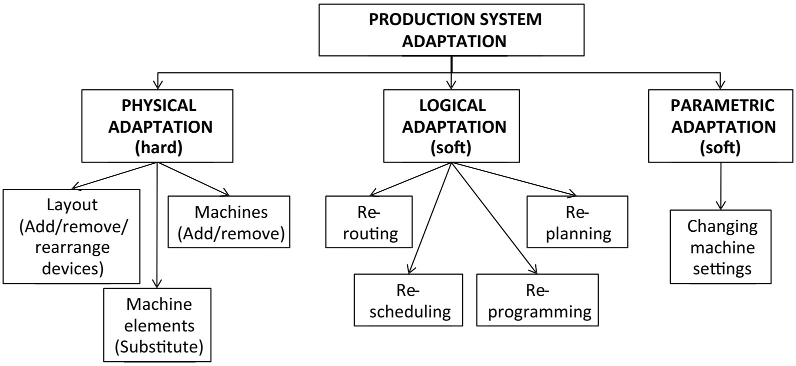


Fig. 1. Types of production system adaptation, modified from (ElMaraghy, 2006).

Adaptation can also be divided into static and dynamic adaptation. Static adaptation is the change of system design during downtime of the system. Dynamic adaptation is the change of system design during the operation of the production system. These dynamic changes be either logical or parametric adaptation. Dynamic adaptation allows the production system to react to changes in its environment in real-time, for example to recover from disturbances on the line and to self-organize itself to balance the production flow. Whereas physical adaptation is usually done on the static level, the logical and parametric adaptation can be either dynamic or static. This means that logical and parametric changes can be executed while the system is running or during its downtime.

2.2 Systems approach for adaptation

Based on the systems approach, in order to achieve adaptability, the system must be able to learn from the experience (Bourgine & Johnson, 2006). The learning is achieved via gaining and understanding the feedback of change – its magnitude and direction. In order to understand the change, the system must be able to compare the past status with the new status of operations. Unfortunately, in the traditional operation environments the knowledge of neither the past nor the present status is in a computer interpretable and comparable form. It would require that the content, context and interaction between those is known. Without this content, context and their interaction information the adaptation to the changes in the environment, from systems approach perspective, is only a theoretical idea without real implementations. Until now, the adaptation has relied on human experience and knowledge, and has therefore been highly subjective.

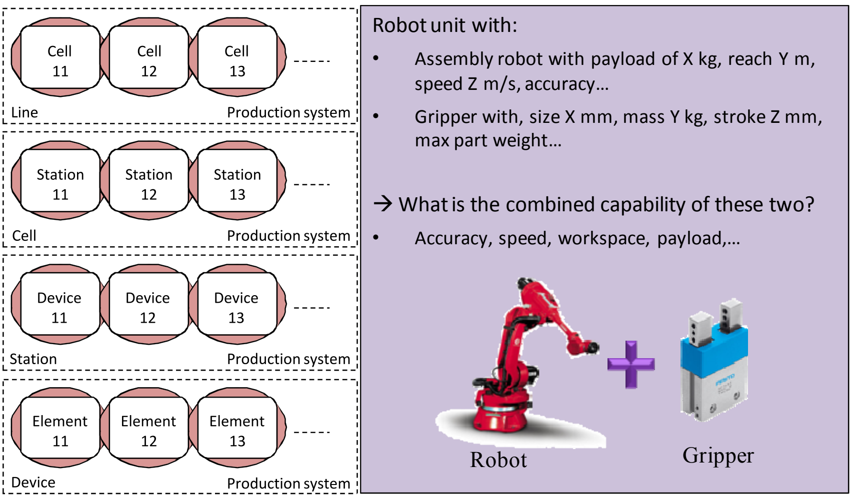
For adaptive production system, feedback loops and understanding the feedback are therefore essential. However, in case of static adaptation, the feedback and its processing don't need to happen in real time. The learning within production system can happen either on human or system level and this learning can turn into adaptivity of any of the types presented in the Figure 1.

2.3 Capabilities in adaptation

When adapting an existing production system for new product requirement two different approaches can be taken. According to Bi et al. (2008) the first approach is to design the new system from scratch and then compare it with the existing system to establish the required changes. This is not very practical method, because it can lead to large and unnecessary changes. The second approach is to start with the original specification of the existing system and to change it until it fits the new requirements. (Bi et al., 2008.)

The adaptation methodology developed by the authors follows the second approach. It is based on comparing and matching the capabilities of the current system with the capability requirements set by the new product. Every device in the manufacturing environment has certain properties and behavior. Some of these properties and behaviors allow the device to perform a technical operation. All of the devices and their properties have certain ranges and constraints. They can be for example technical properties such as maximum torque of the spindle or velocity range of the moving axis, or environmental constraints like maximum allowed humidity and temperature. Automatic matching of available devices against product requirements requires formalized and structured representations of the functional capabilities and constraints of the devices.

However, presenting a simple capability of an individual resource is not enough when designing or adapting complete systems. Adaptation planning problem deals with a heterogenic system level, where a combination of different system levels can be recognized (see Figure 2a). When adapting a production system for a new product, there may be stations that could be utilized as they are without any mechanical changes. However, in order to detect that, there is a need for a formal description of the combined capabilities of the station composed of multiple devices. According to Ueda et al. (2001), the design of production systems follows emergent synthesis where the local interactions between the artifacts of the system form the global behaviour through bottom-up development to achieve the purpose of the whole system. Due to these local interactions, the combined behavior of resources is something else than the sum of the behavior of each individual resource. Consider for example the problem of combination of robot and gripper, illustrated in Figure 2b.



a) b)

Fig. 2. a) Partonomy of different system levels; b) Example of combined capability problem (Järvenpää et al., 2011a).

A method to automatically derive the combined capability of multiple devices would not only enhance the adaptation, but also the original system design. This kind of model would allow the adaptation planning to start from the top, without considering each individual equipment inside the station. During the original system design it would allow to search for those resources, which would together fulfill the capability requirements, without the need to split up the capability requirements into most atomic pieces. In a holonic manufacturing environment (Nylund et al., 2008; Salminen et al., 2009) it would allow the capabilities of holarchies, composed of multiple individual holons, to be formulated based on the capabilities of the individuals.

3. Existing approaches for resource and capability descriptions

Usually, in order to describe production systems, different classification methods are used. Traditionally the devices are classified into groups based on their common properties or functions they provide, e.g. milling machines, lathes and so on. Unfortunately this kind of classification of systems is often limited. If the system is allowed to be member of one class only, then multifunctional system needs to be forced to the most appropriate class, even though it would have functionality fitting to multiple classes. One example of this kind of machine is a multifunctional universal CNC lathe, which is able to perform turning, milling and drilling. Depending on the context, the same device may be used to perform different activities. Therefore, this kind of classification doesn't provide enough expressiveness. To overcome this limitation, instead of classifying devices, the authors classify the functional capabilities that the devices provide. This way one device can have multiple capabilities to be used in different contexts and new capabilities can be assigned for the device when they emerge.

The manufacturing resource information and capability descriptions are considered as a fundamental basis for the various manufacturing activities including process planning, resource allocation, system and facility design as well as planning for system adaptation. By far there haven't been standardized information models to represent the manufacturing resource capabilities. Resource data models and tools for different system components exist, but they are vendor specific or very limited on their scope and not capable for describing the combined capabilities of multiple resources. In this section few existing approaches trying to overcome some of those limitations will be shortly reviewed.

In EUPASS project Skill concept was used to fill the gap between processes and equipment in the ontology. In the EUPASS ontology the skills were divided into basic skills and complex skills. The basic skills are the most fundamental skills, whereas the complex skills are combinations of more simple skills. (Lohse et al., 2008) Emplacement concept was developed to give a standardized description of the EUPASS modules comprising an assembly system (Siltala et al., 2008). Barata et al. (2008) presented a multiagent-based control architecture for shop floor system, where each agent representing a manufacturing component can be aggregated through the Broker Agent GUI to form a coalition of agents that coordinates higher level processes (complex skills) based on the ones available in its members. Based on Cândido and Barata (2007) ontology is used to identify which basic skills are necessary to provide complex skills. Unfortunately, there is no published material about how these two relating approaches solve the emergence of the atomic skills into complex ones, or how they are handling the complex skills within the ontology model.

Based on the example presented in Barata et al. (2008), the skill concept name is used to express the properties of the skill, which indicates that the skills don't have parameters, but just name. No reasoning seems to be done based on the technical properties of the devices. Therefore according to our best knowledge the problem has still remained unsolved.

The work performed at NIST (National Institute of Standards and Technology) by Ameri and Dutta (2008) aims to connect buyers and sellers of manufacturing services in web-based e-commerce environments. The matching is based on semantic similarities of supply and demand in terms of manufacturing capabilities. (Ameri & Dutta, 2008) Smale and Ratchev (2009) proposed a capability-based approach for multiple assembly system reconfiguration. Their work consists of capability taxonomy, capability model and reconfiguration methodology. The capability taxonomy is suited to both equipment specification and requirement definition, whereas the capability model combines the roles of the requirements definition, capability definition and capability comparison. (Smale & Ratchev, 2009.) The underlying problem to be solved in these two above mentioned approaches is somewhat similar to us - matching the existing capabilities with the required ones. However, neither of these approaches considers automatic capturing of the combined capabilities from the individual ones. The available system components are treated as individuals without considering their interfaces or co-operation.

Vichare et al. (2009) developed a Unified Manufacturing Resource Model (UMRM) to represent CNC machining systems and their auxiliary devices, such as workpiece and tool changing mechanisms, fixturing, robotic arms, conveyors, etc., aiming to capture information related to the manufacturing facility and its capabilities. UMRM is based on modeling kinematic chains of machines and is more concentrated on geometric aspects of the system. Therefore it is not adequate for describing the capability of e.g. an assembly system composed of multiple individual devices (not dividing the devices into elements like joints and axes).

4. Capability description method

The core of the developed adaptation planning methodology lies on the capability-based matching of product requirements and system capabilities. Automatic matching of available resources against product requirements requires formalized and structured representations of the functional capabilities, properties and constraints of the resources. This chapter will introduce a novel approach for describing and managing capabilities of manufacturing resources and combined capabilities of multiple co-operating resources with an ontology model. This modeling approach enables matching of products and resources based on their required and provided capabilities and this way supports rapid allocation of resources and adaptation of systems. First the definition of capabilities and combined capabilities will be given. Then the developed method for capability descriptions will be thoroughly discussed followed by the description of the components of the overall resource description.

4.1 Capabilities and combined capabilities

In the proposed approach capabilities are functionalities of resources, such as drilling, milling, moving and grasping (also called as capability concept name). Capabilities have parameters, which present the technical properties and constraints of resources, such as speed, torque, payload, and so on. In other words the concept name of the capability indicates the operational functionality of the resource, whereas the parameters of the capabilities distinguish between capabilities having the same concept name. For example capability with concept name 'moving', has parameters 'velocity' and 'acceleration'. The capability parameters allow determining which resource has the capability that best fits to the given product or production requirement.

Capabilities are divided into simple capabilities, combined capabilities and competences. Combined capabilities are combinations of simple capabilities, usually formed by combination of devices, such as a robot and gripper. Competences are human capabilities. Figure 3 represents the relations between capabilities, competences and combined capabilities.

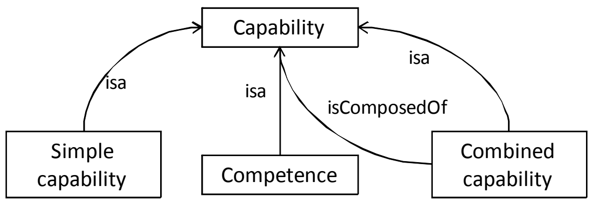


Fig. 3. Relations between capabilities, competences and combined capabilities (Järvenpää et al., 2011a).

There are two types of capabilities:

- **Strong capabilities** are those resource functionalities, which directly provide some kind of process, such as moving, grasping or releasing. Those resource characteristics that are directly related to some simple capability are given as the capability parameters.
- **Weak capabilities** are those properties and characteristics of a resource that do not naturally directly relate to any simple functional capability, but which are important when selecting the resource for a specific application. Weak capabilities are practically additional parameters for the other capabilities to aid in decision making, such as basic device information (containing device dimensions and weight).

4.2 Capability model

Ontologies play an important role in knowledge-based modeling approaches. They give a standardized way to present knowledge from different domains and knowledge sources. Lanz et al. (2008) used ontologies to structure the product, process and system related information, to include the meaning of the content to the models, and allow the information sharing between different applications. In the proposed approach, ontological modeling is used to represent the resource capabilities and constraints. The CoreOntology defined by Lanz (2010) is being used as a basis for describing the product, process and system related information. It allows basic descriptions relating to resources to be formalized. Now in the presented approach, it has been extended for describing the capabilities of the resources, resource interfaces, as well as lifecycle information relating to resources and certain processes. The ontology is saved into a Knowledge Base (KB), described in detail in (Lanz, 2010).

The proposed capability modeling and matching method is based on capability modularization, enabling to build a dynamic link between product requirements and production system. The capabilities are divided into simple and combined capabilities, like discussed in the previous section. The approach is based on functional decomposition of upper level capabilities (combined capabilities) into simple capabilities and assigning these simple capabilities for individual devices in a modular way. In many cases the functionalities of the systems can not be completely decoupled to certain single devices. However, in the presented approach the capabilities are artificially divided on the ontology level to support this capability modularity. When multiple devices are combined, the simple capabilities form combined capabilities. Upper part of the Figure 4b illustrates an example of such division by transportation capability in case of a robot unit consisting of a robot and gripper. The robot alone has only the ability to move its joints within a workspace. When combined with a suitable gripper, together they are able to transport pieces from one place to another.

The modeling of the combined capabilities is handled within the ontology using capability associations as rules how the combinations are formed. In the resource ontology, the devices are assigned the simple capabilities that they posses. Based on the defined capability associations, the device combinations contributing to certain combined capability can be identified and queried. Of course, the devices also need to have matching interfaces to be able to co-operate.

Figure 4a represents the metamodel for defining the combined capabilities. The same combined capability metamodel can be used in different domains. Definition of the domain specific capabilities, as well as the input and output associations for creating combined capabilities, require domain expert knowledge. For example in the context of this research, manufacturing engineering knowledge is required for the definition of the capabilities used in the case studies. Figure 4b represents the usage of the metamodel in the manufacturing domain.

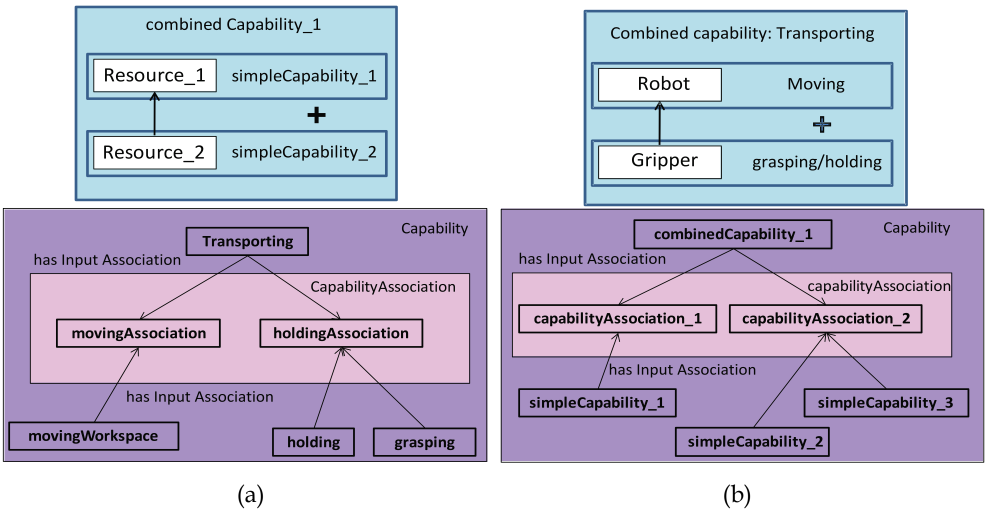


Fig. 4. a) Metamodel for the combined capabilities; b) Instantiated combined capability model.

The principles of the capability model are:

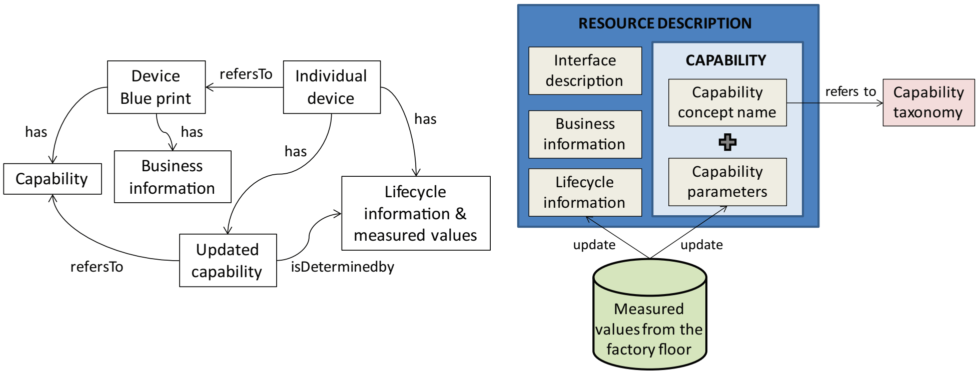
- Resources have simple capabilities, which provide some capability associations as their outputs.
- Combined capabilities require some capability associations as their inputs, e.g. in order to use the combinedCapability_1, both *capabilityAssociation_1* and *capabilityAssociation_2* have to be satisfied.
- Different simple capabilities can provide the same capability association as their output, like simpleCapability_2 and simpleCapability_3 in the Figure 4a.
- When a device or combination of devices provides output for all the required capabilityAssociations (*capabilityAssociation_1* and *capabilityAssociation_2*, in case of Figure 4a) the combined capability (combinedCapability_1) emerges.
- The capability model can have multiple levels, for example the combinedCapability_1 may provide an input association for some other combined capability. In this case the combinedCapability is treated as a simpleCapability from the upper level capability's point of view.
- The capability model can be extended freely upwards and downwards and new capabilities and capability associations can be added through learning processes. The detail level of the model can be further extended without disrupting the whole concept.

Capability model defines the generic capabilities, i.e. a pool of capabilities that can exist in a system. These generic capabilities are assigned to the resources and they become resource specific when filled with resource specific parameter values. With the presented capability model alone it is possible only to merge the capability concept names. On a more detailed level the parameters of the combined capabilities need to be defined, based on the parameters of the individual simple capabilities. The detailed level reasoning with the capability parameters will be handled by holonic reasoning. The holons use a rule base to define the parameters of the combined capabilities.

4.3 Components of the resource description

Manufacturing environment is constantly changing and the condition and capabilities of the resources change during their usage. Therefore it is important that the description of the resource is updated over time. For this reason the devices have two separate, but linked representations within the ontology: device blue prints and individual devices. The device blue print describes the capabilities and properties of one type of device, like given in suppliers' catalogues. The individual devices are presented in a separate class having a reference to the blue print device and presenting the properties of a particular individual device it is a representation of, i.e. individual resource existing on the factory floor. Figure 5a shows the relations between the device blue prints and individual devices.

The representation of individual devices holds the collected and measured lifecycle information of the device, which can be used in the planning process of reuse and adaptation. Collected raw lifecycle and history data is filtered and relevant key figures, such as Mean Time Between Failure (MTBF), Mean Time to Repair (MTTR), maintenance costs, reliability, operation time, estimated remaining lifetime, are calculated and saved as part of the resource description. Also the capabilities of the individual devices can be updated based on the measured values from the factory floor, e.g. if the accuracy of the machine changes. The resource behavior log is constantly recording the operational information of the resources on the factory floor. The knowledge on how a specific resource or system did in a specific process while processing a specific part can later on be used for resource selection for similar products.



(a) (b)

Fig. 5. a) Device blue prints and individual devices; b) Components of the resource description, modified from (Järvenpää et al. 2011a).

Figure 5b represents the components of the resource description. In addition to capability information containing the capability concept name and capability parameters, also collected lifecycle information and interface definition are part of the description. Also business information, such as purchase or rental costs can be added as part of the description to aid in decision making. The capabilities (only functional, i.e. strong capabilities) have a property of "hasCapabilityTaxonomy", by which they refer to the correct level in the capability taxonomy, described more in the next section.

5. Capability based matching of product requirements and systems to support adaptation

The core of the developed adaptation planning methodology lies on the capability-based matching of product requirements and system capabilities. Figure 6 represents the reference architecture and explains the concept of the capability based mapping of product requirements and system capabilities. The mapping is based on capability taxonomy connecting the product and resource domains together. Taxonomy, included into the CoreOntology, is used to make a crude search that matches the resources with required capabilities. The detailed reasoning with the combined capabilities and their parameters is based on holonic negotiation process. Next sections will first discuss about the definition of product requirements and then continuing with the requirement-capability matching based on the ontology definitions and holonic reasoning, as shown in the reference architecture.

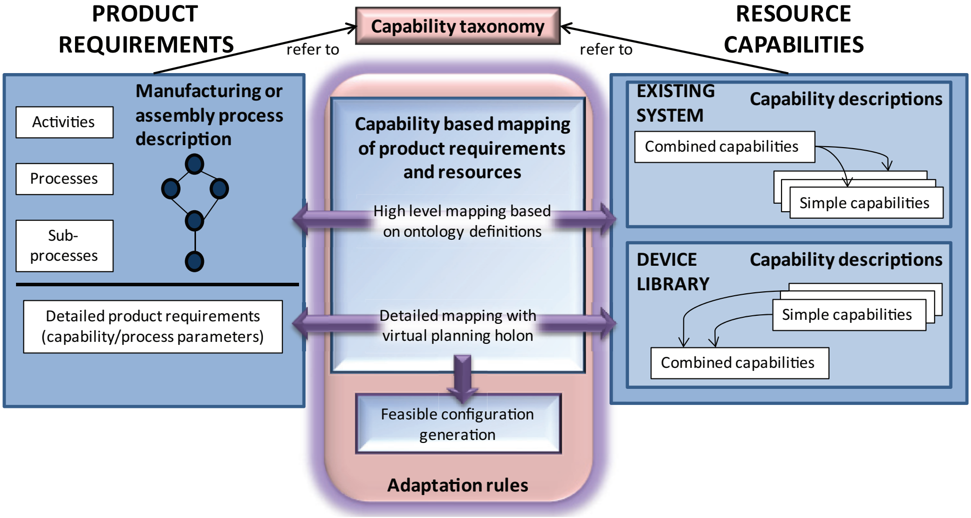


Fig. 6. Reference architecture for capability-based mapping, modified from (Järvenpää et al, 2011a).

5.1 Definition of the product requirements

Product requirements are those product characteristics or features which require a set of processes in order for the product to transform towards the finished product. These processes are executed by the devices and combinations of devices possessing adequate functional capabilities. The product requirements can be expressed by the required capabilities and their temporal and logical order. (Järvenpää et al. 2011b.) In the proposed approach the product requirements are expressed in the form of a pre-process plan, generated by a tool called Pro-FMA Extended (Garcia et al., 2011). Pro-FMA Extended is a software tool, which recognizes the product features from VRML or X3D model and generates the high-level process plan, pre-process plan, based on those features. Basically, the pre-process plan is a generic recipe on how to manufacture the part or product. Each feature contains its characteristics – shape, type, material, tolerance and geometric dimensions – based on which the pre-process plan can be generated. The pre-process plan is an ordered graph of generic activities referring to specific levels on the capability taxonomy stored in the Knowledge Base. (Garcia et al., 2011.)

5.2 Reasoning about the capabilities based on ontology definitions

Capability taxonomy, integrated to the CoreOntology allows, together with the developed capability model, matching between product requirements and resource capabilities on different levels of detail. Like discussed in the previous chapter, the activities in the pre-process plan of the product refer to specific levels on the capability taxonomy. Naturally also the capability instances refer to a certain level in the capability taxonomy enabling the link between products and resources providing the capabilities. See a simplified example of the capability taxonomy in Fig. 7. Due to the limited space, only a small part of the taxonomy with some examples is presented.

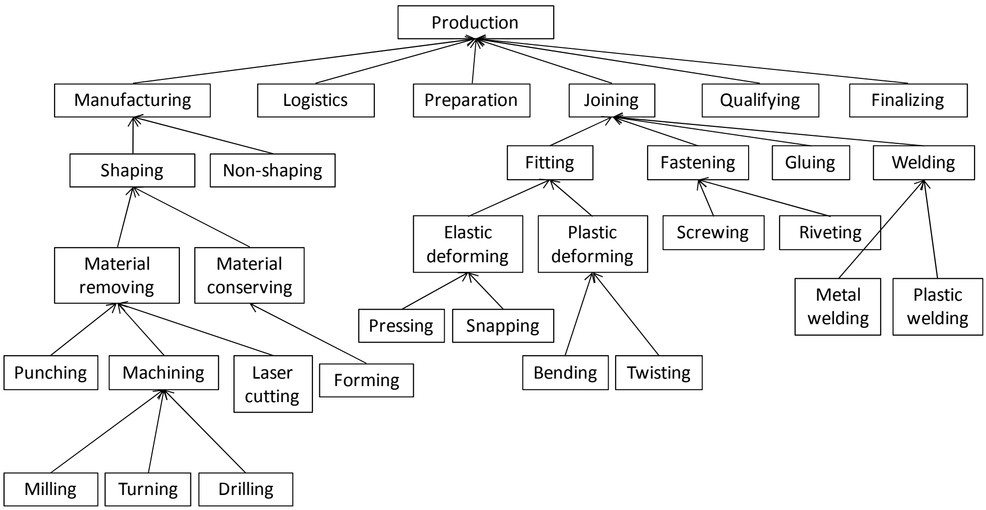


Fig. 7. Simplified capability taxonomy (Järvenpää et al. 2011a).

The taxonomy level where the pre-process plan is referring depends on how detailed information about the required or desired processing methods is available. For example, the product designer may have defined that a specific joining method, such as riveting, should be used to join two parts together or she/he may have only defined that some sort of joining capability is required, leaving possibility to determine the joining method based on the capabilities available on the factory floor. In the first case the product requirement is directed on the particular method in the capability taxonomy, whereas in the latter case the requirement is directed on the joining level in the taxonomy.

The taxonomy allows to search for different devices that are able to perform the same function (e.g. material removing) by different behavior (e.g. milling, turning, drilling,…). The parameters of the capabilities will then determine if the suggested device is able to fulfill the given requirements. For example if the requirement is [material removing, hole of diameter 20mm and depth 50mm, aluminum], the parameters of the capabilities which are subset of the material removing capability (e.g. milling, turning and drilling) will then express, which device combination is able to provide required material removing with required parameters.

As the devices are assigned the simple capabilities they posses, based on the defined capability associations, the device combinations, contributing to certain combined capability, can be identified and queried with SPARQL RDF query language. Similarly it is possible to reason out the capabilities that the resource combinations have. Fig. 8 illustrates how the capability associations (inputs and outputs) are used to make a match between the capabilities existing with the current resources and the required capabilities. By using the associations it is possible to answer e.g. to following questions: "Which devices I need to combine in order to get a certain combined capability? "What combined capabilities a certain combination of devices can have?"

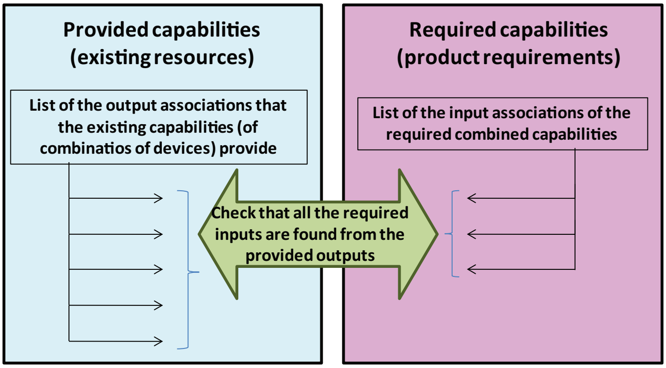


Fig. 8. Matching of capability output and input associations.

Figure 9 gives an example of the matching. The input and output associations are written with italic (e.g. *spinningTool*). As seen in the figure the combined capabilities are formed by hierarchical climbing from lower level combinations to up, e.g. the tool holder and threading cutter combination is considered as threading tool on the next level. Queries are implemented to the Knowledge Base as services that can be called by different applications.

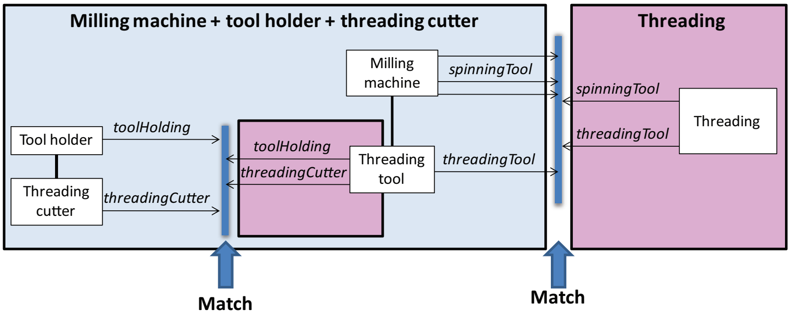


Fig. 9. Example for matching the capability input and output associations.

Summarized, the ontology serves as the representation of taxonomic (is-a) and partonomic (is-part-of) relations between capabilities. The partonomic relations are enabled by the use of the capability associations. As the capability associations are only able to identify the concept name of the combined capability, more intelligent reasoning is required for defining the combined capability parameters. For this purpose, holonic reasoning is applied.

5.3 Holonic reasoning about the capabilities

Ontology is used to make a crude search that matches the resources with required capabilities. The detailed reasoning with the capabilities and their parameters is based on holonic negotiation process. Each different machine on the factory floor has its holon representation that presents that specific machine in a digital world and gets the capability information from the Knowledge Base. When the preliminary matches are found using ontology queries, these scenarios are tested with each holon to make sure that it can actually produce the item in question. This reasoning will output a list of machine-capability pairs that match the request at hand. Since the holons gather the actual lifecycle data of the machines, it can be used in the decision making process. For example, if the accuracy of the machine has not been as good as promised in the data provided by the manufacturer, the holon can adapt new values from the history data and use those for the reasoning.

The holonic reasoning requires rules in which the reasoning can be based on. The rule-base is currently under development, consisting of three types of rules:

Combined capability rules:

- Rules, which determine how the parameters of combined capabilities can be formed from the parameters of individual capabilities. The aim is not to provide detailed analysis of e.g. workspace or kinematics of the device, but to perform scenario modeling of possibly suitable devices and device combinations. For kinematics and detailed workspace definitions virtual simulation tools are used to validate the results got from the reasoning based on the digital information.
- Example rule: When robot and gripper are combined, the payload of the combination is robot payload minus the gripper weight or gripper payload if it is smaller than the previous difference.

Domain expert rules:

- Rules, which define how the capability and its parameter information are applied in each domain when matching with the product requirements. These rules include, e.g. in machining process domain, how the achieved feature depends on the tool shape and type. Our focus is not, however, on defining the domain specific rules in detail, but to demonstrate the concept of how to use those rules and capability definitions in production system design and adaptation in case of changing product requirements and other external and internal changes.
- Example rule: In milling the nose radius of the cutter has to be same as the required rounding inside the machined pocket.

Adaptation planning rules:

- Rules defining how other criteria, such as availability and scheduling, device condition and lifecycle, as well as user and company specific criteria relating for example to costs, eco-efficiency or speed, is used in the final resource selection and configuration generation. These rules are usually given by the user in a specific case and are therefore dynamic in their nature.
- Example rule: If the amount of ordered items is 20 or more, use the fastest machine for the product manufacturing. Otherwise, use the cheapest option.

Despite the generation of rule-base allowing automatic reasoning, human intervention is still required. The goal of the holonic system is not to make everything automatic, but to keep the human involved in the decision making and control loop. In the presented approach it means validation of the automatically generated scenarios and selection of the most desirable solution based on the user specific criteria. Positive aspect with the rule-base is, that it can be extended incrementally on the fly leading to more and more accurate and realistic reasoning results requiring less human intervention.

6. Case implementation for dynamic adaptation

In the context of this research, a dynamic operation environment consisting of the hardware in the TUT machining laboratory environment and modular ICT architecture has been implemented. The build environment utilizes holonic manufacturing paradigm and integrates existing technologies resulted from different projects into one operation environment. The adaptivity of this holonic system rests on SOA (Service Oriented Architecture) -based communication and negotiation between entities through open interfaces, and matching of resource capabilities against product requirements.

The main character of the built holonic system is that the status of the production system and desired goal (defined as order connected to product model) are known, but the steps for reaching the goal, in this case the routing order of the parts in the factory floor, is not predefined. The holonic system follows service oriented architecture (SOA), where the resources provide services through their capabilities. When an order enters to the holonic framework, the system will search for those resources, which can alone, or with some other resource, satisfy the requested service. The holons will then negotiate to determine the best resource for the given situation or the part is directed to first available resource combination that has a capability to produce the part or a specific feature. (Järvenpää et al. 2011c.)

This section will give a detailed description of the implementation of the dynamic operation environment utilizing the capability descriptions. The implementation was built into the Academic Research Environment of TUT. This section will first shortly introduce the hardware part of the environment and then concentrate on the developed modular software system architecture. After those, the process and information flows in the dynamic operation environment will be thoroughly discussed.

6.1 Hardware in the dynamic operation environment

The hardware part of the research environment consists of several manufacturing resources and work pieces as physical manufacturing entities, the real parts (see Fig. 10). Each of them has their corresponding computer models and simulation environments as their virtual parts. The resources of the research environment, offering different manufacturing capabilities, are:

- Machine tools (a lathe and a machining centre) for machining operations
- Robots for material handling and robotized machining operations
- Laser devices for machining, marking and surface treatment
- An automated storage for storing blank parts and finished work pieces
- A punch press, existing only virtually, for the punching of sheet metal parts.

The work pieces are fairly simple cubical, cylindrical, and flat parts in shape. They have several features with parameters that can be altered, such as part dimensions (width, length, and depth), number of holes, internal corner radiuses, sheet thickness as well as material and tolerance requirements of the finished products. These features determine the product requirements for which suitable capabilities need to be found. The system in control level does not distinguish the tasks needed to be done, but the tasks are purely assigned based on the capability requirements.

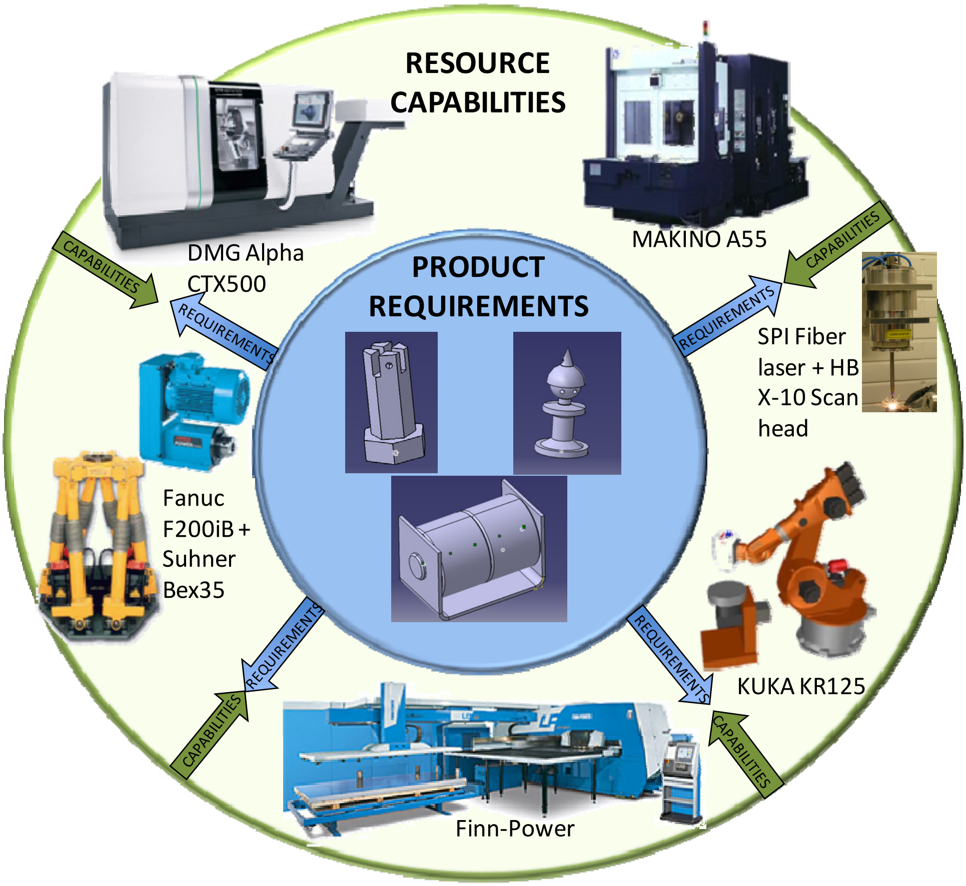


Fig. 10. Resources on the factory floor and manufactured products.

6.2 Modular ICT architecture

The software system architecture, illustrated in Fig. 11, has several different interoperating software modules each providing one or two essential functions for the whole holonic manufacturing system. The architecture follows the dynamic modularization principles being designed in such way that each of the modules can be replaced with a new module if needed without disturbing the whole system. The interoperation of the modules is mainly based on the shared information model and common knowledge representation, the Core Ontology, and modular services. Each of these modules requires specific domain related information and by processing the information they provide a set of services.

Figure 11 shows the simplified information flows between the modules forming the ICT architecture. The communication between the Knowledge Base and the modules is done by RDF and XML messages (depending on the situation) using SOAP. The communication between the DeMO tool, UI/control holon and the machine UIs is done by using XML-RPC calls.

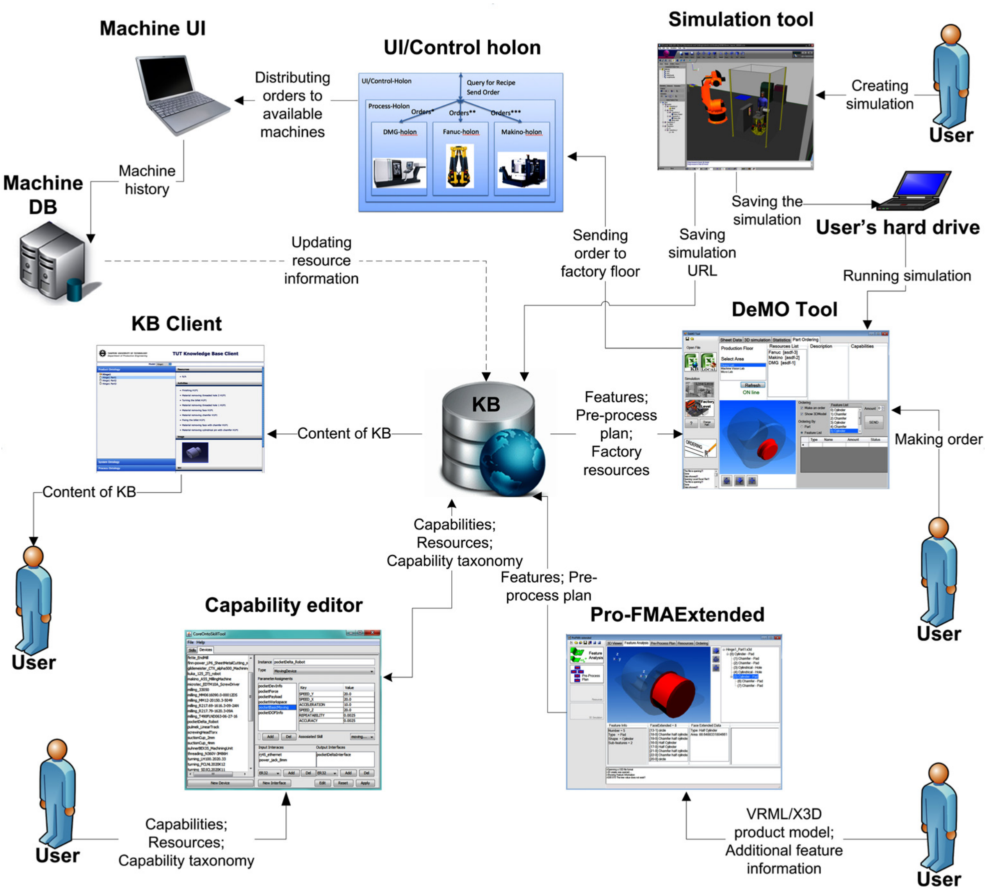


Fig. 11. Modular ICT architecture and simplified information flows during the operation of the dynamic operation environment.

In the following is listed the roles of different ICT tools in the dynamic operation environment:

- **Pro-FMA Extended** is used to define the product requirements from the product model given in VRML or X3D format (Garcia et al., 2011).
- **Capability Editor** allows user to add resources to the ontology and assign them capabilities and capability parameters. It is also used for adding new generic capabilities to the KB, as well as for creating associations between the capabilities. (Järvenpää et al. 2011c.)
- **Simulation tool** is used for verification and creation of the manufacturing or assembly scenarios.
- **Decision Making and Ordering Tool (DeMO tool)** is used for setting up the orders in this environment. The main function of the DeMO tool is to verify connection to factory floor and forward the orders to the holonic UI.

- **UI/Control holon** manages the process flow and distribution of tasks to each manufacturing or assembly cell. This system distributes the tasks to suitable and available cells or stations based on the capability requirements defined by Pro-FMA Extended.
- **The Knowledge Base (KB)** store the information created by Pro-FMA, Capability Editor and DeMO tool in a form of common knowledge representation, Core Ontology (Lanz, 2010). This system serves also as a reference architecture, since it can handle closed models as references.
- **Web-based KB client** is used for human friendly knowledge browsing and content verification. This tool serves as online product data management (PDM) user interface (UI).

6.3 Information and process flows during a case scenario

This section will explain step-by-step the activities taken when new product comes to the production in the developed dynamic operation environment. It will also explain how new resources and new capabilities are integrated into the system on a digital level. Fig. 11 gave an overall view of the information flows in the developed dynamic operation environment. Fig. 12 represents the flow of the activities and reasoning process in the environment when all the available resources are already described in the Knowledge Base. Before discussing in more detail those activities shown in the figure, the resource description activities will be explained. This means that in the next process description, the phases a and b are not shown in the graph below. The process description is divided into activities on digital, virtual and real levels.

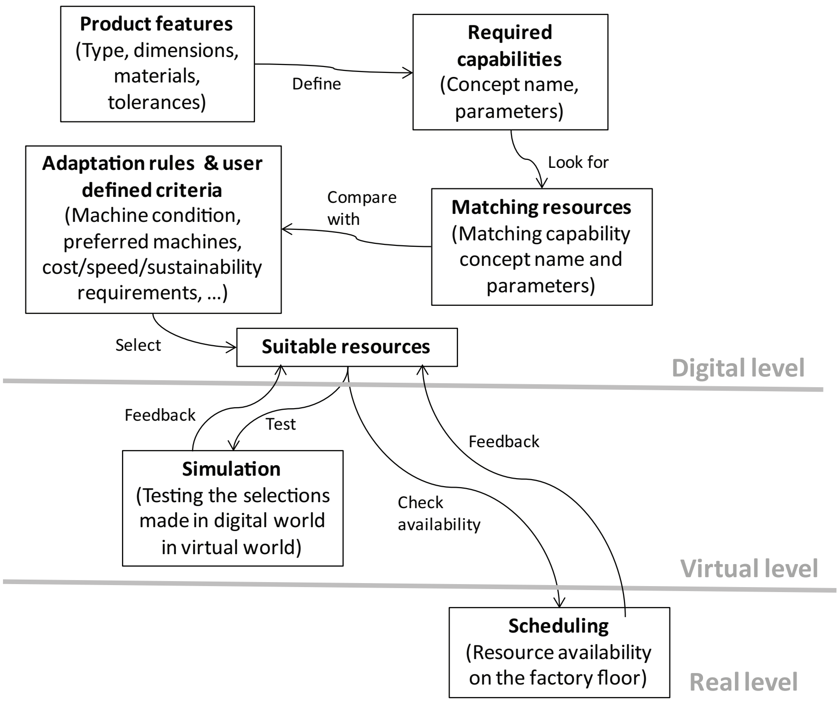


Fig. 12. Process flow in the holonic manufacturing system framework.

Activities on a digital level

a. Adding new resources to the Knowledge Base

New resources are added to the Knowledge Base with the Capability Editor. In the Capability Editor, the user can specify the name and type of the resource and describe its mechanical, control and energy related interfaces. The main task when describing the resources is to assign them capabilities. The user will select the pre-defined simple generic capabilities from the Capability Editor and define the resource specific parameters for the capability, making the capabilities unique for each resource. If a suitable generic capability is not pre-defined, a new capability has to be defined first and then assigned to the resource (see the phase b). Fig. 13 shows the user interface of the Capability Editor for adding resources to the KB. In the ParameterAssignments box, the resource specific instances of the generic capabilities are created and parameter values given.

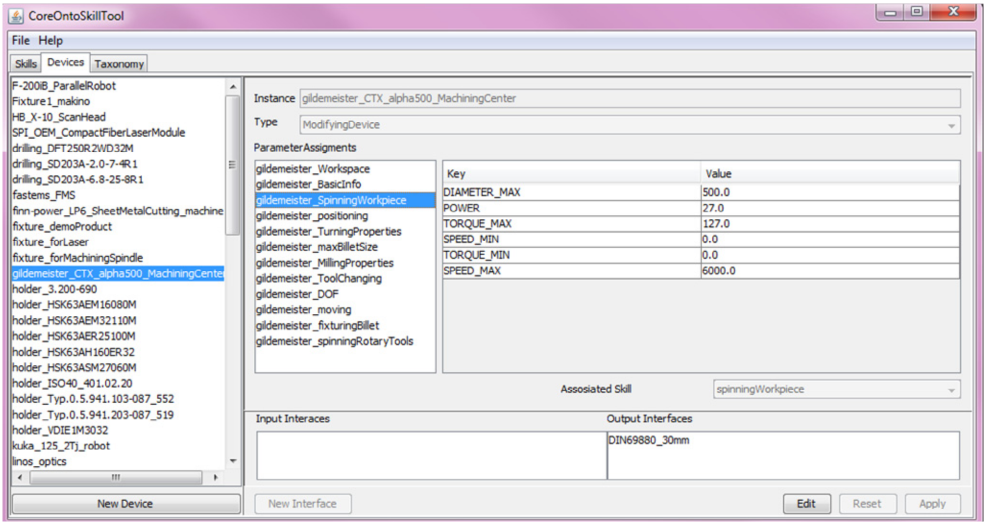


Fig. 13. Capability Editor – Adding resources and capabilities to the Knowledge Base.

b. Adding new generic capabilities to the Knowledge base

If suitable generic capabilities are not pre-defined in the KB, the user needs to first define those with the Capability Editor. Definition of the generic capabilities includes definition of the capability concept name and capability parameters. In case of the strong (functional) capabilities, also the link to the capability taxonomy will be created. Figure 14 shows the user interface for adding capabilities in the Capability editor.

1. Definition of product features

The product model is first sent in VRML or X3D format to Pro-FMA Extended software, which will recognize the features that need to be manufactured. It recognizes the shape, type and dimensions of the feature. The user can manually give the material and tolerance information, as well as special finishing information for the features. This feature information is then sent to the Knowledge Base for further use. The upper part of the Figure 15 presents the user interface of the Pro-FMA Extended with the recognized features and their parameters.

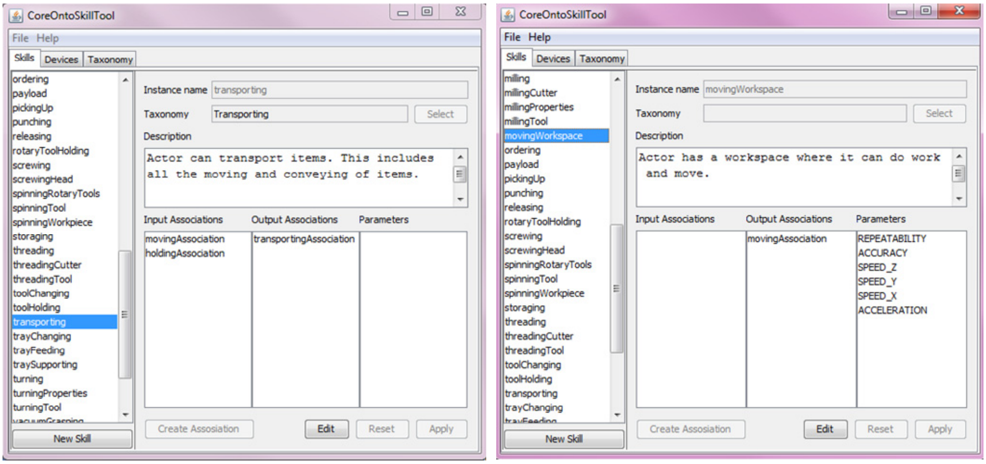


Fig. 14. Capability Editor – Adding new capabilities to the Knowledge Base.

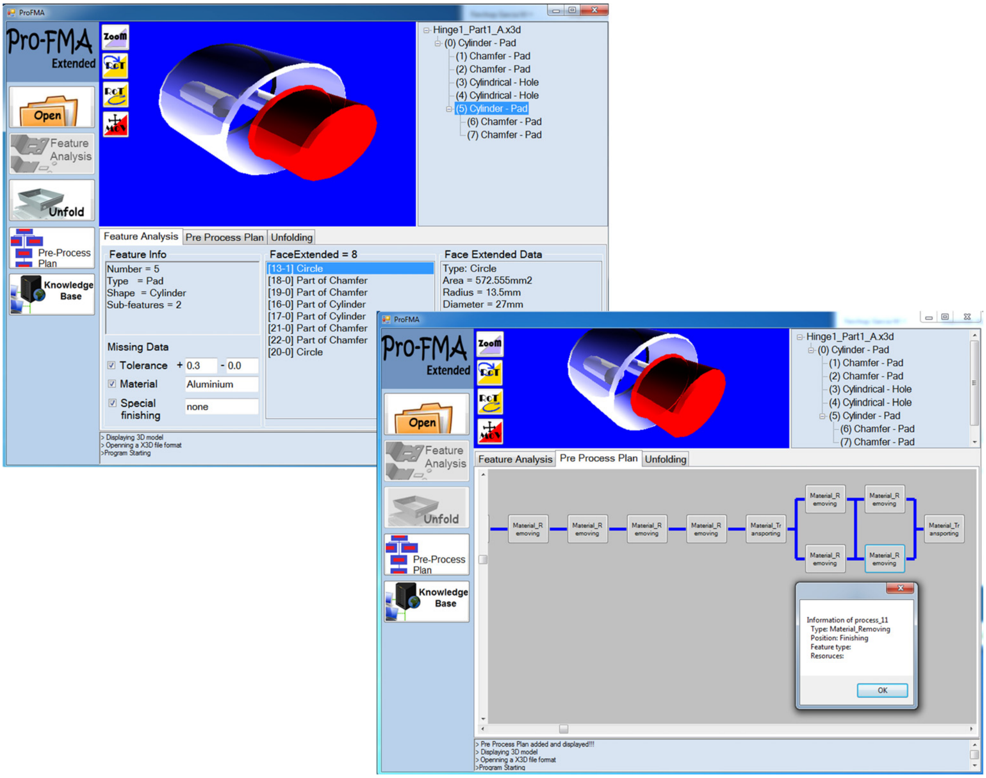


Fig. 15. Pro-FMA Extended – analyzing the product features and creating the pre-process plan, modified from (Garcia et al., 2011).

2. Definition of pre-process plan

Based on the recognized features the Pro-FMA Extended defines a pre-process plan for the part/product manufacturing (see lower part of the Figure 15). The pre-process plan is created based on the rules embedded into the software. The pre-process plan determines the processes on high level (e.g. material removing, material adding) leaving possibility to use different methods (capabilities) for manufacturing the part/product. The pre-process plan is then sent to the Knowledge Base, where the steps of the pre-process plan are linked with the generic capabilities in the capability taxonomy. Pre-process plan together with the feature recognition data defines the capability requirements for the part/product manufacturing in the following way. Generic process names in the pre-process plan refer to certain capabilities in the capability taxonomy. Each step in the graph represents one generic capability requirement. The capability parameters relating to the product features, e.g. size and material requirements remains in the product side in the ontology. Other capability parameters, e.g. required speed, relate to the order specific parameters, like production volume and delivery date. These are defined in later phase when creating the order with DeMO Tool.

3. Matching the capability requirements and resources

The matching of the required capabilities and suitable resources is done based on the capability taxonomy. Both the pre-process plan and the resource specific capabilities have a reference to the capability taxonomy allowing the matching. Definition of combined capabilities is implemented into the KB as services, i.e. other software can use those services to determine the capabilities of device combinations and to search for suitable devices.

4. Applying context specific criteria and rules for the resource selection

The DeMO Tool (Decision Making and Ordering Tool) is used to place the order and send it to the factory floor, as well as to select the preferred resources and apply user defined criteria. First the product to be ordered is selected. The feature information related to that specific product is retrieved from the Knowledge Base. The different manufacturing stations existing currently on the factory floor and defined in the KB are shown in the DeMO tool. While selecting one station, the tool shows the individual machines and tools comprising the station. It also shows the capabilities and combined capabilities of the selected resources. By comparing the required capabilities and provided capabilities the matching resources can be identified. This limits the options of useable stations. Based on the context specific criteria (including adaptation and user specific criteria) the user can then further limit the amount of suitable resources by selecting the most desirable stations. The criteria can relate e.g. to sustainable performance metrics, machine condition, or speed requirements. Figure 16 shows the user interface of the DeMO Tool when sending the order to the specific machine on the factory floor.

Activities on a virtual level

5. Test manufacturing by simulations

With DeMO tool, it is possible to run pre-created simulations of the processes and view the statistics before placing the orders. Because it is very difficult to determine some capabilities, e.g. workspace of the device combination, accurately on a digital level, simulations may be required to validate the feasibility of the matched capabilities.

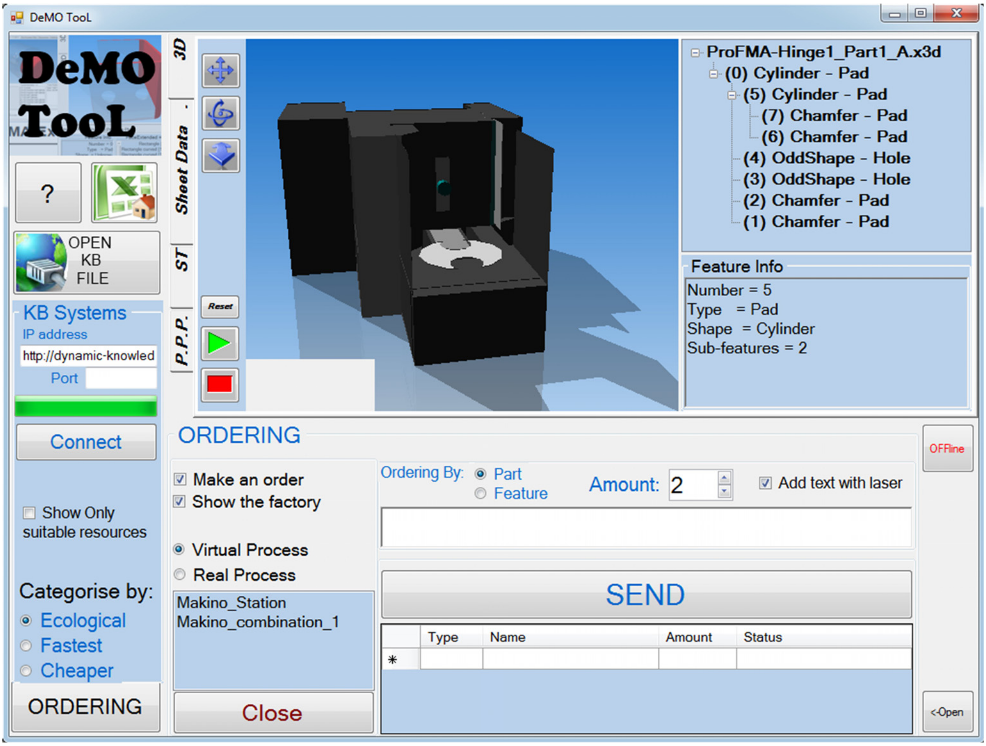


Fig. 16. DeMO Tool – Sending order to the factory floor.

Activities on a real level

6. Checking the availability of resources and routing the order

When the order is placed by the user, the UI/Control holon (scheduling holon) will check the availability and status of the suitable resources on the factory floor by discussing with the machine UIs and then communicate the information with the DeMO Tool. The order is then routed to the first available and suitable manufacturing station (taking into consideration the context specific criteria). The Machine database will save the resource behavior log information, including the status of the resource, completed orders, possible measured values, and so on, which can later on be used for decision making.

7. Future work

In order to make the concept and operation of the dynamic operation environment more intelligent and autonomous, some future developments have been envisioned for the tools forming the environment.

Capability Editor: Currently the Capability Editor is used to manage only the device Blue Print information and the individual devices and device combinations need to be created with the Protégé ontology editor. However, because Protégé is not an optimal tool for manipulating large amounts of data, this function will be later on implemented to the Capability Editor.

Pro-FMA Extended: Currently the pre-process plan defines only the very high level capability requirements, e.g material removing or material adding. In the future the algorithms for the pre-process planning will be further developed to enable more intelligent reasoning based on the recognized features and the additional user given information.

DeMO Tool: Currently the DeMO tool shows only those device combinations that exist on the factory floor (i.e. those, which are created to the KB). Later on it should be possible to display all the possible device combinations having the required capability (also those, which are not existing currently on the factory floor). If suitable capabilities don't exist currently on the factory floor, new combinations of existing devices would be created and shown in the DeMO Tool. The DeMO Tool would discuss with the UI/Control holon about the availability of the devices. Only available devices can make new combinations.

Usage of history data: The history data collected by the Machine DB can be later on used by the UI/control holon e.g. to evaluate how well the resource performed in different situations. This information can then be utilized when resources for similar applications are needed. The history data can also be used for updating the capability information in the Resource KB, e.g. the accuracy. The history data should be handled by the role engine, so that it could be associated with the specific roles that were used while collecting the data. This way the content and context information could be connected enabling knowledge to emerge. This knowledge can then be utilized for successful adaptation.

Rule-base for the holonic reasoning: The development of the rule-base for detailed reasoning about the capabilities has been started, but it is not yet implemented as a part of the system. The implementation of the rule-base will allow more automatic reasoning and more accurate results requiring less human intervention.

8. Conclusions

The operation and business environment changes rapidly. Ability to quickly adapt itself to new requirements has become a crucial enabler for the industry to gain operational flexibility. Support for adaptation is required from all operation levels. A crucial enabler for this kind of dynamic operation environment is modular ICT architecture following the holonic principles. This chapter presented a concept and case implementation of a new kind of dynamic operation environment based on holonic framework. The presented approach enables the step towards more intelligent and adaptive production systems by applying four technical solutions: service oriented architecture allowing the customers to place their orders and resources to advertise their capabilities; open interfaces enabling interoperability; common language and structure for the communication based on the ontology; and holonic negotiation process allowing to make the match between requests and offerings and utilize other criteria for the final decision making.

The introduced capability description method is a crucial enabler for the operation of the presented dynamic operation environment. Formalized capability descriptions allow the holons to advertise their capabilities to other holons in the system, and to autonomously organize the production based on the available capabilities. Capability descriptions allow also automatic methods to find suitable system components and to build alternative scenarios for different product requirements. As the matching based on the capability descriptions can not assure optimal accuracy of found solutions, human intervention is still needed to check the feasibility of the generated scenarios and to select the best one for the given situation.

The presented implementation of the dynamic operation environment provides information of the content and context of a manufacturing and assembly system. However, for achieving real intelligence and adaptivity, the past versus present status needs to be taken into consideration. In the current implementation this comparison is not yet done. More research actions towards a "learning factory" is required in the future.

9. References

Ameri, F. & Dutta, D. (2008). A Matchmaking Methodology for Supply Chain Deployment in Distributed Manufacturing Environments. *Journal of Computing and Information Science in Engineering*, Vol. 8, No. 1, p. 9, ISSN 1530-9827.

Barata, J, Camarinhamatos, L. & Candido, G. (2008). A multiagent-based control system applied to an educational shop floor. *Robotics and Computer-Integrated Manufacturing*, Vol. 24, No. 5, pp. 597-605, ISSN 0736-5845.

Bi, Z.M., Lang, S.Y.T., Shen, W. & Wang, L. (2008). Reconfigurable manufacturing systems: the state of the art. *International Journal of Production Research*, Vol. 46, No. 4, pp. 967-992, ISSN 0020-7543.

Bourgine, P. & Johnson, J. (Eds) (2006). *Living roadmap for complex systems science*, Open Network of Centres of Excellence in Complex Systems, 71 p.

Cândido, G. & Barata, J. (2007). A multiagent control system for shop floor assembly, In: *Holonic and Multi-Agent Systems for Manufacturin,* A. W. Marik, V, Vyatkin, V. Colombo (Eds.), pp. 293–302, Springer, ISBN 3-540-74478-9.

ElMaraghy, H.A. (Ed.) (2009). *Changeable and Reconfigurable Manufacturing Systems*, Springer, ISBN 978-1-84882-066-1, London.

ElMaraghy, H. A. (2006). Flexible and reconfigurable manufacturing systems paradigms. *International Journal of Flexible Manufacturing Systems*, Vol. 17, No. 4, pp. 261-276, ISSN 0920-6299.

Garcia, F., Lanz, M., Järvenpää, E. & Tuokko, R. (2011). Process planning based on feature recognition method, *Proceedings of IEEE International Symposium on Assembly and Manufacturing (ISAM2011)*, ISBN 978-1-61284-343-8, 25-27 May, 2011, Tampere, Finland.

Harms, R., Fleschutz, T. & Seliger, G. (2008). Knowledge Based Approach to Assembly System Reuse, *Proceedings of ASME 2008 9th Biennial Conference on Engineering Systems Design and Analysis (ESDA2008)*, pp. 295-302, ISBN 978-0-7918-4835-7, 7-9 July, 2008, Haifa, Israel.

Jovane, F., Westkämper, E. & Williams, D. (2009). *The ManuFuture Road – Towards Competitve and Sustainable High-Adding-Value Manufacturing*. Springer, ISBN 978-3-540-77011-4.

Järvenpää, E., Luostarinen, P., Lanz, M. & Tuokko, R. (2011a). Presenting capabilities of resources and resource combinations to support production system adaptation, *Proceedings of IEEE International Symposium on Assembly and Manufacturing (ISAM2011)*, ISBN 978-1-61284-343-8, 25-27 May, 2011, Tampere, Finland.

Järvenpää, E., Lanz, M., Siltala, N. & Tuokko, R. (2011b). Activities and Corresponding Information Flows for Production System Adaptation Planning, *Proceedings of the 4th International Swedish Production Symposium, SPS11*, pp. 124-133, 3-5 May, 2011, Lund, Sweden.

Järvenpää, E., Luostarinen, P., Lanz, M., Garcia, F. & Tuokko, R. (2011c). Dynamic operation environment — Towards intelligent adaptive production systems, *Proceedings of IEEE International Symposium on Assembly and Manufacturing (ISAM2011)*, ISBN 978-1-61284-343-8, 25-27 May, 2011, Tampere, Finland.

Koren, Y. (2006). General RMS Characteristics, Comparison with Dedicated and Flexible Systems, In: *Reconfigurable Manufacturing Systems and Transformable Factories*, A. I. Dashchenko (Ed.), pp. 27-46, Springer, ISBN 3-540-29391-4.

Koren, Y., Heisel, U., Jovane, F., Moriwaki, T., Pritschow, G., Ulsoy, G., & Van Brussel, H. (1999). Reconfigurable Manufacturing Systems. CIRP Annals - Manufacturing Technology, Vol. 48, No. 2, pp. 527-540, ISSN 0007-8506.

Lanz, M. (2010). Logical and Semantic Foundations of Knowledge Representation for Assembly and Manufacturing Processes. *PhD Thesis,* Tampere University of Technology, ISBN 9789-52152-3939, Tampere, Finland, p.138.

Lanz, M., Kallela, T., Velez, G. & Tuokko, R. (2008). Product, process and system ontologies and knowledge base for managing knowledge between different clients, Proceedings of the *IEEE International Conference on Distributed Human-Machine Systems (DHMS2008),* ISBN 978-80-01-04028-7, pp. 508-513, 9-12 March, 2008, Athens, Greece.

Lohse, N., Maraldo, T. & Barata, J. (2008). EUPASS std-0007: Assembling Prcess Ontology Specification, *EUPASS project report*, 38 p.

Mehrabi, M.G., Ulsoy, A.G. & Koren, Y. (2000). Reconfigurable manufacturing systems: key to future manufacturing. *Journal of Intelligent Manufacturing*, Vol. 11, No. 4, pp. 403–419, ISSN 0956-5515.

Nylund, H., Salminen, K. & Andersson, P. (2008). Digital Virtual Holons – An Approach to Digital Manufacturing Systems, In: *Manufacturing Systems and Technologies for the New Frontier*, M. Mitsuishi et al. (Eds.), pp. 103–106, Springer, ISBN 978-1-84800-266-1.

Salminen, K., Nylund, H. & Andersson, P.H. (2009). Role based self-adaptation of a robot DiMS based on system intelligence approach, *Proceedings of the 19th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2009),* pp. 964-970, ISBN 978-0-9562303-2-4, *July 6-8, 2009*. Middlesbroug, UK.

Siltala, N., Hoffmann, A., Gengenbach, U., Traivani, E. & Vollmer, H. (2008). EUPASS std-0004: Eplacement Specification and Guideline. *EUPASS project report*, 61 p.

Smale, D. & Ratchev (2009). A Capability Model and Taxonomy for Multiple Assembly System Reconfigurations, *Proceedings of the 13th IFAC Symposium on Information Control Problems in Manufacturing*, pp. 1923-1928, ISBN 9783902661432, 3-5, June, 2009, Moscow, Russia.

Tolio, T. & Valente, A. (2006). An Approach to Design the Flexibility Degree in Flexible Manufacturing Systems, *Proceedings of the 16th Int. Conference on Flexible Automation and Intelligent Manufacturing (FAIM2006),* pp. 1229-1236, *26 – 28, June,* Limerick, Ireland.

Ueda, K., Markus, a, Monostori, L., Kals, H. & Arai, T. (2001). Emergent Synthesis Methodologies for Manufacturing. *CIRP Annals - Manufacturing Technology*, Vol. 50, No. 2, pp.535-551, ISSN 0007-8506.

Vichare, P., Nassehi, A., Kumar, S. & Newman, S.T. (2009). A Unified Manufacturing Resource Model for representing CNC machining systems. *Robotics and Computer-Integrated Manufacturing*, Vol. 25, No. 6, pp. 999-1007, ISSN 0736-5845.

Wiendahl, H. et al. (2007). Changeable Manufacturing - Classification, Design and Operation. *CIRP Annals - Manufacturing Technology*, Vol. 56, No. 2, pp. 783-809, ISSN 0007-8506.

Wiendahl, H.P. & Heger, C.L. (2004). Justifying Changeability. A Methodical Approach to Achieving Cost Effectiveness. *Journal for Manufacturing Science and Production*, Vol. 6, No. 1-2, pp. 33-40, ISSN 0793-6648.

6

Consideration of Human Operators in Designing Manufacturing Systems

Namhun Kim[1] and Richard A. Wysk[2]
[1]Ulsan National Institute of Science and Technology
[2]North Carolina State University
[1]Korea
[2]USA

1. Introduction

A manufacturing system normally includes various types of automated/computer controlled system resources such as material processors (e.g., CNC machines), material handlers (e.g., robots), and material transporters (e.g., AGVs) (Joshi et al., 1995). However, in most cases, to implement fully automated systems where the human is not involved is impractical (Brann et al., 1996), because of both economic and technical reasons. Furthermore, in human-involved automated manufacturing systems, a human can act as one of the most flexible and intelligent system resources in that he or she can perform a large variety of physical tasks ranging from simple material handling to complex tasks such as inspection, assembly, or packaging (Altuntas et al., 2004). From this argument, integrating a human into the system operation is a critical aspect in the design of practical manufacturing systems.

To represent the logical flows of systems' behavior, finite state automaton (FSA), formalism for discrete event-based systems, is widely used in modeling and building a control algorithm of automated manufacturing systems. While FSA-based models can be partially well suited to represent routine human activities, the vast majority of research on control models of human-involved manufacturing systems using FSA tends to consider a human as a system component that can perform tasks without considering dynamic and perceptual conditions of system constraints on human capabilities. (Shin et al., 2006b; Shin et al., 2006c). It is desirable, therefore, to include flexible and dynamic human decision making/tasks in the control of manufacturing systems with consideration of human capabilities and the corresponding system's physical conditions in human-machine co-existing environments.

Under ideal conditions, human operators should be allowed to access all physical components capable of being manipulated in the system (Altuntas et al., 2004). In this sense, a human operator can be considered a distinctive component of the system that is capable of affecting both the logical and physical states of the system. In reality, however, the human can be restricted in affecting the system components given what is afforded (e.g., offered) (Gibson, 1979) by the task environment (e.g., a part on a conveyor may be moving too fast for a human operator to grasp it). To incorporate human capabilities into the system representation, one must consider the control opportunities offered to humans by the system environment as well as the judgment demands placed on human operators.

In this chapter, a framework to develop formalisms for human-machine co-existing manufacturing systems is introduced and illustrative examples are provided in the last section.

2. Modelling of manufacturing systems

The discrete event-based modeling formalism of FSA is introduced in section 2.1. Section 2.2 presents modeling of manufacturing system control using message-based part state graph and its extended version of including human tasks into manufacturing system operations.

2.1 Finite state automata representation of DES

The fundamental physical properties of nature are considered to be continuous in that they can be expressed using real values as time changes. As systems have become more complex, event-driven approaches have become commonplace for a variety of models. Several computer technologies employ discrete methods to control complex systems such as communication networks, air traffic control, automated manufacturing systems, and computer application programs (Cassandra and Lafortune, 1999; Zeigler, 1976). Discrete event-based system modeling is a common tool to represent physical behaviors of systems, including continuous systems that are broken into discrete models which are suitable for DES-based software applications.

One of popular formalisms used to represent the logical behavior of discrete systems is based on the theories of languages and automata. This approach is based on the notion that any discrete event system can be modeled with discrete states and an underlying event set associated with it. An automaton, formalism for discrete systems, is an atomic mathematical model for finite state automata (FSA). It consists of a finite number of states and transitions that enable the model to jump between states via predetermined rules. These jumps are incurred by transition functions. These transition functions determine which state to go to next, given the current state and a current input symbol. An FSA is an effective technique capable of representing a language according to well-defined rules, which means it is rule-based and the state of the system is tractable (Sipser, 2006).

A commonly used FSA in practice is a *Deterministic Finite Automaton* (DFA), which can be defined as a 5-tuple (Hopcroft, 2001);

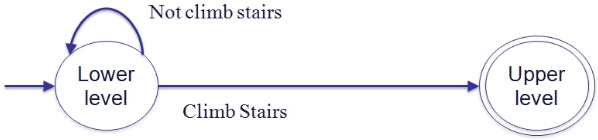
$$M^{DFA} = < \Sigma, Q, q_0, \delta, F>,$$

where;

Σ is a set of input alphabets (a finite non-empty set of symbols),
Q is a finite and non-empty set of states,
q_0 is an initial state such that $q_0 \in Q$,
δ is a state transition function such that δ: $Q \times \Sigma \rightarrow Q$, and
F is a set of final states such that $F \subseteq Q$.

For example, a representation of the 5-tuple FSA for the person-climbing-stairs system is shown in Figure 1. A transition from a lower level to an upper level occurs immediately following the action of 'climb stairs' which is an input symbol to a current state "lower level."

This model only represents the physical aspects of systems behavior without considering the resource availability, and a person's attention and capability to accomplish a specific action (e.g., climb stairs). To better model human participation, it is essential to take into account the conditions required for human actions which consist of affordances (walk-on-ability) and effectivities (capability to walk) in the systems as will be explained in Section 3.



where;
$\boldsymbol{\Sigma}$ = {Climb Stairs, Not climb stairs},
$\boldsymbol{Q}$ = {Lower level, Upper level},
q_0 = Lower level,
δ (Lower level, Climb Stairs) = Upper level; δ (Lower level, Not climb stairs) = Lower level, and
$\boldsymbol{F}$ = Upper level.

Fig. 1. An FSA representation for the person-climbing-stairs system.

2.2 Control model of manufacturing systems

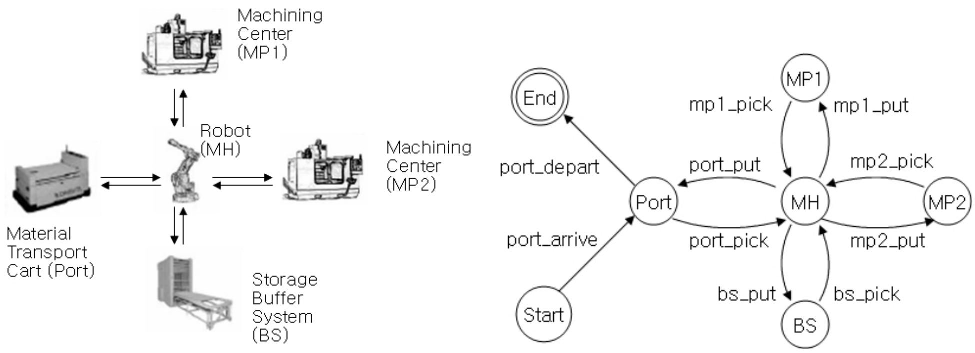
2.2.1 Message-based part state graph (MPSG)

In the 1990's, a formal model for control of discrete manufacturing systems was developed based on FSA, and called MPSG which is an acronym for Message-based Part State Graph (Smith et al., 2003). It is a modified deterministic finite automaton (DFA) similar to a Mealy machine. The MPSG model consists of sets of vertices (nodes) and edges (transitions) which correspond to the part states and the command messages, respectively. The trace of a part advancing trough an automated manufacturing system is described by its part flow diagram, which shows the sequence of part processing states in the system. As shown in Figure 2, the part state graph of a part is represented with a set of vertices and a set of edges. A vertex represents a part position in the part state graph and an edge corresponds to an operation associated with the part.

Fig. 2. An example of a part state graph for a MP class.

A MPSG describes the behavior of a controller from the parts' point of view, and each part within the domain of the controller is in a particular 'state' as described by the MPSG for that controller. The MPSG model provides no information about the system states; it determines which controller events are 'legal' with respect to that part and how to make a transition when one of these legal events occurs.

In the MPSG, all equipment level manufacturing resources are partitioned into material processors (MP; such as numerical control (NC) machines), material handler (MH; such as robots), material transporters (MT; such as automated guided vehicles (AGVs)), automated storage devices (AS), and buffer storage (BS), based on the types of their functionalities. We can create simplified physical connectivity graphs based on the MPSG controller. Figure 3 depicts a physical connectivity graph of a system, which consists of two MPs, MH, and BS, that represents physical interactions and accessibilities among the pieces of equipment. From a system's point of view, the connectivity graph is quite similar to the automaton that consists of states and transitions. However, for the individual resources, more detailed and sophisticated state transition mechanisms need to be considered, and the MPSG enables to describe the states of the entities (parts) in the system by means of the physical connectivity graph.



(a) Physical layout of the system. (b) Representation of connectivity graph.

Fig. 3. Connectivity graph with two MPs, MH, BS, and port.

The MPSG $\boldsymbol{M}$ is defined formally as an eight-tuple, $\boldsymbol{M}=<\boldsymbol{Q}_M, q_0, \boldsymbol{F}, \boldsymbol{\Sigma}_M, \boldsymbol{A}, \boldsymbol{P}_M, \delta_M, \gamma>$, where definitions of the components are as follows:

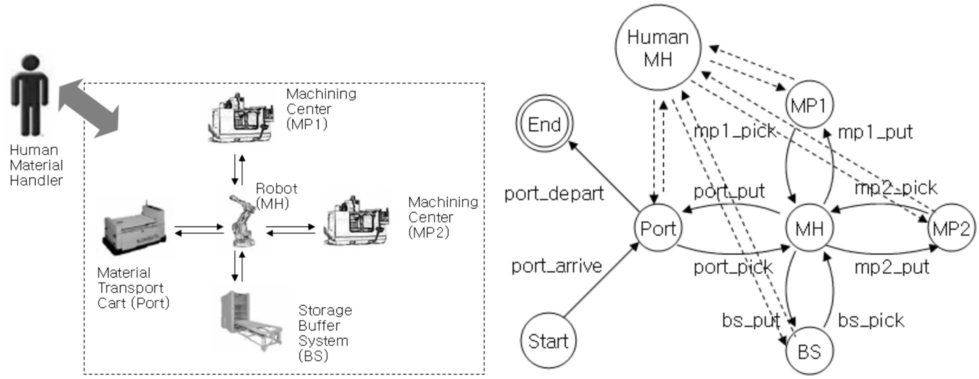
$\boldsymbol{Q}_M$: Finite set of states,
$q_0 \in \boldsymbol{Q}_M$: Initial or start state,
$\boldsymbol{F} \subseteq \boldsymbol{Q}_M$: Set of final or accepting states,
$\boldsymbol{\Sigma}_M$: Finite set of controller events,
$\boldsymbol{A}$: Finite set of controller actions,
$\boldsymbol{P}_M$: Physical preconditions,
$\delta_M : \boldsymbol{Q}_M \times \boldsymbol{\Sigma}_M \rightarrow \boldsymbol{Q}_M$: State transition function, and
$\gamma : \boldsymbol{Q}_M \times \boldsymbol{\Sigma}_M \rightarrow \boldsymbol{A}$: Controller action transition function.

2.2.2 Extended MPSG for human-involvement in manufacturing systems

The MPSG is a formal representation of a shop floor controller and assumes that all the resources are run in an automated way without any human involvement. To incorporate human characteristics into an automated manufacturing systems, Shin et al. investigated human-involved manufacturing systems and developed a novel formal representation by adding the tuples associated with a human element to the MPSG model (Shin et al., 2006b).

The extended MPSG model enables a human operator to cooperate with the automated pieces of equipment.

In Figure 4, solid arcs represent connections between two pieces of equipment made by automated MH equipment, whereas dotted arcs are newly created ones made by a human operator who plays as a material handler. In general, when a human operator who performs material handling tasks in a system that consists of n pieces of equipment is considered, $2 \times n$ of arcs for human transitions are created (Altuntas et al., 2004). It should be noted that the complexity of the connectivity graph increases in a linear manner.



(a) Physical layout of the system. (b) Representation of connectivity graph.

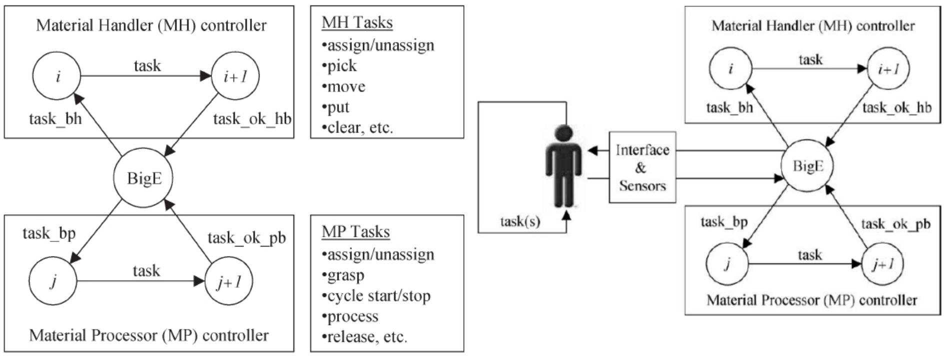
Fig. 4. Change of connectivity graphs with consideration of a human MH.

In order to express the newly created transitions by incorporating a human operator, the representation of part states is extended by incorporating information about a part location within a system and a part handling subject such that it becomes $\boldsymbol{Q} = \boldsymbol{Q}_M \times \boldsymbol{L} \times \boldsymbol{I}(p)$, where $\boldsymbol{L}$ represents a set of physical locations in the system and $\boldsymbol{I}(p)$ is an interaction status with a human. In this way, an extended MPSG with a human operator, denoted by $\boldsymbol{M^E}$, is constructed. It is defined formally as also the eight-tuple, $\boldsymbol{M^E} = <\boldsymbol{Q}, q^E_0, \boldsymbol{F_E}, \boldsymbol{\Sigma}_E, \boldsymbol{A}_E, \boldsymbol{P}_E, \delta_E, \gamma_E>$, where the definitions of the components are as follows:

$\boldsymbol{Q = Q_M \times L \times I(p)}$: Finite set of states, where the set of state $\boldsymbol{Q_M}$ is the state of the original MPSG controller,
$\boldsymbol{q^E_0 \in Q}$: Initial or start state,
$\boldsymbol{F_E \subseteq Q}$: Set of final or accepting states,
$\boldsymbol{\Sigma_E = \Sigma_M \cup \Sigma_H}$: Finite set of controller events, where $\boldsymbol{\Sigma_M}$ is a set of messages for a machine operation and $\boldsymbol{\Sigma_H}$ is a set of messages associated with human actions,
$\boldsymbol{A_E = A \cup}$ ***{actions caused by human activities}*** : Finite set of controller actions and human actions,
$\boldsymbol{P_E}$: Set of Preconditions of the extended controller,
$\boldsymbol{\delta_E : Q \times \Sigma_E \rightarrow Q}$: State transition function,
$\boldsymbol{\gamma_E : Q \times \Sigma_E \rightarrow A_E}$: Controller action transition function,
$\boldsymbol{L}$: Set of all physical locations in the system, and
$\boldsymbol{I(p)}$: Indicator function of interaction status with a human. If a human is dealing with a part $\boldsymbol{p}$, $\boldsymbol{I(p)=1}$,. Otherwise, $\boldsymbol{I(p) = 0}$.

In the concept of the human-involved semi-automated system, its control depends on the complexity of a system, since a controller should recognize current status of the system and provide a proper set of commands for possible tasks based on the logical and physical preconditions. Hence, when a human material handler (human MH) performs tasks during system operation, assessment of the part flow complexity of the system needs to be conducted in developing an effective and efficient control mechanism for the system. The part flow complexity represents the possible number of tasks with a part and the possible outcomes of the tasks in terms of part states (Shin et al., 2006a).

Using this point of view, the major difference of the control schemes between the automated system and the human-involved semi-automated system is whether a human act as a passive resource of the system or a supervisory controller. The human MH can play a role as a self-regulating component which does not subordinate to the computer controller whereas other automated components perform operations in response to a given command for the controller. As such, the human MH can be considered to act as a supervisory controller, and he or she shares the system information via interfaces and sensors as shown in the Figure 5. This perspective will be further developed to expand the human's participation in complex systems.



(a) MPSG controller. (b) Human-involved extended MPSG.

Fig. 5. Control scheme of the MPSG and extended MPSG controllers (Shin et al., 2006a).

3. Modelling and control of human-machine cooperative manufacturing systems

In section 3.1, a representation of human-involvement considering prospective human action opportunities (affordance) is introduced. The modeling basis and formal control model for affordance-based human-machine cooperative manufacturing system are presented in section 3.2 and 3.3, respectively. The example of affordance-based MPSG system control with a simple and typical manufacturing cell is illustrated in section 3.4.

3.1 Human-involvement in system representation

In dynamic situations, the interactions between humans and environs play a key role in achieving an ecosystem's goal. Identifying opportunities for interactions between them is important to the modeling and operation of human-involved systems in an effective way. In this section, we introduce a formal modeling methodology that combines human actions into the system control scheme in formal mathematical FSA.

The concept of affordances implies that human-involved systems are composed of two or more related objects including at least one human and one environmental component, (an affordance complementary property consisting of the dual relationship between animals (humans) and their environs). The terms of affordance and effectivity are treated as an environmental reference and the animal's capability to take actions in the environment. In the sense of a formal representation of affordances, the environmental and animal components are combined together so that they incur a different property to be activated (Turvey, 1992).

Turvey presents a formal definition of affordances mathematically using a juxtaposition function as follows;

> Let $W_{pq}=j(X_p,Z_q)$ be a function that is composed of two different objects X and Z, and further p and q be properties of X and Z, respectively. Then, p refers to an affordance of X and q is the effectivity of Z, if and only if there exists a third property r such that:
>
> i. $W_{pq}=j(X_p,Z_q)$ possesses r,
> ii. $W_{pq}=j(X_p,Z_q)$ possesses neither p nor q, and
> iii. Neither X nor Z possesses r, where r is a joining or juxtaposition function.

For example, a person (Z) can walk (q), stairs (X) that can support something (p), and they together yield a climbing property (r) as shown in Figure 6**.** This formal definition corresponds to a mathematical formalism of an FSA in that it describes properties as discrete states, and the juxtaposition function can be mapped to the state transition function in the FSA. The existence of a formal definition of an affordance provides a foundation that the concept of an affordance can be combined with software engineering and systems theory.

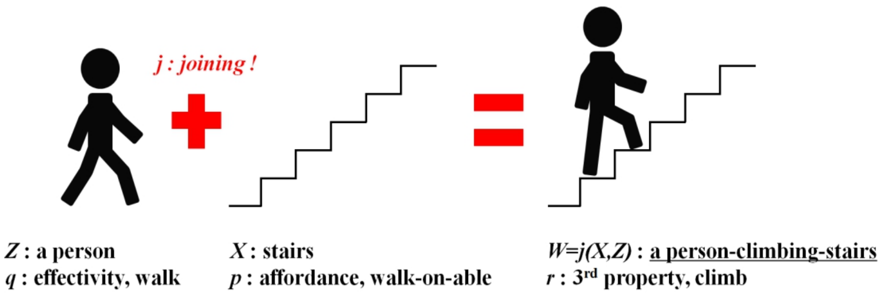


Fig. 6. An example of a 'person-climbing-stairs' system.

If we regard the states of the environmental system as discrete ones and consider the transitions among the states which are triggered by possible actions of animals or other system resources, an ecosystem of an environment and humans can be represented by an FSA (Kim et al., 2010). The theory of automata corresponds to the ecological sense of affordances for at least the following two reasons: 1) an environmental system can be defined as a set of nodes and arcs which describe discrete states of the system and the transitions between states, respectively, and 2) a set of transitions between states represents a set of potential properties (affordances) of the environmental system which can be triggered by certain human activities and lead to the next states. Therefore, affordance-effectivity combinations can be considered conditions for identifying possible human actions using FSA representations.

There is a set of physically connected transitions from one state to another, which corresponds to a set of dispositional properties of affordances in the system. The set of feasible transitions is triggered if and only if the input symbol is taken as a parameter of a transition function in the environmental system. This input symbol is considered an effectivity. Next, the circumstances need to be specified in order for a human transition to occur in terms of the general representation of the FSA, $M^{DFA} = <\Sigma, \mathbf{Q}, q_0, \delta, \mathbf{F}>$. The conditions that allow humans to make transitions within a system can be represented by a four-tuple, $<\mathbf{X}_p, \mathbf{Z}_q, J, \mathbf{W}_{pq}>$, which comes directly from Turvey's definition of affordance. By merging these two sets of tuples, an extended automaton for incorporating affordances of a system and effectivities of humans within the system can be constructed. The new representation for the formal model of affordance and effectivity in FSA is $M^{DFA'}=<\Sigma, \mathbf{Q}, q_0, \delta, \mathbf{F}, \mathbf{X}_p, \mathbf{Z}_q, J, \mathbf{W}_{pq}>$,where;

J is a Juxtaposition function such that $J: \mathbf{X}_p \times \mathbf{Z}_q \rightarrow \mathbf{W}_{pq}$,
$\mathbf{X}_p$ is a set of affordances in the system,
$\mathbf{Z}_q$ is a set of effectivities of human in the system,
$\mathbf{W}_{pq}$ is a set of possible human actions in the system, and
all other definitions of tuples are the same as those of M^{DFA}.

The graphical representation of the affordance-based FSA, $M^{DFA'}$, for the person-climbing-stairs system is shown in Figure 7. Transition from a lower level to an upper level occurs, *if and only if a human is able to 'walk ($\mathbf{X}_p$)' and the stairs are 'walk-on-able ($\mathbf{Z}_q$)' for human, which means 'Climb Stairs (system input of human action) ∈ W_{pq}.'*

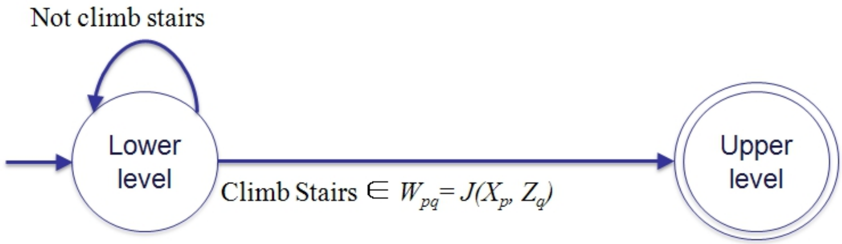


Fig. 7. Affordance-based FSA for the person-climbing-stairs system(Kim et al., 2010).

From an ecosystem's perspective, if the set of all transitions among the system states can be considered Σ. A state transition occurs only when the transition for some input alphabet is included in the set of transitions in the system, $a \in \Sigma$, where a represents an input alphabet. From the human's point of view, he or she has a set of effectivities (capabilities or available actions) regardless of the transitions included in Σ. Thus, transitions occur if and only if the set of possible human actions, W_{pq}, are executed (the results of juxtaposition between specific affordance and effectivity). Component, $h \in W_{pq} \subseteq \Sigma$, represents the possible set of actions for a human to actualize on the environmental system, causing state transitions. In this sense, the dispositional properties that come from joining the properties of affordance and effectivity are considered possible human actions on the human-environmental system. In many unstructured instances, the set of actions can be an infinite set, but for well-structured environs the set of actions can be a very small set. The relationship among affordances of a system, effectivities, and actions of human can be depicted as shown in Figure 8.

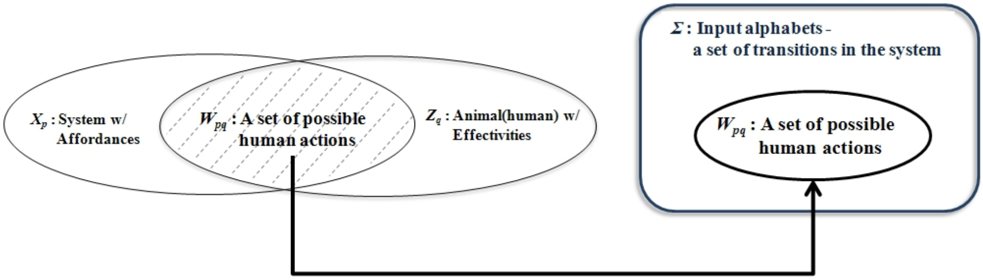


Fig. 8. System affordances, human effectivities, and actions in the ecological point of view (Kim et al., 2010).

The FSA-based modeling formalism for manufacturing systems control can take human activities into account, where source and sink state nodes are defined within the system state behaviors. However, the existing control model of human-involved manufacturing systems lacks prospective control perspectives. It only considers human operators as flexible system components acting like robots, rather than animals which have nondeterministic natures of recognitions and physical limitations. To develop the formal model of human-machine cooperative systems, the ecological sense of system affordances and human effectivities should be included in the model for the seamless control of the systems.

Special care needs to be taken for human operators since their actions are those of a nondeterministic autonomous agent that perceives, measures, and makes a judgment in the system in consideration of other resources and environmental aspects. For this reason, affordances for a human operator in the system need to be considered carefully for human-machine cooperative systems. It can then contribute to assess the human effects on the system in a more effective way.

3.2 Representation of affordances in human-involved manufacturing systems

For a formal control model of human-involved manufacturing systems, it is necessary to incorporate affordances within a system that accounts for possible human actions with regard to at least the material handling processes with consideration of physical limitations for the actions, such as size, weight, and temperature. This corresponds to distinguishing possible human actions from human capable actions (effectivities). We remark that the set of possible human actions are a part of the collection of human *potentially* capable actions considering that human may or may not take actions due to his or her cognitive recognition of actions or physical limitations imposed by an environment. From the viewpoint of the manufacturing system with a human material handler, the affordance can be described as;

Define W_{pq} as a set of *possible human actions in manufacturing system*. Let X_p be a physical state of a part or a piece of equipment in a system where p is a human accessibility (affordance) to it, and Z_q be a human material handler where q is a human action if and only if there is the third property r such that:

i. $W_{pq}=j(X_p,Z_q)$ possesses r : human material handing action (location change),
ii. $W_{pq}=j(X_p,Z_q)$ possesses neither p nor q, and
iii. Neither X nor Z possesses r, where r is a joining or juxtaposition function.

In order to incorporate system affordances into the manufacturing system controller, a formal representation of occurrences of the third properties, called dispositions, needs to be established. For some typical possible human actions for material handling (e.g., *access*, *pick*, *move*, and *put*), the corresponding circumstance can be specified as follows (Kim et al., 2010);

The specific classes of human activities to be addressed include the following:

1. For a human material handler to be able to **access** a machine (resource), the machine should be stopped before the human starts to work and other MHs (robots) should not run on it. Also, the human needs to perceive that he or she can work on the machine.
2. For a human material handler to be able to **pick** a part, at least one degree of freedom (DOF) of the part needs to be released and the human can separate the part from the position.
3. For a human material handler to be able to **move** a part, he or she should be holding the part, and the central position of the part can be changed in the global Cartesian coordinate system as a result of the ***move*** action.
4. For a human material handler to be able to **put** a part on a machine, the machine should be stopped and the number of parts on the machine should not exceed the capacity of the machine. Also, the machine needs to support the part (fixture without slip).

As described above, there are some obvious state transitions for a human material handler in the system. Based on the examples, we decompose human material handling tasks into four types of actions (*access*, *pick*, *move*, and *put*), and also define system affordances and human capable actions corresponding to them as follows;

<Affordances in the manufacturing system with human material handlers>

The specific classes of human activities to be addressed include the following:

1. A machine is **accessible**; the machine is stopped and waits to process a part, and no other MH is working on it. The machine volume should be within the human's access ranges.
2. A part is **pickable**; the chuck or fixture holding the part is open, and at least one DOF of the part is available. The part should weigh less than maximum lifting force, and should be less than maximum grapping width for a human material handler.
3. A part is **movable**; the part is held by a human, and the location of the part can be changed by human actions. There are no *substantial* obstacles from a starting point to an ending position of the human.
4. A part is **putable**; the machine stops working, and it can support the part upright without slip.

<Human MH's capable actions (effectivities) in manufacturing systems>

1. A human material handler can **access** a piece of equipment.
2. A human material handler can **pick** a part from a piece of equipment.

3. A human material handler can **move** a part to a piece of equipment.
4. A human material handler can **put** a part to a piece of equipment.

The third property in Turvey's affordance formalism is mapped on a subset of possible human actions. By doing this, the juxtaposition function can be formulated based on its definition. In the definition of the set of possible human actions, denoted by $W_{pq}=j(X_p,Z_q)$, j is the joining or juxtaposition function. If X_p and Z_q have multiple dispositions, the juxtaposition function j needs to filter p and q from the dispositions possessed by X_p and Z_q to realize the possible actions of W_{pq}. To construct a juxtaposition function to address this, X_p and Z_q are expressed as row matrices that consist of '0' and '1', which represent a certain property *exists* ('1') or *not* ('0') in the system. Thus, in a manufacturing system with human, the sets of P and Q can be expressed as following equation (1) and (2), respectively;

$$P = (\textit{Accessible, Pickable, Movable, Putable})\text{: Properties of the system} \tag{1}$$

$$Q = (\textit{Can Access, Can Pick, Can Move, Can Put})\text{: Properties of a human} \tag{2}$$

By multiplying each component in the matrices, the juxtaposition function of this problem is defined as in equation (3) to obtain the third properties and possible state transitions;

$j : X_p \times Z_q \rightarrow W_{pq}$ and $\pi : P \times Q \times C \rightarrow PA$, where P is a set of affordance status for a part state, Q is a set of action capability (effectivity) status to the human operator, PA is a set of possible human actions in the system, and C is a set of physical action conditions (preconditions for human actions).

Suppose, $P = \{p_i : i=1,2,3,4\}$, $Q = \{q_j : j=1,2,3,4\}$ where p_i and q_j are binary numbers, then

$$PA = \begin{cases} \phi & \text{if } p_1q_1=0. \\ \{(p_2q_2 \times '\textit{pick}'),(p_3q_3 \times '\textit{move}'),(p_4q_4 \times '\textit{put}')\} - \{0\} & \text{if } p_1q_1=1 \text{ \& } C \text{ is true.} \end{cases} \tag{3}$$

Note that the empty set refers to a situation that a human operator cannot access resource.

3.3 Formalism for human-machine cooperative systems: Affordance-based MPSG

As mentioned in the previous section, some human actions become available depending on the environmental affordances, and transitions made by human actions can be realized by satisfying both system affordances and corresponding human effectivities (capable actions). Affordances should have ontological assumptions related to space and time as in Gibson's ecological definition (Gibson, 1979). In the sense of system controller such as the MPSG, it is one of the key factors to build formal representation of the affordance concepts that imposing quantifiable metrics on affordances.

From the MPSG point of view, the supervisory controller, called a Big-E, has a module to generate possible transitions based on the logical validation of preconditions as shown in the Figure 9. The existing MPSG controller generates process plans based on the fully automated systems that are assumed to properly operate as planned beforehand. In this sense, as long as the system is working without *critical* failures, the human action is not necessary and a human is allowed to intervene between machine operations whenever he or she decides to do so. However, when an unanticipated incident occurs (e.g., machine down, oversized part), which is usually beyond the controller's resolution capability, human involvement is required. In this case, the Big-E controller notifies a human of the case so that a human operator can step in the process for preceding the system to the next available proper transition (Kim et al., 2010).

From the viewpoint of a human operator, he or she could make a transition in the system to move a part forward toward completion (one of the feasible ways to proceed when the system requires some human action). The set of feasible transitions are mapped into the Big-E controller, which can generate possible alternative action commands based on the logical validation modules. It is worth note that this exactly corresponds to the set of system affordances for this case. It should also be noted that not all feasible transitions are available for the human operator because the system affordances for the human operator have ontological assumptions of physical and time domains, as mentioned above.

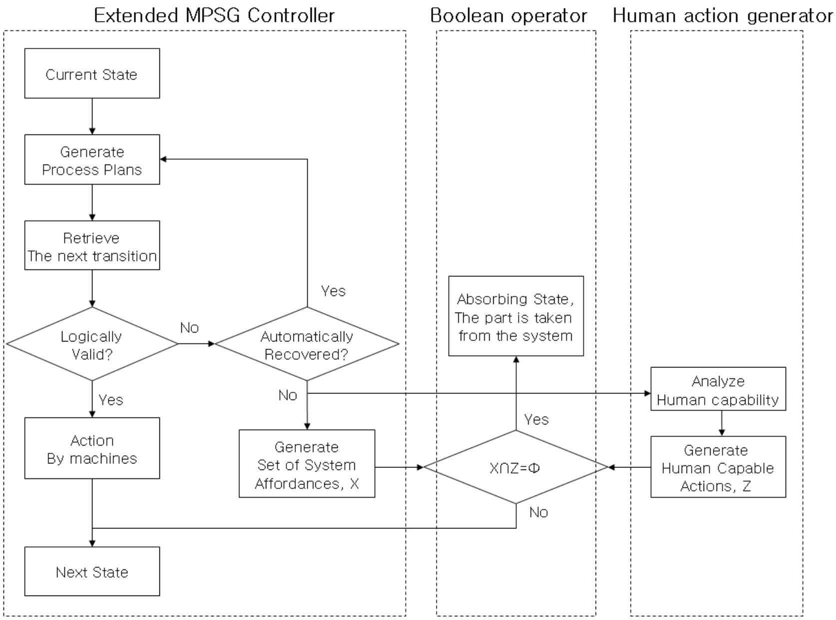


Fig. 9. Control flow of human-involved automated system with consideration of affordances (Kim et al., 2010).

In order to realize a human cooperative system in the ecological sense, generation modules for two distinctive logical sets and the Boolean operator for juxtaposing these two logical sets need to be constructed for a human operator to cooperate with the controller with consideration of affordances as shown in Figure 9.

Considering the formal representation of affordances, the extended MPSG for human-involved system control can be improved in such a way that it can consider more realistic transitions by human possible actions. In this chapter, the affordance-based MPSG, denoted by M^A, is defined as a 12-tuple, which comprises eight-tuples from the initial extended MPSG model, M^E , and four-tuple from the affordance representation. It is defined formally as $M^A = <Q, q^E_0, F_E, \Sigma_E, A_E, P_\alpha, \delta_E, \gamma_E, X, Z, J, W>$, where the definitions of the components are as follows:

J is a juxtaposition function such that *J*: $\boldsymbol{X} \times \boldsymbol{Z} \rightarrow \boldsymbol{W}$,
$\boldsymbol{X}$ is a set of affordances,
$\boldsymbol{Z}$ is a set of effectivities (human capable actions),
$\boldsymbol{W}$ is a set of possible human actions,

where;

$$\mathbf{J}(x(p,\mathbf{l}), z(p,\mathbf{l}))=\mathbf{W}$$

$$=\begin{cases} \varnothing & if \quad x_1 z_1 = 0 \\ j(X,Z)=\begin{Bmatrix} x_2 z_2 \times pick\_p\_from\_l_1, \\ x_3 z_3 \times move\_p\_from\_l_1\_to\_l_2, \\ x_4 z_4 \times put\_p\_on\_l_2 \end{Bmatrix} - \{0\} & if \quad x_1 z_1 = 1 \end{cases},$$

where;

$\mathbf{l} \subseteq \mathbf{L}$, $x(p, \mathbf{l}) \in X$, $z(p, \mathbf{l}) \in Z$,
$x(p, \boldsymbol{l})=x(p, \{l_1,l_2\})$= (a location set $\{l_1,l_2\}$is accessible, part *'p'* is 'pickable' at l_1 , part *'p'* is movable from l_1 to l_2 , part *'p'* is 'putable' on l_2), and
$z(p, \boldsymbol{l})=z(p, \{l_1,l_2\})$= (access to a location set $\{l_1,l_2\}$, pick the part *'p'* at l_1 ,move the part *'p'* from l_1 to l_2 ,put the part *'p'* on l_2)'
δ_E is a state transition function such that $\delta_E : \boldsymbol{Q} \times \boldsymbol{\Sigma}_E \rightarrow \boldsymbol{Q}$.,

where;

$$\delta_E((v,l,I(p)),a,W) = (\delta_M(v,a),l,0), \; if \; a \in \Sigma_M \quad (by\ a\ MH)$$

$$\delta_E((v,l,I(p)),a,W) = \delta_H((v,l,I(p)),a,W), \; if \; a \in W \subset \Sigma_H \; (by\ a\ human)$$

$$\delta_E((v,l,I(p)),a,W) = (v,l,I(p)) \; if \; a \notin W \subset \Sigma_H \quad (no\ transition),$$

where δ_M is a state transition function by automated MHs (robots) and δ_H is a state transition function by a human material handler, and all other definitions of tuples are the same as those of $\boldsymbol{M^E}$.

Based on the above definition, the juxtaposition function can generate a set of possible human actions under a particular circumstance when system affordances are defined as environmental situations, time limitation, physical layout of a system, and part properties, e.g., size, volume, and speed. The human transition set of the affordance-based MPSG is a subset of that of the extended MPSG as shown in Figure 10.

Thus, the complexity of the MPSG controller of human cooperative systems can be reduced when the concept of affordances are taken into account. In the extended MPSG controller, $\boldsymbol{M^E}$, physical preconditions, $\boldsymbol{P}_\alpha$, are evaluated so that some impossible transitions can be prevented. However, the physical preconditions may account for only a small part of system affordances that can be measured by pre-installed sensors, while most of possible human transitions are determined by human cognitions. It is noteworthy that system affordances for humans have much greater impact on the operations and control of the human-machine cooperative systems.

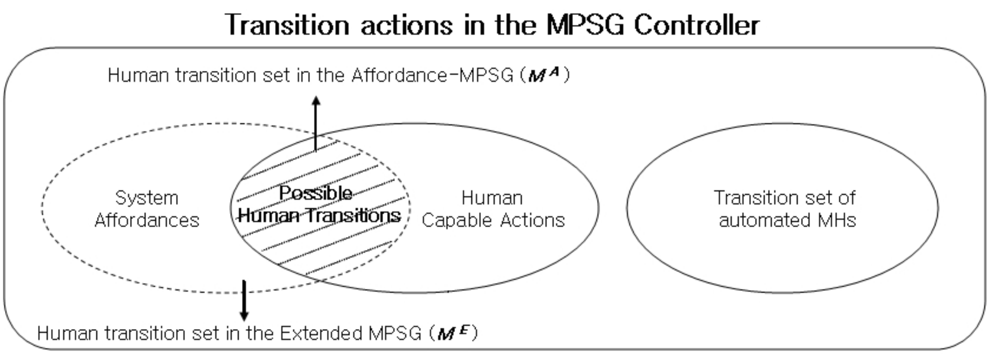
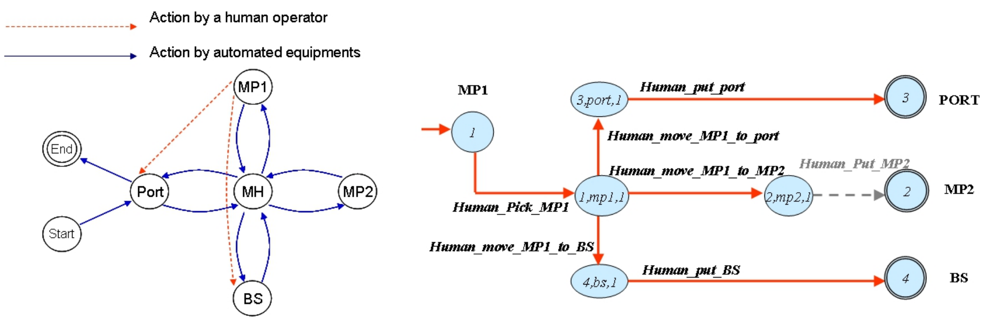


Fig. 10. Transition action sets in the Affordance-based MPSG controller (Kim et al., 2010).

3.4 Illustrative example: Affordance-based MPSG model

This section presents an application example to illustrate the proposed manufacturing control model with affordances. As shown in Figure 11, two types of graphs are constructed to represent the system's physical configuration and the logical control logic. The first graph shows the relationship among the resources in a system and possible path for parts. Based on this connectivity graph in Figure 11(a), the affordance-based FSA representation in Figure 11(b) can be created to develop a control scheme for the system. This is then used to generate an affordance-based MPSG controller that incorporates operations of each piece of equipment and possible human actions (Kim et al., 2010).



(a) Connectivity graph (b) Affordance-based MPSG representation

Fig. 11. FSA representation; MP2 has no affordance of put-ability for human operators (Kim et al., 2010).

Specifically, Figure 11 depicts a case in which a human operator can move a part from 'MP1' to anywhere when the part is not 'putable' on 'MP2', i.e., the MP2 is located so far from the operator that he or she cannot see if the MP2 is empty. The affordance and effectivity matrices between 'node 1' and 'node 2' are expressed with the proposed model as follows,

$$x(part,\{\text{MP1},\text{MP2}\}) = (x_1, x_2, x_3, x_4) = (1,1,0,1) \text{ and}$$

$$z(part, \{\text{MP1}, \text{MP2}\}) = (z_1, z_2, z_3, z_4) = (1,1,1,1).$$

So, the juxtaposition function can be,

$$W = \left\{ \begin{array}{c} 1 \times pick\_part\_from\_\text{MP1}, \\ 1 \times move\_part\_from\_\text{MP1}\_to\_\text{MP2}, \\ 0 \times put\_part\_on\_\text{MP2} \end{array} \right\} - \{0\}$$

$$\therefore W = \{pick\_part\_from\_\text{MP1}, move\_part\_from\_\text{MP1}\_to\_\text{MP2}\}$$

If the human material handler wants to make a transition between 'MP1' and 'MP2', he or she needs to take three actions (pick, move, and put) between the nodes. However, the action, 'put', is not available in the system. It means that by taking the affordances in the system, the complexity of the graph in terms of the number of possible human actions in the FSA representation can be reduced.

The eligible affordances and effectivities of the example can be expressed as follows;

The affordance chart at time *t*:

$$x(part, \{\text{MP1}, \text{MP2}\}) = (1,1,0,1)$$

$$x(part, \{\text{MP1}, \text{BS}\}) = (1,1,1,1)$$

$$x(part, \{\text{MP1}, \text{PORT}\}) = (1,1,1,1)$$

The effectivity chart at this point:

$$z(part, \{\text{MP1}, \text{MP2}\}) = (1,1,1,1)$$

$$z(part, \{\text{MP1}, \text{BS}\}) = (1,1,1,1)$$

$$z(part, \{\text{MP1}, \text{PORT}\}) = (1,1,1,1)$$

From the above affordances and effectivities relationships, we obtain;

$$\begin{aligned} W = \{ & Pick\_part\_from\_\text{MP1}, Move\_part\_from\_\text{MP1}\_to\_\text{MP2}, \\ & Move\_part\_from\_\text{MP1}\_to\_\text{BS}, Move\_part\_from\_\text{MP1}\_to\_\text{PORT}, \\ & Put\_part\_on\_\text{BS}, Put\_part\_on\_\text{PORT}\} \end{aligned}$$

If the controller is to allow a part transition between MP1 and MP2 by a human material handler, *W* should contain a complete set of actions which is composed of *pick*, *move*, and *put* between MP1 and MP2. However, *W* does not have '*put*' on MP2 actions in itself in this example. Thus, the human operator cannot make the part transit between MP1 to MP2 as a material handler.

4. Function allocation between human and machine

Dynamic task allocation control scheme for realization of human-machine cooperative systems is introduced in section 4.1. Classification of errors and their recoveries in human-machine cooperative systems using affordance-based MPSG are presented in section 4.2.

4.1 Work allocations in human-machine cooperative systems

Sheridan (2000) discusses a list to assert "what men are better at" and "what machines are better at" (MABA-MABA) as follows;

<Humans are usually conceived to be better at>:

1. Detecting small amount of visual, auditory, or chemical energy.
2. Perceiving patterns of light or sound.
3. Improvising and using flexible procedures.
4. Storing information for long periods of time, and recalling appropriate parts.
5. Reasoning inductively.
6. Exercising judgment.

<Machines are better at>:

1. Responding quickly to control signals.
2. Applying great force smoothly and precisely.
3. Storing information briefly, erasing it completely.
4. Reasoning deductively.

The gaps between 'what machines are better at and what humans are better at' are getting narrower as machines are replacing human more and more with the development of artificial intelligence technologies. However, the complete replacing humans with the automated machines are almost impossible and impractical partly because of both economic and technical reasons (Brann et al., 1996).

In this sense, the function allocations between machines and humans in the human-involved automated system are one of the vital factors to control the system in effective and flexible ways. As pointed out in the previous section 3.1, human actions are available depending on the environmental affordances, and transitions by human can be realized by satisfying both system affordances and corresponding human effectivities Thus, from the system point of view, the controller needs to differentiate the set of actions that humans are better at from actions that machine are better at with consideration of availability of human actions identified by the model.

Suppose that the material handling time (e.g., time for picking up, moving, and putting a part) and material lifting capability (e.g., part weight, volume, size, and temperature) can be critical factors to allocate work between human material handlers and robots in a manufacturing cell. If the controller is able to evaluate the availability of a resource (either human or machine) which can reduce a processing time for a material handling job at a certain point of time and space, the whole system works faster and more intelligent to increase its productivity. For example, if we consider a simple human-machine cooperative manufacturing cell shown in Figure 4 with following characteristics;

Time for a robot to move a part from a resource to a resource = 10 ± 0.5 sec.,
Time for a human to move a part between adjacent resources = 5 ± 2 sec.,
Time for a human to move a part between facing resources = 10 ± 5 sec., and
Human capable part size and weight ≤ 3ft × 3ft × 3ft and 20 LB.

The controller is able to evaluate expected average processing time for each task and allocate the task between a human operator and a robot based on information of the dynamic location of a part and the human operator, and system working status. For instance, if the human operator is waiting for a message from Big-E within three seconds walking distance from MP1, and a part, whose size and weight are 1 ft × 1 ft × 1 ft and 10 LB, needs to be moved from MP1 to BS, the expected average time of the human task to move the part is eight seconds and that of the robot is 10 seconds. Thus, the human operator is supposed to be faster than the robot to accomplish this specific task, and the controller will allocate the task to the human operator as shown in Figure 12. In this case, the external transition function in affordance-based MPSG needs to be revised as follows,

$$\delta_E : \boldsymbol{Q} \times \boldsymbol{\Sigma}_E \rightarrow \boldsymbol{Q}$$

$\delta_E((v, l, I(\text{part})), a) = \delta_H((v, l, I(\text{part})), a)$ if $a \in PA \subseteq \boldsymbol{\Sigma}_H$ and the human is expected to perform a task faster than the automated MH, and
$\delta_E(v, l, I(\text{part})), a) = (\delta_M(v, a), l, 0)$, otherwise.

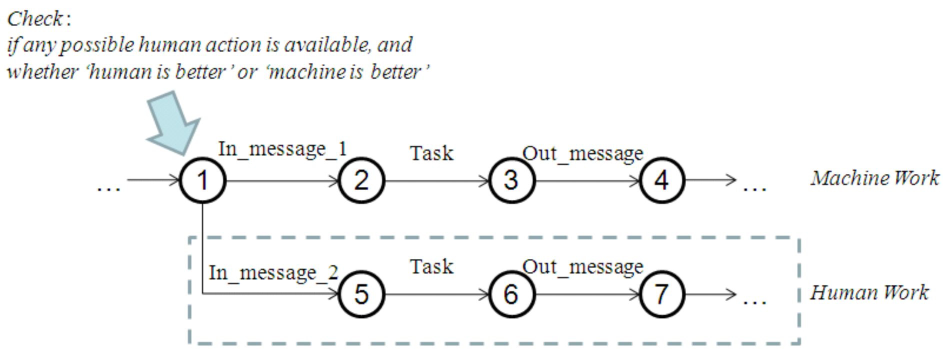


Fig. 12. Task allocation between human and machine in affordance-based MPSG.

4.2 Classification of errors

In the perspectives of systems theory and controls, a human agent is neither completely controllable nor perfectly predictable because of his or her nondeterministic and complex behaviors. For this reason, human-machine interactive system models need to harness dynamic human decision making processes into discrete system contexts. The level of modeling grains is defined with respect to the modeling purposes and modelers' perspectives on the systems. Representation of systems using finite numbers of states and transitions poses a lot of challenges to make a model complete by itself. Thus, comprehensive definition and classification of errors and error states in discrete system models can increase modeling easiness, simplicity and completeness.

For instance, the human-involved automata model of 'a semi-automated manufacturing system' illustrated in section 3.4 should contain an additional system state of the absorbing (error) state. In this modeling representation, human actions and system transitions that may not lead to the desired states, which come from the goal of the human-involved system, directly go to the absorbing state. Only valid interactions between a human and a system can be parts of a human-involved or human-machine cooperative process that change system states from a current to a next state which is placed within the process to the desired goal states.

It may not be critical to investigate errors in descriptive system representation as mentioned above. However, in the perspectives of system control models, system recoveries from errors are important to accomplish the seamless and complete modeling of human-involved systems. Thus, investigation of errors and their proper classification in systems are one of keys to develop control models for human-machine cooperative systems.

4.2.1 Error classification in extended MPSG systems

In human-involved systems, human errors are considered important factors from a control point of view because sometimes system status is significantly changed by the human errors which are not within traceable states. It is well known that there are a number of topics to be addressed in terms of human errors. Shin et al. (2006c) investigated human operational errors concerning the human material handling in extended MPSG controls. In the authors' research, only human operational errors that are directly related with physical material handling tasks are considered, and human operational errors are classified into two separate categories; location errors and orientation errors.

A location error means that a human material handler made a mistake to pick or put a part on a wrong resource location. A human may commit a location error during his or her material handling task by loading or unloading a specific part on some equipment (resources) which are not in the proper process plans for the part. An orientation error is the case of not properly placing a part on equipment. For example, a human may commit an orientation error when he or she places and fixes the part on the controllable vice. Even if the human operator places the part on the right equipment (location), he or she may make an orientation error because of placement of the part in wrong directions and fixation of the part improperly.

Location and orientation errors may hinder the system from starting a proper operation in processes, and this failure causes the system to stop and wait for a recovery action. Every part in system operations is represented by its own unique state that is specified in a part-state graph with electronic sensors that can check the physical precondition of a system operation a, $\rho_a \in \boldsymbol{P}_E$. Therefore, location and orientation errors are checked by sensors installed on equipment before the system starts a process.

4.2.2 Error classification in affordance-based MPSG systems

The location and orientation errors stated in the previous section 4.2.1 are taxonomies under physical preconditions regarding coordination states of a part in systems. However, in the ecological definition of affordances, properties of affordances, effectivities, and possible human actions in systems should have ontological assumptions related with space and time (Turvey, 1992), and the failure to satisfy these assumptions can lead a system state to undesirable states or make an improper transition. The cases of failing to satisfy assumptions in space dimension fall into the category of location and orientation errors. The cases of failing to satisfy assumptions in time domains, however, were not investigated.

In the perspectives of control models, actual system status and behaviors should coincide with representation of states and events within the same time and space domains. The detection of location and orientation errors can be easily performed by using sensors installed on resources (equipment) in control systems, while the detection of failing to satisfy time constraints cannot be considered in the existing extended MPSG control systems. Specifically, the automated equipment in systems run based on the logical preconditions within systems, but a human in the system tends to take an action relying on his or her perception-based actions which are available within a specific space dimension and time duration containing the affordance-effectivity duals for those actions. For this reason, one additional error classification for a human needs to be considered; a set of transition errors with respect to time and space constraints between a human and a part. A human may commit transition errors if he or she missed to perform a desired task within a specific time range.

An example of transition errors can be expressed in affordance-based MPSGs as shown in Figure 13. The errors can be detected and checked when a specific human action is not taking within the time and space conditions described in a set of action conditions, *C*. The action conditions can be estimated based on the information of the relative properties between a human material handler and a part, such as size and weight of the part, lifting and moving capabilities of the human material handler, relative distance between the part and human. The size and weight of a part can be detected by sensors installed on equipment, the human capabilities are pre-programmed based on the personal information, and the location, viewing, and moving direction of the human material handler can be detected by a vision sensor installed in the shop floor system. The representation of affordance-based MPSG systems contains time-related tuples which can measure and check the time constraints for existence of possible human actions. The time advances are checked within control programs for equipment and the system allows a human material handler to perform human tasks only within a specific time range.

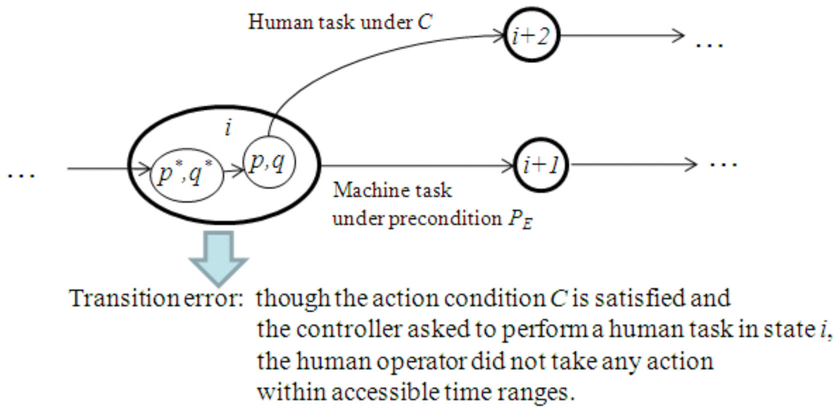


Fig. 13. Examples of location and orientation errors in affordance-based MPSG systems.

4.2.3 Error recovery

The detection and classification of human errors in human-machine cooperative systems are crucial to validate the control processes of human-involved systems. The analysis of error status in systems can guarantee the prompt and proper recovery of the systems from undesirable system states.

When location and orientation errors are occurred, the system will stop and wait for recovery action by either incurring automatic recovery module or calling human material handlers. In case of a transition error, the system can simply recover it by re-allocating the human task to a machine without stopping and recalling an error recovery module as shown in Figure 14, if the desired task can be performed by either a human or a machine. If a desired human task is failed to be performed within an eligible time range, machine can take an action instead of a human. However, if the required task for a specific system transition can be done only by a human, it should be recovered by human operators. The transition error recovery process is described as shown in Figure 15.

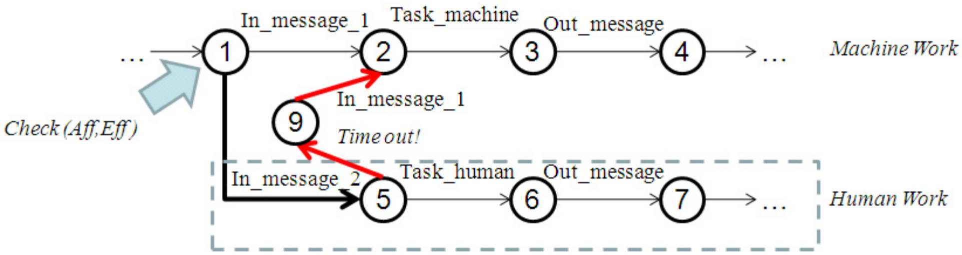


Fig. 14. Recovery of a transition error by re-allocating a human task to machine.

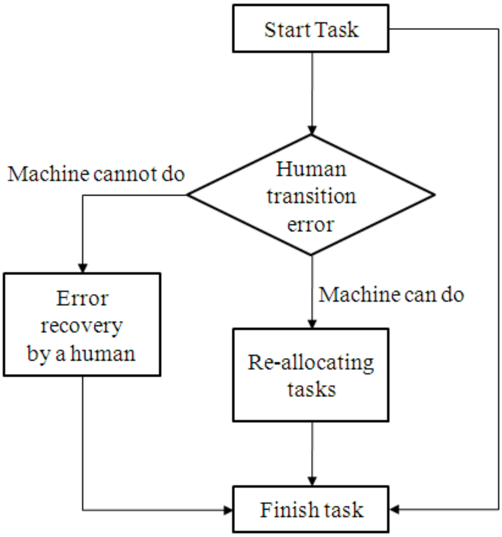


Fig. 15. Human transition error recovery procedure.

5. Summary

This chapter presents the modeling concept and formal representation of human-involved manufacturing control systems called affordance-based MPSG. With consideration of affordances in manufacturing systems, a human can participate in system operations and dynamic task allocation between a human and a machine is available.

Investigation of errors and their classification are also discussed. In regard to human transition errors, the automatic task reallocation to machine is a solution to solve the errors in easy ways. However, if the original task for a specific transition was only available for a human operator, an error recovery task by a human should be incurred to solve it.

6. References

Altuntas, B.; Wysk, R. A. & Rothrock, L. (2007). Formal approach to include a human material handler in a computer-integrated manufacturing (CIM) system. *International Journal of Production Research*, Vol. 45, No. 9, pp. 1953-1971, ISSN 0020-7543

Brann, D. B.; Thurman, D. A. & Mitchell, C. M. (1996). *Human interaction with lights-out automation: a field study*, IEEE, ISBN 0-8186-7493-8, Dayton, OH, USA

Cassandra, C. G. & Lafortune, S. (1999). *Introduction to Discrete Event Systems*, Springer, ISBN 0792386094, Boston, MA

Fitts, P. M.; Viteles, M. S.; Barr, N. L.; Brimhall, D. R.; Finch, G.; Gardner, E.; Grether, W. F.; Kellum; W. E. & Stevens, S. S. (1951). *Human engineering for an effective air navigation and traffic control system*, Ohio state University Fundation Report, Columbus, OH

Gibson, J. J. (1979). *The ecological approach to visual perception*, Psychology Press, ISBN 0898599598, Boston, Houghton Mifflin.

Hopcroft, J. E. (2001). *Introduction to automata theory, languages, and computation*, Addison-Wesley, ISBN 0201441241, Boston, MA

Joshi, S. B.; Mettala, E. G.; Smith, J. S. & Wysk, R. A. (1995) Formal models for control of flexible manufacturing cells: physical and system model. *IEEE Transactions on Robotics and Automation*, Vol. 11, No. 4, pp. 558-570, ISSN 1042-296X

Kim, N.; Shin, D. ; Wysk, R. A. & Rothrock, L. (2010) Using finite state automata (FSA) for formal modelling of affordances in human-machine cooperative manufacturing systems. *International Journal of Production Research*, Vol. 48, No. 5, pp. 1303-1320, ISSN 0020-7543

Shin, D.; Wysk, R.A. & Rothrock, L. (2006a). Formal Model of Human Material-Handling Tasks for Control of Manufacturing Systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, Vol. 36., No. 4, pp. 685-696, ISSN 1083-4427

Shin, D.; Wysk, R. A. & Rothrock, L. (2006b). A formal control-theoretic model of a human-automation interactive manufacturing system control, *International Journal of Production Research*, Vol. 44, No. 20, pp. 4273-4295, ISSN 0020-7543

Shin, D.; Wysk, R. A. & Rothrock, L. (2006c). An investigation of a human material handler on part flow in automated manufacturing systems, *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, Vol. 36, No. 1, pp. 123-135, ISSN 1083-4427

Smith, J. S.; Joshi, S. B. & Qiu, R. G. (2003). Message-based Part State Graphs (MPSG): A formal model for shop-floor control implementation, *International Journal of Production Research,* Vol. 41, No. 8, pp. 1739-1764, ISSN 0020-7543

Turvey, M. T. (1992). Affordance and Prospective Control: An Outline of the Ontology. *Ecological Psychology*, Vol. 4, No. 3, pp. 173-187, ISSN 1040-7413

Zeigler, B. P. (1976). *Theory of Modeling and Simulation*, John Wiley & Sons, ISBN 0-12-778455-1, New York, NY

7

Towards Adaptive Manufacturing Systems - Knowledge and Knowledge Management Systems

Minna Lanz, Eeva Jarvenpaa, Fernando Garcia,
Pasi Luostarinen and Reijo Tuokko
Tampere University of Technology
Finland

1. Introduction

Today the changes in the environment, be those business related or manufacturing, are both frequent and rapid. Industry has talked about the adaptation to meet the changes over a decade. Adaptation as a word has gained quite a reputation. Adaptation is expected in design of products and processes and in the realization of processes. The adaptation in the field of manufacturing sector is commonly understood as operational flexibility and reaction speed to the changes and/or opportunities. However, in order to achieve the required level of adaptability a company must be able to learn. Learning is achieved through gaining and understanding feedback of a change: its quantity and direction. Gaining and understanding the feedback a company must be able to compare the past status to the new status of actions. Unfortunately, the knowledge of neither the past nor the present is in computer interpretable and comparable form. Thus, the achieved and/or imagined flexibility is slightly above non-existent in reality.

This chapter discusses the possibilities of a modular and more transparent knowledge[1] management concept that provides means for representing and capturing needed information as feasible as possible while understanding that it is also the software systems that need to adapt to the changes along the physical production systems. The research approach discussed here aims to introduce new ideas for the companies knowledge management and process control by facilitating the move from technology based solutions to configurable systems and processes where the digital models and modular knowledge management systems can be configured based on needs - not based on closed legacy systems. The case implementation chosen here to illustrate this approach divides the knowledge management system into three separate layers: data storing system, semantic operation logic (the knowledge representation) and services that utilize the commonly available knowledge. The modular approach in

[1] In the knowledge management literature, three levels of knowledge - data, information and knowledge - are commonly distinguished. Awad and Ghaziri (2004) define data as unstructured facts, which in IT terms are usually considered as just raw bits, bytes, or characters. Information is structured data and attributes which can be communicated, but which may only have meaning locked inside proprietary software. Knowledge is seen as information that has meaning for more than just one actor and it can be used to achieve results.

ICT allows also software vendors to enhance their production to be more modular and configurable thus allowing the service oriented operation model to be realized. Once the storing method is separated from the logic and services, the new concepts can emerge. It is seen also that the vendors can make new business opportunities based on modular system solutions and configuration of those instead of highly tailored solutions which cannot be re-used later on.

The chapter is structured as following: the section 2 will illustrate the challenges industrial world is facing today. Section 3 summarizes the state of the art in field of knowledge modeling. Section 4 outlines needs for modularity in systems and introduces one possible solution candidate. Section 5 introduces a case implementation. Section 6 concludes the chapter and section 7 discusses about the challenges and future trends.

2. Set of challenges for the new decade

2.1 From simple to complex operation environment

For society to sustain and prosper, it needs along with societal, structural and organizational values a steady flow of income. For most of societies manufacturing has been and still is one of the biggest source of income. However, global competition has changed the nature of European manufacturing paradigms in past decades, see Figure 1. A turbulent production environment, short product life-cycles, and frequent introduction of new products require more adaptive systems that can rapidly respond to required changes whether or not the changes are based on product design changes or changes in the production itself. However, the technological leap in the mid 20th century, provided the means to venture towards more capable systems with very highly performing components. Today the acute problem is to take full advantage of their specific capabilities. These new systems, called complex systems, are no longer reducible to simple systems like complicated ones described by Descartes, Cotsaftis (2009).

Technical developments in recent years have produced stand-alone systems where high performance is routinely reached. This solid background has allowed the extension of these systems into networks of components, which are combined from very heterogenous elements, each in charge of only a part of the holistic action of the system. As the systems are process oriented instead of knowledge oriented systems, the interaction between tasks cannot be modeled, thus the effect of single interactions and relationships cannot be represented in the full systems scale. The types of interactions are changing into a complex network of possibilities within certain limits instead of a steady and predefined process flow. This situation is relatively new and causes pressures to define the role of intended interaction. According to Chavalarias et al (2006), there is no doubt that one of the main characteristics of complex and adaptive production platforms in the future will be the ever- increasing utilization of ICT. However, while the industrial world has seen the possible advantages, the implementations fall short as a result of the required changes to the whole production paradigm, going from preplanned hierarchical systems to adaptive and self-organizing complex systems, Chavalarias et al (2006) and Cotsaftis (2009).

Chavalarias et al (2006) stated that complex systems are described as the new scientific frontier which has been advancing in the past decades with the advance of modern technology and the increasing interest towards natural systems' behavior. The main idea of the science in complex systems is to develop through a constant process of reconstructing models from constantly improving data. The characteristics of a multiple-component systems is to evolve and adapt due to internal and external dynamic interactions. The system keeps becoming a different system. Simultaneously, the connection between the system and its surroundings evolves as well. When multiple-component system is manipulated it reacts via feedback, with the manipulator and complex system inevitably becoming entangled.

Paradigm	Craft Production	Mass Production	Flexible Production	Mass Customization and Personalization	Open Complex and Adaptive Production Systems
Paradigm started	~1850	1913	~1980	2000	2020
Society needs	Customized Products	Low cost products	Variety of products	Customized products	Customized on-demand products
Market	Very samll volume per product	Steady demand	Smaller volume per product	Global manufacturing and fluctuating demand	Global manufacturing and fluctuating demand
Business Model	Pull Sell-design-make-assemble	Push design-make-assemble-sell	Push-Pull design-make-sell-assemble	Pull design-sell-make-assemble	Pull design-sell-make-assemble
Technology Enabler	Electricity	Interchangeable parts	Computers	Information technology	Information and communication technology
Process Enabler	Machine tools	Moving assembly line	Flexible Manufacturing Systems, robots	Reconfigurable Manufacturing System	Self-organizing agents

Fig. 1. Paradigm shift, adapted and modified from ManuFuture Roadmap published by European Commission (2003)

In complex systems, reconstruction is searching for a model that can be programmed as a computer simulation that reproduces the observed data 'well'. The ideal of predicting the multi-level dynamics of complex systems can only be done in terms of probability distributions, i.e. under non- deterministic formalisms. An important challenge is, contrary to classical systems studies, the great difficulty in predicting the future behavior from the initial state as by their possible interactions between system components is shielding their specific individual features. In this sense, reconstruction is the inverse problem of simulation. This naturally indicates that the complex system cannot be understood as deterministic system, since the predictions from Complex Systems Science do not say what will happen, but what can happen, Valckenaers et al (1994), Chavalarias et al (2006), Cotsaftis (2009) and Lanz (2010).

In general, complex systems have many autonomous units (holons, agents, actors, individuals) with adaptive capabilities (evolution, learning, etc), and show important emergent phenomena that cannot be derived in any simple way from knowledge of their components alone. Yet one of the greatest challenges in building a science of such systems is precisely to understand this link - how micro level properties determine or at least influence properties on the macro level. The current lack of understanding presents a huge obstacle in designing systems with specified behavior regarding interactions and adaptive features, so as to achieve a targeted behavior from the whole, Chavalarias et al (2006).

Due to the complexity of the system behavior and the lack of tangible and implementable research results on how complex systems theory can bring revenue to a company; implementations at the moment are scarce and acceptance varies. In order to meet the new requirements set by the evolving environment several new manufacturing paradigms have been introduced, which follow characteristics of natural systems. These paradigms are:

- Bionic Manufacturing System (BMS): The BMS investigates biological systems and proposes concepts for future manufacturing systems. A biological system includes autonomous and spontaneous behavior and social harmony within hierarchically ordered relationships. Cells as an example are basic units, which comprises all other parts of a biological system and can have different capabilities from each other, and are capable of multiple operations. In such structures, each layer in the hierarchy supports and is supported by the adjacent layers. The components, including the part, communicate and inform each other of the decisions, Tharumarajah et al. (1996) and Ueda et al. (1997).
- Fractal Factory (FF): The concept of a fractal factory proposes a manufacturing company composed of small components or fractal entities. These entities can be described by specific internal features of the fractals. The first feature is self-organization that implies freedom for the fractals in organizing and executing tasks. The fractal components can choose their own methods of problem solving including self-optimization that takes care of process improvements. The second feature is dynamics where the fractals can adapt to influences from the environment without a formal organization structure. The third feature is self-similarity understood as similarity of goals among the fractals to conform the objectives in each unit Tharumarajah et al. (1996) .
- Holonic Manufacturing System (HMS): The core of HMS is derived from the principles behind the term 'holon'. The term holon means something that is at the same time a whole and a part of some greater whole Koestler (1968). The model of integrated manufacturing systems consists of manufacturing system entities and related domains, the structure of individual manufacturing entities, and the structuring levels of the entities. A manufacturing system is, at the same time, part of a bigger system and a system consisting of subsystems. Each of the entities posses self-description and capability for self-organization and communication, Valckenaers et al (1994) and Salminen et al (2009).

2.2 The meaning of knowledge

It is said that the world is surrounded by knowledge. Knowledge is saved into knowledge-bases and managed by knowledge management systems is something what has been stated over and over again. However, today, no matter what the vendor flyers express with colorful pictures and highly illustrative arrows, knowledge - as computers can understand it and reason with it - is not saved. The majority of the research and design effort is never captured or re-used. The interpretation of, for example a technical drawing is entirely based on the human perception and this perception may vary. ***"The meaning of knowledge is not captured and therefore not utilized as it has been intended."***

The need today is the capability for rapid adaptation to the changes in environment based on the previously acquired knowledge. However, the challenge is precisely the input knowledge or to be more accurate: the lack of it. In a large-scale company there can be up to hundreds of different design support systems, versions, and ad-hoc applications, which are used to create the information of the current product, process, and/or production systems. The majority of systems are using proprietary data structures and vaguely described semantics. This leads to challenges in information sharing since none of those are truly able to share data beyond geometrical visualizations. The design knowledge - the design intention - if even created, remains locked inside the authoring system, Ray (2004), Lanz (2010), Jarvenpaa et al. (2010), Lohse (2006) and Iria (2009).

3. The state of the art

3.1 Data modeling

As product, process and manufacturing system design have become more and more knowledge-intensive and collaborative, the need for computational frameworks to support much needed interoperability is critical. Academia and industrial world together have provided multiple different standards for product, process and resource models ranging from conceptual models to very formal representations. However, there are some serious shortcomings in the current representations:

- Firstly, none of these can represent the needs of the industry, not even industrial sector as whole.
- Secondly these standards do not form a knowledge architecture due to the missing critical parts (such as life-cycle information of products, processes and factory systems, history of past events and occurances).
- Thirdly, there does not exist a study that would outline the overlapping between these standards, Lanz et al. (2010).

Table 1 summarizes several different languages to represent data models that exist today. The list is not complete, nor it is intend to be, but it will summarize examples of standards, de facto standards and other models that are used today by industry and academia.

There have been three main approaches used to create a knowledge exchange infrastructure. They are a "point-to-point" customized solution, where dedicated interfaces are created between the design tools; a "one size fits all" solution decided by the original equipment manufacturer (OEM)'s proprietary interface for design and planning and knowledge exchange between parties; and the third solution is the a "neutral and open reference architecture" based on published standards. The first approach is expensive and time-consuming for the OEM, while the second option is very cost-efficient for the OEM, but expensive for partners who are working with several OEMs. The third option has never been fully implemented, Ray (2004), Lanz (2010) and Lohse (2006).

3.2 Knowledge capture

Second large problem area is the knowledge capturing. Currently there are very few systems that can be called knowledge capturing systems. By the definition information becomes knowledge, once other parties exist, which can understand the meaning of the information and can use it for their own purposes. In large scale organizations, data regarding activities and tasks are routinely stored in an unstructured manner, in the form of images and natural language used in e-mails, word-processed documents, spreadsheets and presentations. Over

Language/standard	Used in projects and standards	Description and use
EXPRESS	Standard for the Exchange of Product model data (ISO 10303 STEP), Open Assembly Model (OAM), Core Product Model (CPM), Krima et al (2009)	Defining the connections between the artifacts
CommonCADS	EUPASS Ontology, Lohse (2006)	Definition of interdependencies between classes
Web Ontology Language, Description Logics (OWL DL)/ Resource Description Framework (RDF)	Core Ontology Lanz (2010), ontoSTEP, Krima et al (2009)	Definition of interdependencies between classes and artifacts
First Order Logic (FOL)	Core Ontology, Lanz (2010)	Definition of interdependencies between classes
Common Logic Interchange Format (CLIF)	Process Specification language (PSL)	Describing what actually happens when a process specification executes and for writing constraints on processes, Bock & Gruninger (2005).
eXtensive Mark-up Language (XML)	Core Manufacturing Simulation Data (CMSD)	Used for the exchange manufacturing resource data
Automation ML	Knowledge Integration Framework ROSETTA (2010)	Representation Language of entities ROSETTA (2010)
Pabadis Promise Product and Production Process Description Language (P5DL)	OWL based language in FP6 Pabadis'Promise (2006)	P5DL used for description of products (as STEP) with their commercial and control relevant data and their necessary control applications and description of manufacturing processes with their hosting resources and necessary control functions, FP6 Pabadis'Promise (2006).

Table 1. Means for representing domain knowledge

time, large unstructured data repositories are formed, which preserve valuable information for the organization, if this information can ever be found or used.

Thus, a challenging research issue is to consider how information and knowledge is spread across numerous sources, and how it can be captured and retrieved in an efficient manner. Unfortunately, traditional information retrieval (IR) techniques not only tend to underperform on the kinds of domain-specific queries that are typically issued against these unstructured repositories, but they are also often inadequate, Iria (2009). The capturing of knowledge should start already from the creation of knowledge, where the engineer knows the meaning of the models and documents he/she is creating. This meaning should be captured in a form of computer readable format, such as a formal ontology, for further use.

3.3 Knowledge and meaning

According to DoHS (2008) increasing trend can be found from ongoing research in different domain contexts on using emerging technologies such as ontologies, semantics and semantic web (Web 2.0), to support the collaboration and interoperability. In recent years there have been a lot of activities concerning the domain and upper ontologies for manufacturing. As a result for the FP6 EUPASS project Lohse (2006) defined the connection between processes and resources for modular assembly systems. FP6 Pabadis'Promise (2006) project resulted in a manufacturing ontology (P2 ontology) and, reference architecture focusing on factory floor control.

Borgo & Leit (2007) developed the ADACOR ontology for distributed holon-based manufacturing focusing on processes and system interaction descriptions. ADACOR was later extended with an upper ontology Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE). Research done in the FP6 IP-PiSA project resulted an ontology, called Core Ontology, for connecting product, process and system domains under one reference model, Lanz (2010). The main goals these approaches generally try to achieve are: improved overall access to domain knowledge and additional information. However, none of these developed ontologies fully consider the needs above their narrow domain, Ray (2004), Lanz (2010), Jarvenpaa et al. (2010).

Ray (2004) introduced a roadmap from common models of data to self-integrating systems. The table 2 shows the 4 levels of representation. The table shows the logical steps for reaching first the creating of meaningful models (as in computational sense) to achieving finally systems that can autonomously exchange knowledge and operate based on shared knowledge.

According to the guidelines envisioned by Ray (2004), Lanz (2010) developed a common knowledge representation (KR) and semantics, called as Core Ontology, that allowed different design tools to interoperate across the design domains. The structure of the KR was formed on the basis of the requirements set by the knowledge management and integration challenges between different design tools, and the requirements set by the dynamic and open production environment. The developed model formalized the knowledge representation between product, process, and system domains utilizing fractal systems theory as a guideline. The surrounding system, be it the design environment or adaptive production system, can focus on the reasoning at different levels of abstraction, while the KR remained neutral for these

LEVELS OF REPRESENTATION	DESCRIPTION OF CHARACTERISTICS
Common Models of Data	In the lowest level the current state of the art, where the XML-based standards are utilized with relative ease within the IT sector, but not fully utilized in more conservative industry sectors.
Explicit and Formal Semantics	The second step, formal semantics, offers the generation of standardized representation that is formal enough to be parsed with computers.
Self-describing Systems	The third step is self-describing systems, where the systems can provide formal descriptions of their content and interfaces. This requires a formal semantic definition language that is rigorous enough to support logical inference.
Self-integrating Systems	The fourth level that Ray (2004) proposes is self-integrating systems. These systems are intelligent enough to be able to ask others for a description of their interfaces and, on the basis of the information thus acquired, adjust their own interfaces to be able to exchange information.

Table 2. The evolution of representational power towards formal semantics, and the systems integration capabilities that could follow (Ray, 2004)

reasoning procedures, but force the saved information to be consistent across the models. This approach differs from the traditional approaches by the fact that these tools are all utilizing already existing information as well as contributing specific information to the same model from different perspectives. The main objective of the developed KR was to achieve level two of the knowledge roadmap illustrated in the table 2. Other similar models exist, which utilize the complex nature of production systems. Most of these approaches are in the field of autonomous systems and control science.

4. Towards modular knowledge architecture for the dynamic environment

4.1 Understanding the Life-cycles of systems

All systems have their own life-cycles. In an open and complex operation environment the life-cycles play a very important role. The life-cycles that products can have are the most well-known life-cycle phases. These are such as in design, approved, in manufacturing, obsolete and such. These life-cycle phases represent the status of the design information.

The resource units also have their own specific life-cycle phases. These life-cycle phases describe the essential part of individual system units. Fore example, in the case of manufacturing resources the payload of a robot, accuracy of the tool and joints or tolerances do change over the life-cycle of the machine. It may happen that the capabilities of a system decline when it proceeds along its life-cycle. An example could be the capability for manufacturing certain surface with tight tolerances is possible when the machine is relatively new, but once the operating hours exceed a certain level the capability to reach the needed tolerances is no longer possible. Another example is the combined capability of an advanced manufacturing center and its operator. The machine may have dormant capabilities to perform advanced operations, which can be obtained once the operator has achieved needed knowledge in this particular case. Now the combined system's capabilities have increased.

In the case of modular knowledge architecture, ICT has also its own life-cycle. It is accepted from the start that the business field may change. When the change happens the ICT architecture must also adapt to the change. The change can happen also in the technological side when new technologies replace old ones. This means that some of the services may become obsolete and new services need to be added. In order to keep the architecture maintainable one solution is to offer independent service modules that operate over one information model without direct integration to the underlying databases.

4.2 Layers of operation

One of the approaches divides the knowledge management system into three separate layers: databases, semantic operation logic (the knowledge representation) and services that utilize commonly available knowledge. The modular approach in ICT allows also the software vendors to enhance their production to be more modular and configurable thus allowing the service oriented operation model to be realized. Once the storing method is extracted from the logic and services, the new concepts can emerge. It is also seen that vendors can make new business strategies based on new modular system solutions and configuration of those instead of highly tailored solutions, which cannot be re-used later on.

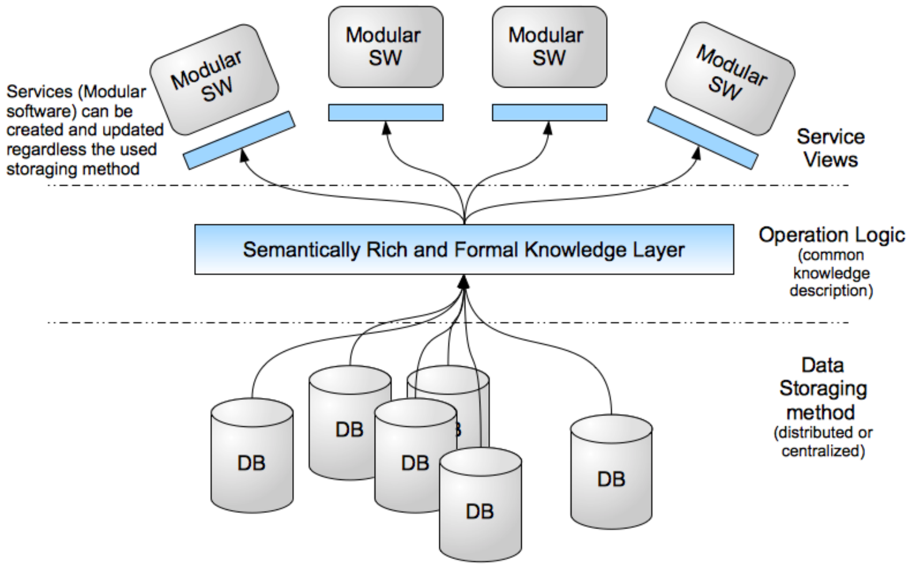


Fig. 2. Modular ICT

The ultimate goals in this particular research effort were to provide an information architecture, which allows different utilization of domain knowledge, while keeping the core information consistent and valid throughout the life-cycles of that particular set of information. The primary requirements that were defined together with industry are:

1. The model needs to represent the function of products and systems;
2. The model needs to connect different domains under one representation;
3. It must contain the history of changes applied to different instances;
4. The model must serve as an input source for automated information retrieval and reasoning in the traditional and in holon-based operation environment;
5. The model must be independent of the database implementation and services;
6. The model must allow as well as facilitate the generation of different services; and
7. The model must be extendable without disrupting the validity and consistency of the core domains.

5. Implementation of a modular ICT system

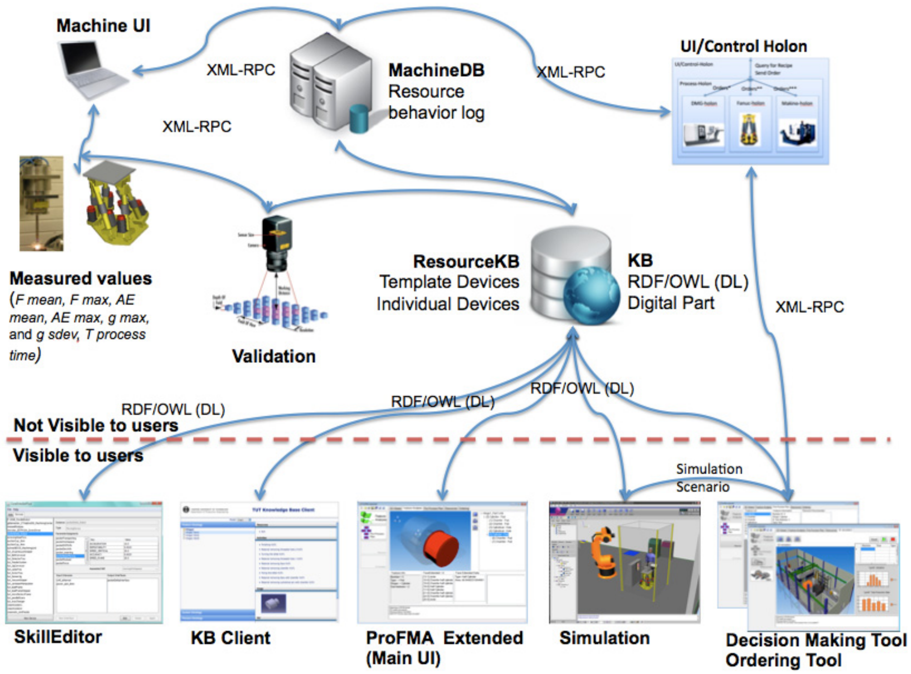


Fig. 3. implementation

The developed system, used here as an example, was based on the common knowledge representation and modular services would look as illustrated in figure 3. The clients contributing to the knowledge base are both commercial and university built existing systems and beta versions. Each of these tools requires specific domain related information and by processing the information they provide a set of services. However, the core of the system, the Knowledge Base (KB), needs to be extended to allow the capture and storing of semantically richer knowledge Lanz (2010) and Jarvenpaa et al. (2011).

The utilized knowledge representation (KR) can capture the meaning of classes via relationships that are defined between the classes. This technology allows semantic richness to be embedded into the model. Several service providers can use the meaning of stored information for their own specialized purposes. The model is divided into three separate layers as illustrated in figure 2. By dividing the data reserves, operation logic and services into separate layers connected with interfaces the upgrading of layers becomes independent of each others. This allows services to be extended, replaced and modified throughout their life-cycles.

In this case study the whole system architecture, illustrated in figure 3 has several different interoperating software modules each providing one or two essential functions for the whole holonic manufacturing system. The architecture is designed in such way that each of the modules can be replaced with a new module if needed. The connection of the modules is mainly based on the shared information model, the Core Ontology, described in detail in Lanz (2010), Lanz et al. (2011) and in Jarvenpaa et al. (2011).

The tools in the environment are designed by keeping the modularization principles in mind. Each of the tools are contributing their specific information to the common information model. The tools provide one or two main functionalities to the software environment. The modular design of the software allows changes to be applied to the tools with minimum disturbances. For example the holon user interface (UI), which controls the actual production can be replaced with a commercial tool that provides queueing functionality for the system.

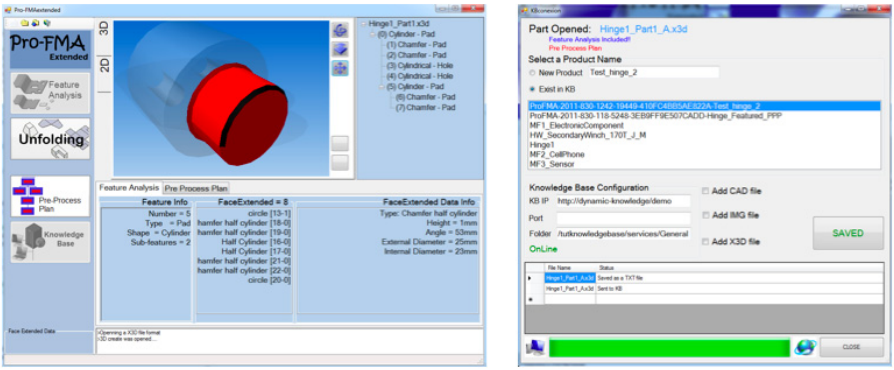


Fig. 4. Pro-FMA tool

The tools are:

Content creation: Pro-FMA illustrated in figure 4 is used to define the product requirements from the product model given in virtual reality modeling language (VRML) or eXtensive 3D (X3D) format. Product requirements are those product characteristics or features that require a set of processes for product to be assembled or manufactured. Features can be geometrical or non-geometrical by nature. These processes are executed by devices and combination of devices possessing adequate functional capabilities, Garcia et al. (2011).

Context creation: The Capability Editor, illustrated in figure 5, allows user to add devices to the ontology and assign them capabilities and capability parameters and enables creating associations between the capabilities. In other words it creates rules about which simple capabilities are needed to form combined capabilities, Jarvenpaa et al. (2011).

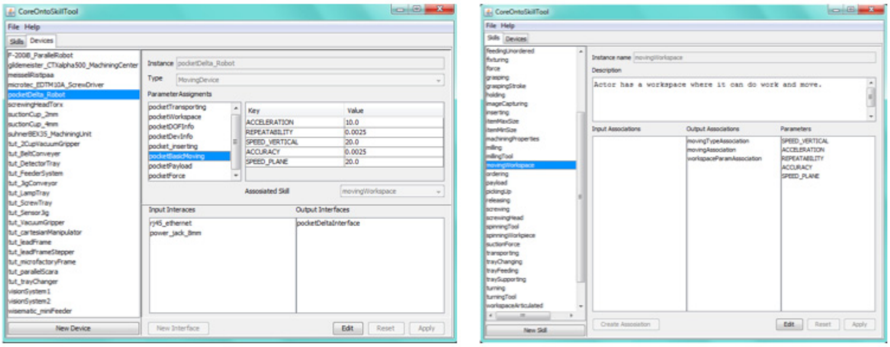


Fig. 5. Editor for Capabilities

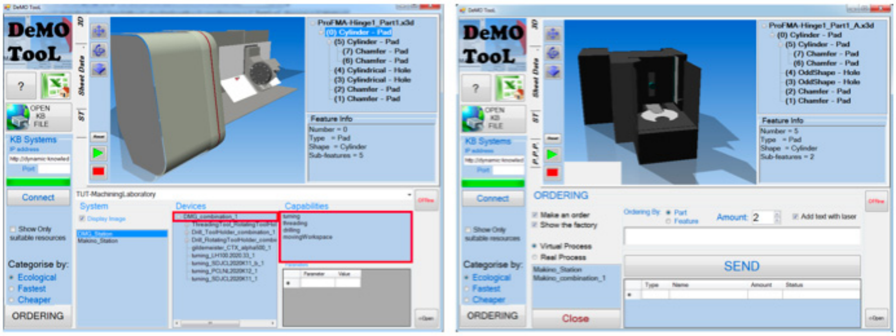


Fig. 6. Decision Making and Ordering Tool

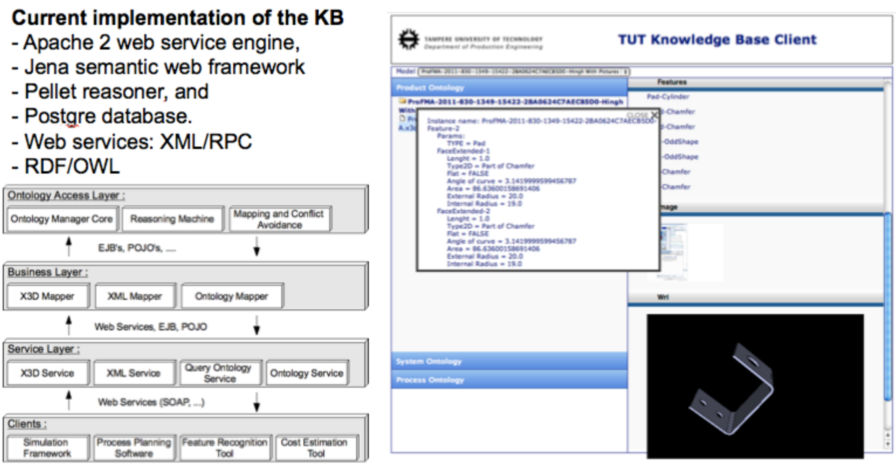


Fig. 7. Knowledge Base and Knowledge Base Web Client

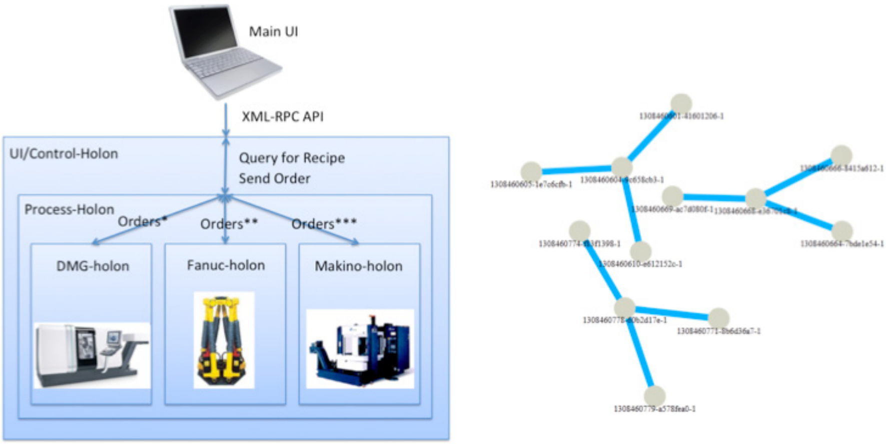


Fig. 8. Holonic control based on Kademlia, the right side of the figure shows the messages sent between holons

Verification: A simulation tool is used for creating the manufacturing or assembly scenarios. Since the environment is holonic by nature, it is accepted that the simulation only expresses possible solutions. The operation principle inside the simulation also follows holonic guidelines. This means that part or product is routed to the first available and capable cell.

Ordering: The Decision Making and Ordering Tool (DeMO tool), in figure 6, is used for setting up orders in this environment. The tool supports the viewing of the simulation as its minor function. The main function of the DeMO tool is to verify the connection to the factory floor and forward the orders to the holonic UI, Garcia et al. (2011).

Common Knowledge Representation: The KB and ResourceKB, shown in figure 7, store the information created by Pro-FMA, Capability Editor and DeMO tool. This system serves also as reference architecture, since it can handle closed models as references. The knowledge representation used in this case is based on OWL DL. The simulation model can be attached to product definition if needed. Similarly closed sub-programs and Computer-Aided-x (CAx) models can be associated with the part/product/resource description, Lanz (2010).

Content Verification: A web-based KB client, shown in figure 7 is used for human friendly information browsing. This tool serves as product data management (PDM) system's web-based user interface (UI). The client allows only limited set of changes to be applied tot he ontology. These changes are for example a new name for a product, part or other instance. For more details, please see Lanz (2010).

Operations Management: The process flow and distribution of tasks to each manufacturing or assembly cell is done with the Control Holon, see figure 8. The control holon observes the status of the system and available capabilities of system units (manufacturing resources in this case). The manufacturing resources can enter to and leave from the network without disturbing the whole system. This holonic control system distributes the tasks to suitable and available cells or stations based on the capability requirements defined by Pro-FMA earlier.

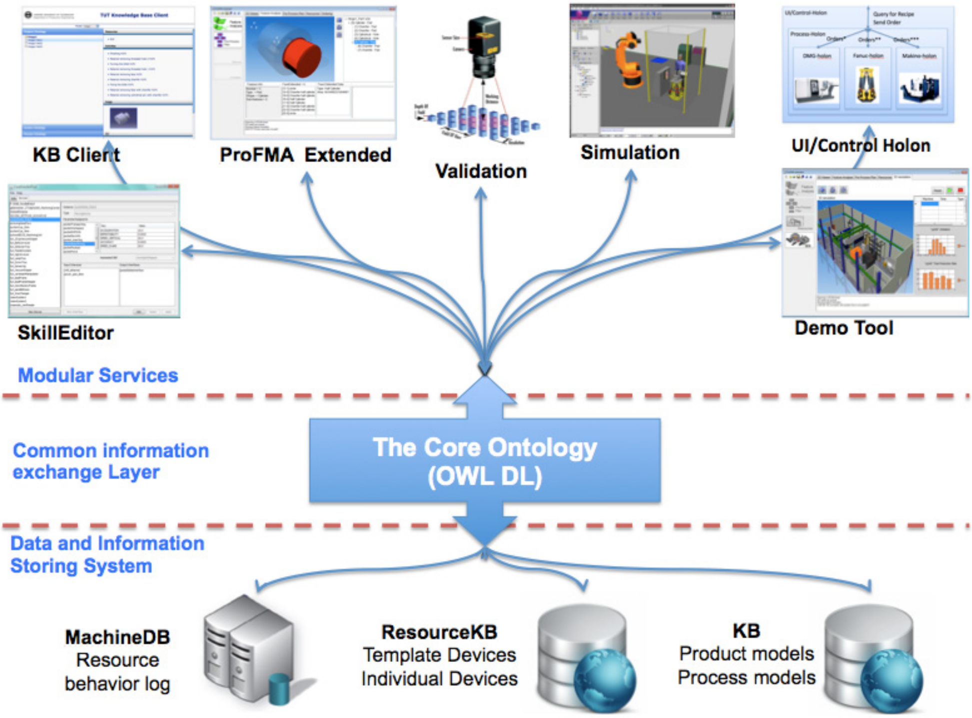


Fig. 9. The implementation is formed based on the modular ICT concept

The tools are divided into the layers described in previous chapter, see figure 2. The implemented environment, in figure 9, allows the addition of new services which can contribute and /or utilize already existing information, thus proving the modular ICT concept feasible for an adaptive, open and complex manufacturing environment. These tools constitute the necessary core for a modular system. There are additional services that could be added to this environment. These are traditional operation management module for production orchestration and machine vision based validation module. Both services are seen as extra for the core system.

6. Conclusion

Manufacturing after all is the backbone of each and every society, and in order for a society to be sustainable in long run the manufacturing has to be sustainable as well. From another point of view, manufacturing systems are shifting from being to becoming. This means that as the intelligence and cooperativeness advances the system will become a society where the rules, possibilities and constrains of a society as we know it will also apply. In order to achieve goals in the manufacturing society this research effort will contribute tremendous assets for securing the paradigm shift while keeping the manufacturing industry sustainable, flexible and adaptive. Without acceptance, further concept developments and implementation of the open and complex system approach the industry will not meet the challenges of the evolving environment.

It is seen that one partial solution is to develop these kind of modular ICT architectures that support the evolution of systems. However, it is understood that there is a lot of developments and solutions needed, since the industry cannot adopt partial solutions. Industry will require a concept that allows several data sources to be combined under one coherent and valid representation that facilitate the design and utilization of intelligent services in open and dynamic operations environment.

This paper introduced the context and operation principles of a dynamic system, and what is needed to support this kind of system from the knowledge management perspective. The article emphasized the challenge of dynamic systems from the life-cycle perspective as well, since all of the system parts be those software or hardware have their specific life-cycle phase. The division of architecture does provide tremendous possibilities for service development in future. As a proof of concept one type of modular ICT architecture and its core tools were introduced.

These results introduced here can be utilized in other fields than manufacturing engineering as well. The field of constructed environment and urban development has already seen the potential of an open world system where the input can be delivered in formal representation and services can be created independently of each other.

7. Discussion

When discussing about the holonic concepts with different people in seminars, workshops and conferences, a common comment/question has been: "Holonic manufacturing systems were developed 20-30 years ago and they didn't work then. How could they work now?" Shortly put, the answer could be technological and methodological development of knowledge and information management. Reasoning needed in the holonic systems relies on information and knowledge. Even though the concept of holonic manufacturing has remained similar throughout the years, information technology has made huge leaps enabling the implementation of these concepts in a feasible way. The novel methods to manage and distribute knowledge, such as semantic web and web service technologies, as well as semantic knowledge management systems, have been paving the way for the successful implementation of holonic systems.

Another question, which often arises in discussions has been: "Why holons? What advantages we gain by implementing holonic architecture? The implementation seems to be a huge task." Holons are autonomous and self-describing entities having well defined interfaces and the ability to communicate and co-operate with other holons. The modularity and self-organization ability enables the holonic systems to be extendable and adaptable. New holons, be they software system modules, new manufacturing resources or human workers, can enter and leave the system without disturbing the operation of the whole system. Each holon, module, knows its own purpose and the inputs and outputs, making the operation more transparent. In a holonic system it is possible to make changes in individual modules without the need to change/re-program the whole system. Until recently the holonic paradigm has only been implemented to physical devices and immediate control architecture of those. The design, operations management and supporting ICT systems have been ignored. However, as the ICT is expected to adapt to the changes in the production environment the holonic paradigm provides operation principles for this side as well.

Manufacturing is not the only domain, where the holonic paradigm could be applied. Actually, it could be applied almost anywhere, like in a medical and logistical domains. A good example can be found from city logistics. Cities, and their design, are not centrally controlled organized systems, but they are characterized by some level of chaos and the continuous threat of the chaos to expand to other operational areas. This chaos is controlled by hierarchical control systems where the control is coming from the top. From this viewpoint chaos is always considered as a negative element. This kind of systems need always be implemented as closed systems in order to prevent chaos.

The problem here is that innovations do not happen in order and harmony. The innovation always causes temporary chaos. Hierarchical control naturally strangles innovation. Therefore, what is needed is a control system where chaos is not a matter of crisis, but a normal event the system can handle in a flexible and efficient way. This kind of control system can be called as "chaordic system" (chaos + order). "Chaordic system" is self-organizing system which can always find a new equilibrium when the situation changes. The holonic control architecture can answer to the requirements of the "chaordic system". This idea has been presented to experts in the field of city logistics with very good feedback. The experts saw significant development potential for their business in holonic architecture in ICT and following the "open system" principles. However, all of this will be just theoretical discussion unless the surrounding ICT does support the change.

8. References

Awad, E.M. & Ghaziri, H.M. (2004).*Knowledge management*, Upper Saddle River, NJ, Pearson Education Inc.

Bock, C. & Gruninger, M. (2005). PSL: A Semantic Domain for Flow Models, *Journal of Software and Systems Modeling*, 4:2

Borgo, S. & Leit, P. (2007). Foundations for a core ontology of manufacturing, *Integrated Series in Information Systems*, vol 14

Chavalarias, D.; Cardelli, L.; Kasti, J. et al., (2006).Complex Systems: Challenges and opportunities, *an orientation paper for complex systems research in fp7*, European Commission

Cotsaftis, M. (2009). A passage to complex systems, in C*omplex Systems and Self-organization Modeling*, C. Bertelle, G. H. E. Duchamp, and H. Kadri-Dahmani, eds., Springer, 2009

European Commission (2003). *Working Document For The MANUFUTURE 2003 Conference*

Gruver, W. (2004). Technologies and Applications of Distributed Intelligent Systems, *IEEE MTT-Chapter Presentation*, Waterloo, Canada

Dept. of Homeland Security (DoHS), NAT, Cyber Security Division: Catalog of Control Systems Security, (2008). *Recommendations for Standards Developers*

FP6 Pabadis'Promise 2006. D3.1 Development of manufacturing ontology, project deliverable, The PABADIS'PROMISE consortium

Garcia, F.; Jarvenpaa, E.; Lanz, M. & Tuokko, R. (2011). Process Planning Based on Feature Recognition Method, *Proceedings of IEEE International Symposium on Assembly and Manufacturing (ISAM 2011)*, 25-27th of May, 2011, Tampere, Finland

Iria, J., (2009), Automating Knowledge Capture in the Aerospace Domain, *Proceedings of K-CAPÕ09*,Redondo Beach, California, USA

Jarvenpaa, E.; Lanz, M.; Mela, J. & Tuokko, R. (2010). Studying the Information Sources and Flows in a Company Ð Support for the Development of New Intelligent Systems. *Proceedings of the FAIM2010 Conference*, July 14-17, 2010 California, USA, 8 p.

Jarvenpaa, E.; Luostarinen, P.; Lanz, M.; Garcia, F. & Tuokko, R. (2011).Presenting capabilities of resources and resource combinations to support production system adaptation, *Proceedings of IEEE International Symposium on Assembly and Manufacturing (ISAM 2011)*, 25-27th of May, 2011, Tampere, Finland

Jarvenpaa, E.; Luostarinen, P.; Lanz, M.; Garcia, F. & Tuokko, R. (2011). Dynamic Operation Environment Ð Towards Intelligent Adaptive Production Systems, *Proceedings of IEEE International Symposium on Assembly and Manufacturing (ISAM 2011)*, 25-27th of May, 2011, Tampere, Finland

Koestler, A. (1968). *Ghost in the Machine*, Penguin, ISBN-13: 978-0140191929, p. 400

Krima, S.; Barbau, R.; Fiorentini, X.; Sudarsan, R. & Sriram, R.D., 2009, ontoSTEP: OWL-DL Ontology for STEP, *NIST Internal Report*, NISTIR 7561

Lanz, M.; Lanz, O.; Jarvenpaa, E. & Tuokko, R. (2010). D1.1 Standards Landscape, *KIPPcolla: internal project report*

Lanz, M. 2010. *Logical and Semantic Foundations of Knowledge Representation for Assembly and Manufacturing Processes*, PhD thesis, Tampere University of Technology

Lanz,M.; Rodriguez, R. & Tuokko,R. (2010) Neutral Interface for Assembly and Manufacturing Related Knowledge Exchange in Heterogeneous Design Environment, published as book, Svetan M. Ratchev (Ed.): *Precision Assembly Technologies and Systems*, 5th IFIP WG 5.5 International Precision Assembly Seminar, IPAS 2010, Chamonix, France, February 14-17, 2010. Proceedings. IFIP 315 Springer 2010, ISBN 978-3-642-11597-4, France

Lanz, M.; Jarvenpaa, E.; Luostarinen, P.; Tenhunen, A.; Tuokko, R. & Rodriguez, R. (2011). Formalising Connections between Products, Processes and Resources Ð Towards Semantic Knowledge Management System, *Proceedings of Swedish Production Symposium 2011*, Sweden

Lohse, N. (2006), *Towards an Ontology Framework for the Integrated Design of Modular Assembly Systems*, PhD thesis, University of Nottingham

ROSETTA Project Presentation May 2010. FP7 ROSETTA Project consortium, URL: http://www.fp7rosetta.org/public/ ROSETTA_Project_ Presentation_May_2010 _webpage.pdf

Ray, S. (2004). Tackling the semantic interoperability of modern manufacturing systems, *in Proceedings of the Second Semantic Technologies for eGov Conference*

Salminen, K.; Nylund, H. & Andersson,P. (2009). Role based Self-adaptation of a robot DiMS based on system intelligence approach, *proceedings of Flexible Automation and Intelligent Manufacturing (FAIM 2009)*, US

Tharumarajah, A.; Wells, A.J. & Nemes, L. (1996). A Comparison of the Bionic, Fractal and Holonic Manufacturing Concepts, *International Journal of Computer Integrated Manufacturing*, vol.9, no.3/1996

Ueda, K.; Vaario, J. & Ohkura, K. (1997). Modelling of Biological Manufacturing Systems for Dynamic Reconfiguration, *Annals of the CIRP*, 46/1: 343-346/1997

Valckenaers, P.; van Brussel, H.; Bongaerts, L. & Wyns, J. (1994). Holonic manufacturing execution systems, *CIRP Annals - Manufacturing Technology*, nro 54/1994

PLC-Based Implementation of Local Modular Supervisory Control for Manufacturing Systems

André B. Leal[1], Diogo L. L. da Cruz[1,3] and Marcelo da S. Hounsell[2]
[1]*Department of Electrical Engineering, Santa Catarina State University*
[2]*Department of Computer Science, Santa Catarina State University*
[3]*Pollux Automation*
Brazil

1. Introduction

Developing and implementing control logic for automated manufacturing systems is not a trivial task. Industrial production lines should be able to produce many types of products that go through a growing number of processes given the needs of the market, and there is an ever growing flexibility demand because of it. To keep up with it a faster way to develop control logic automation for the production lines is required. And this should be done in such a way to easy development and to guarantee that the control is correct in terms of making the system to behave as it should. To this end, the use of formal modelling tools seems to help raise the abstraction level of specifying systems' behaviour at the same time that it provides ways to test the resulting model.

The Supervisory Control Theory (SCT) of Ramadge and Wonham (1987, 1989) is an appropriate formal tool for the control logic synthesis of automated systems because it ensures the achievement of an optimal control logic (minimally restrictive and nonblocking), and that meets control specifications. Regardless of its advantages for automated manufacturing systems control and troubleshooting, this theory and its extensions have not been broadly used in industrial environments so far. The main reason for this resides in some difficulties that exist in dealing with complex problems. According to Fabian and Hellgren (1998) another important reason is the difficulty in implementing a pragmatic solution obtained from SCT theoretical result, i.e., bridging practice and theory.

This chapter presents a methodology, named DECON9, that aims to reduce the gap between this promising theory and real world applications, i. e., it presents a methodology for the implementation of the SCT into Programmable Logic Controllers (PLCs). The local modular approach (Queiroz & Cury, 2000a, 2000b) is used for the supervisors' synthesis and the implementation in PLC is performed in the ladder diagram language. Local Modular approach is used because systems of greater complexity that have a big amount of machines (and then, events) usually can be modelled as many concurrently interacting and simpler subsystems.

PLC implementation of supervisory control was also discussed in (Ariñez et al., 1993; Lauzon, 1995; Leduc & Wonham, 1995; Leduc, 1996; Qiu & Joshi, 1996; Lauzon et al. 1997; Fabian & Hellgren, 1998; Dietrich et al., 2002; Hellgren et al., 2002; Liu & Darabi, 2002; Music &

Matko, 2002; Queiroz & Cury, 2002; Chandra et al., 2003; Hasdemir et al., 2004; Manesis & Akantziotis, 2005; Vieira et al., 2006, Morgenstern & Schneider, 2007; Noorbakhsh & Afzalian 2007a&b; Afzalian et al., 2008; Hasdemir et al., 2008; Noorbakhsh, 2008; Silva et al., 2008; Leal et al., 2009; Possan & Leal, 2009; Uzam et al., 2009). In most of these works the monolithic approach (Ramadge & Wonham, 1989) for the supervisors' synthesis is used, in which a single and usually large supervisor is computed to control the entire plant. According to (Queiroz & Cury, 2002), this approach is not adequate for most real problems because they involve a large number of subsystems. In order to overcome this problem, in some works the synthesis of supervisors is performed according to the local modular approach (Queiroz & Cury, 2000a), which reduces the computational complexity of the synthesis process and the size of supervisors by exploiting specifications modularity and the decentralized structure of composite plants. Thus, instead of a monolithic supervisor for the entire plant, a modular supervisor is obtained for each specification, taking into account only the affected subsystems.

In almost all these work the implementation is held on ladder diagram, which is a well-known PLC programming language in industrial environments. But most existing proposals can only tackle one event per PLC scan cycle, which represents a problem when handling large scale plants (Vieira et al. 2006). Just a few of those proposals, at the best situation, can process one event per supervisor at each PLC scan cycle, a situation that can certainly be improved. Finally, just a few of them proposed solutions for the broad spectrum of problems that arise when implementing supervisory control in a PLC-based control system, as will be detailed later.

A contribution of the DECON9 methodology is that it allows dealing with various events at each PLC scan cycle, regardless if these events are controllable (can be disabled by control action) or not. Moreover, DECON9 provides a standardized approach and solution to many problems that arise while implementing SCT into PLCs.

The remaining of this chapter is organized as follows: In section 2, basic notations of the Supervisory Control Theory (SCT) for Discrete Event Systems (DES) control are introduced altogether with Monolithic and Local Modular approaches; Section 3 details the problems that arise while implementing SCT into a PLC; Section 4 presents the general assumptions behind the proposed methodology as well as its step-by-step detailed functioning; Section 5 presents a case study and how it can be solved by the methodology, and; Finally, section 6 concludes this chapter.

2. Supervisory control of discrete event systems

In the solution of manufacturing automation problems through the Supervisory Control Theory (SCT), the shop floor plant can be modelled as a Discrete Event System (DES) and finite-state automata are used to describe plant, specifications and supervisors. In this section, we introduce basic SCT notations. More details can be found in (Wonham, 2011) and in (Cassandras & Lafortune, 2008).

2.1 Discrete event systems

A Discrete Event System (DES) is a dynamic system that evolves in accordance with the abrupt occurrence of physical events at possibly unknown irregular intervals (Ramadge & Wonham, 1989). Application domains include manufacturing systems, traffic systems, software engineering, computer networks and communication systems, among others.

According to (Cassandras & Lafortune, 2008) and (Ramadge & Wonham, 1987, 1989) the free behaviour of a DES can be described through automata. An automaton can be represented by the 5-tuple $(Q,\Sigma,\delta,q_0,Q_m)$, where Q is the set of states, Σ is the alphabet of events, $\delta: Q\times\Sigma\rightarrow Q$ is the (partial) state transition function, q_0 is the initial state and $Q_m \subseteq Q$ is the set of marked states (Vieira et al., 2006). Σ^* is used to denote the set of all finite length sequences of events from Σ. A string (or trace) is an element of Σ^* and a language is a subset of Σ^*. A prefix of a string s is an initial subsequence of s, i.e. if r and s are strings in Σ^*, u is a prefix of s if $ur = s$. For a language L, the notation $\overline{L}$, called the prefix-closure of L, is the set of all prefixes of traces in L. L is said to be prefix-closed if $L = \overline{L}$ (Afzalian et al., 2010).

Consider that an automaton G represents the free behaviour of the physical system. Two languages can be associated with it: the closed behaviour $L(G)$ and the marked behaviour $L_m(G)$. The language $L(G)$ is the set of all sequence of events that can be generated by G, from the initial state to any state of G. Thus, $L(G)$ is prefix-closed because no event sequence in the plant can occur without its prefix occurring first. It is used to describe all possible behaviours of G. The language $L_m(G) \subseteq L(G)$ is the set of all sequence of events leading to marked states of G, each of them corresponding to a completed task of the physical system. A DES represented by G is said to be nonblocking if $\overline{L_m(G)} = L(G)$, i.e., if there is always a sequence of events which takes the plant from any reachable state to a marked state (Afzalian et al., 2010).

The concurrent behaviour of two or more DESs is captured by the synchronous composition of them. Thus, for two DES, G_1 and G_2, the synchronous composition is given by $G = G_1 \| G_2$. This expression can be generalized for any number of DES by $G = \|_{\forall i \in I} G_i$.

The automata can also be represented by transition graphs (see Figure 1), where the nodes represent the states and the arcs labelled with event names represent transitions. Usually, the initial state is identified by an ingoing arrow whereas a marked state is denoted by double circles.

2.2 Supervisory control theory

In the Ramadge & Wonham (1989) framework, the set of plant events Σ is partitioned into $\Sigma = \Sigma_c \dot{\cup} \Sigma_u$, two disjoint sets where Σ_c is the set of all controllable events and Σ_u is the set of all uncontrollable events. An event is considered to be controllable if its occurrence can be disabled by an external agent (named supervisor), otherwise it is considered uncontrollable. The necessary and sufficient conditions for the existence of supervisors are established in (Wonham, 2011).

A supervisor, denoted S, determines the set of events to be disabled upon each observed sequence of events. It is a map from the closed behaviour of G to a subset of events to be enabled $S : L(G) \rightarrow 2^{\Sigma}$. The controlled system is denoted by S/G (S controlling G) and is modelled by the automaton $G\|S$. The closed and the marked behaviour of the system under supervision are respectively represented by the following languages: $L(S/G) = L(S\|G)$ and $L_m(S/G) = L(S\|G) \cap L_m(G)$.

Further, S is said to be nonblocking if $L(S/G) = \overline{L_m(S/G)}$, i.e., if each generated trace of the controlled plant can be extended to be a marked trace of the controlled plant. Consider that a language $K \subseteq L_m(G)$ represents a control specification over the plant G. K is said to be controllable with respect to G (or simply controllable) if its prefix-closure $\overline{K}$ doesn't change under the occurrence of uncontrollable events in *G*. In other words, *K* is controllable if and only if $\overline{K}\Sigma_u \cap L(G) \subseteq \overline{K}$. Given a discrete event plant *G* and a desired nonempty specification language $K \subseteq L_m(G)$, there exists a nonblocking supervisor S such that $L_m(S/G) = K$ if and only if *K* is controllable with respect to *G* (Wonham, 2011).

However, if K is not controllable, the *supremal controllable sublanguage of K* with respect to G, denoted by $SupC(K,G)$, must be computed. In this case $L_m(S/G) = SupC(K,G)$ (Ramadge & Wonham, 1989).

In order to differentiate the controllability of events in the graph representation of automata, usually the state transitions due to controllable events are indicated by a short line drawn across the transitions (Chandra et al. 2003). Figure 1 represents an automaton, where the event *A* is controllable and the event *B* is uncontrollable.

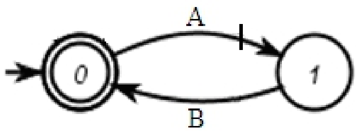


Fig. 1. A graph representation of an automaton

2.3 Monolithic approach

In the monolithic approach for the supervisors' synthesis (Ramadge & Wonham, 1989), the objective is to design a single supervisor that will coordinate the plant behaviour. Thus, all subsystems models G_i (where *i* is related to the number of subsystems), are composed in order to compute an automaton (generator) *G* that represents the free behaviour of the entire plant. In the same way, all control specifications E_j (where *j* is related to the number of control specifications) are composed into a global control specification *E*. From these models, one obtains the closed loop desired behaviour (known as target language) computing $K = G \| E$ and, consequently, obtaining a single supervisor *S* that marks the *supremal controllable sublanguage of K*, that is, $L_m(S) = L_m(S/G) = SupC(K,G)$.

According to (Queiroz & Cury, 2000a), in the monolithic approach the number of states of *G* grows exponentially with the number of subsystems. So this approach has the following drawbacks: the amount of computational effort when performing asynchronous composition of several automata, and; the use of supervisors with too many states in control platforms (usually a PLC) may generate extensive programs, where understanding, validation and maintenance will therefore, become difficult. In some cases the size of the program can render them unfit to be used in certain platform, either because of the storage or processing capacity.

In order to resolve these difficulties, Queiroz & Cury (2000a) propose using the local modular approach, as introduced in the next subsection.

2.4 Local modular approach

This approach is an extension to the monolithic approach and explores both the modularity of the plant and of the control specifications. It allows determining rather than a single and usually large supervisor, many local supervisors whose joint action guarantees the attendance of all the control specifications. Figure 2 illustrates the structure of local modular supervisory control for two supervisors.

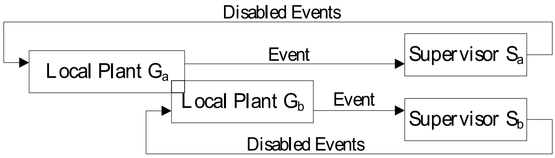


Fig. 2. Local modular supervisory control architecture (Queiroz & Cury, 2002)

In this approach, the physical system behaviour is modelled by a Product System (PS) representation (Ramadge & Wonham, 1989), *i.e.*, by a set of asynchronous automata $G_i = (Q_i, \Sigma_i, \delta_i, q_{0_i}, Q_{mi})$, with $i \in N = \{1,2,...,n\}$, all of them having disjoint alphabets Σ_i. In turn, each specification imposed by the designer is represented by an automaton Ej with an alphabet $\Sigma j \subseteq \Sigma$, $j \in \{1,..., m\}$, where m is the number of specifications. For each specification Ej a local plant G_{locj} is obtained, which is computed by the synchronous composition of all subsystems that share some events with the associated specification.

After determining all local plants it should be calculated the so-called local specification, which consists of performing synchronous composition between a given specification with its own local plant, *i.e.*, $K_{loc_j} = E_j \| G_{loc_j}$. Thus, the supremal controllable local sublanguages of K_{loc_j}, denoted by $SupC(K_{loc_j}, G_{loc_j})$, can be computed. Finally, it is possible to perform the synthesis of a local supervisor S_{loc_j} for each specification defined in the project. If at least one local supervisor disables the occurrence of an event, then the occurrence of this event is disabled in G (Vieira et al., 2006). To ensure that the system under the joint action of local supervisors is nonblocking, it should be guaranteed that the supervisors are nonconflicting, what is verified when the $\|_{j=1}^{m} \overline{L_m(S_{loc_j} / G_{loc_j})} = \overline{\|_{j=1}^{m} L_m(S_{loc_j} / G_{loc_j})}$ test holds. According to (Queiroz & Cury, 2000b), this condition ensures that the joint action of local supervisors is equivalent to the action of a monolithic supervisor that addresses all specifications simultaneously. Some computational tools can be used to assist in the synthesis of supervisors. For each one of them the models of subsystems and control specifications should be introduced in order to obtain synthesized supervisors, automatically. Among these tools *IDES* (Rudie, 2006), *TCT* (Feng & Wonham, 2006) and "*Grail for Supervisory Control*" (Reiser et al., 2006) can be mentioned.

3. Supervisory control implementation

3.1 Problems

According to (Fabian & Hellgren, 1998), "*the supervisor implementation is basically a matter of making the PLC behave as a state machine*". However, this is not trivial task and can lead to many problems (Fabian & Hellgren, 1998):

Causality: SCT assumes that all events are spontaneously generated by the plant and that supervisors should only disable events generated by the plant. However, controllable events on practical applications are not spontaneously generated by the physical plant, but as responses to given PLC commands. Thus, for implementation purposes, "who generates what?" must be answered.

Avalanche Effect: occurs when a change on the value on a given PLC input signal is registered as an event that makes the software jump over an arbitrary number of states within the same PLC scan cycle. This may occur particularly if a specific event is used to trigger many successive state transitions, thus producing an avalanche.

Simultaneity: Due to the cyclical nature of the PLC processing in which input signals readings are performed only at the beginning of each scan cycle, the occurrence of uncontrollable events from the plant is recognized by the PLC once there are changes in the input signals values. Therefore, if in between successive scan cycles two or more signals change, they will all be recognized as simultaneous uncontrollable events regardless of their exact timing. As a result, the PLC is unable to recognize the exact order of uncontrollable events that happen in between scan cycles.

Figure 3 shows an automaton that is sensitive to the sequence of *B* and *C* events. Notice that depending on the order *B* and *C* events happens, the control decision is different, which highlights the problem that if *B* and *C* are recognized altogether in the same scan cycle, we would not be able to determine the actual state and, what should be the control action: *E* or *F*.

In order to avoid the simultaneity problem, the system must present the "interleave insensitivity" property (Fabian & Hellgren, 1998), which requires that after any interleaved uncontrollable events the "control decision" must necessarily be the same.

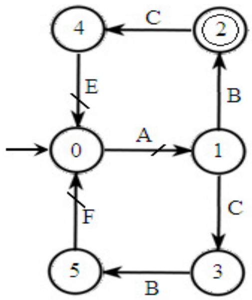


Fig. 3. Automaton that attempts to distinguish between the interleaving of events

Choice Problem: the supervisors obtained by the SCT are required to be "minimally restrictive", which means that the supervisors might provide alternative paths for the plant to choose from. Often a supervisor presents more than one possible controllable event from a single state. Thus, before producing a signal-change in the PLC outputs it may have to choose only one among them because according to Fabian & Hellgren (1998), generating more than one controllable event in a scan cycle can be contradictory and catastrophic.

Inexact Synchronization: during the program execution a change in any PLC input signal may occur and, this change will only be recognized at the beginning of the next scan cycle. The control reasoning is always performed on old frozen data. Therefore the communication between the PLC and the plant is subject to delays due to periodic reading of the input signals (Balemi, 1992). This inexact synchronization (Fabian & Hellgren, 1998) can be a problem when a change in a PLC input signal invalidates a control action (the choice made by the program, which corresponds to the generation of a controllable event).

3.2 Related work

Many researches have dealt with producing PCL programs from TCS. Some attempts (Fabian & Hellgren, 1998) did not propose a methodology but focused on solving particular situations which is far from a generic approach such as (Hasdemir et al., 2008). In the following, we briefly discuss some of these proposals.

In order to solve the choice problem, the solutions adopted in the literature follow the idea that for practical PLC implementation purposes, a deterministic controller must be statically extracted from the supervisor. Fabian & Hellgren (1998) also show that if a choice is not taken, the sequential execution of the program within the PLC will choose and the chosen transition will always be the same in a particular state according to the ordering of the rungs.

Moreover, Malik (2002) shows that depending on the choice taken (or deterministic controller extraction) the controlled behaviour may be blocking, even when the supervisor is nonblocking, which in that work is named *determinism problem*. In (Dietrich et al., 2002) three properties are given which, when satisfied, ensure that any controller for the system is necessarily nonblocking. In (Malik, 2002) a more general property is introduced and an algorithm is given which checks whether every deterministic controller generated from a given model is nonblocking. However, no controller can be constructed from those works in case the DES model does not satisfy these conditions. In particular, a valid controller may exist, even if the conditions of (Malik, 2002) and (Dietrich et al., 2002) do not hold. But, in (Morgenstern & Schneider, 2007) another approach to generate deterministic controllers from supervisors is presented and a property named *forceable nonblocking* is introduced.

In (Queiroz & Cury, 2002), the authors introduce a general control system structure based on a three level hierarchy that executes the modular supervisors' concurrent action and interfaces the theoretical model with the real system. They also propose a ladder-based PLC implementation of TCS but do not discuss the above mentioned problems.

In (Hasdemir et al., 2004) the authors propose the use of two bits for each state in order to solve the avalanche effect, but none of the other problems are discussed. In addition, only a single event is processed per supervisor at each PLC scan cycle.

Vieira (2007) presents a methodology that considers some of the problems but the program is structured as Sequential Function Charts (SFC), which is not so widespread among PLC programmers so far. Also, this methodology requires to change the automaton model in order to remove self-loops and there is no solution to the choice problem as well.

Most of the proposals found in the literature (Leduc, 1996; Hellgren et al., 2002; Queiroz & Cury, 2002) implements SCT in the ladder language. They have the same drawback: they deal with one single event per scan cycle. Thus, if between two scan cycles "*n*" changes occur in the PLC inputs, the program will take "*n*" scan cycles to deal with them. The best proposals so far handle one event per supervisor at each PLC scan cycle. This constraint help ensure existing approaches to deal with the avalanche effect and the choice problem (*determinism problem*). However, in this way the supervisor's update rate and actions will be lower than that obtained via traditional solution, without the use of the SCT.

The related work presented above shows that a ladder-based PLC complete methodology that solves recurring problems on implementing TCS supervisory control is still missing. Below we present a methodology that fills this gap.

4. DECON9 methodology

This section presents DECON9 (which comes from the main idea: DEcomposing the CONtrol Implementation DEpending on the CONtrolability of the events), a nine steps ladder-based PLC implementation of SCT supervisory control methodology that treats multiple events in the same scan cycle and also solves the avalanche effect and the choice problem. The methodology was inspired by the work of Queiroz & Cury (2002) but the Product System (PS, the asynchronous sub-plant models) and supervisors' implementations are decomposed into blocks of events according to their controllability.

At the beginning of each PLC scan cycle, all signal changes in the PLC inputs are registered as uncontrollable events in the PS level, and state transitions due these events are processed in the PS automata. Immediately after that, the supervisors also perform the state transitions due uncontrollable events. In this way the treatment of uncontrollable events are prioritized, and PS and supervisors are maintained in synchronization with the plant.

From the current state of the supervisors all events that are still disabled are verified through a disabling map. Thus, if at least one local supervisor disables a certain event, then the occurrence of this event is globally disabled. Thereafter, from the list of non-disabled events the choice problem (determinism problem) is inferred. If there is more than one enabled event in a current local supervisor state an event is randomly selected. In opposition to the other proposals in which a deterministic controller is statically extracted out of the supervisor (offline procedure) before being implemented, in our methodology the supervisors are fully implemented and the decisions on which controllable event may be executed are dynamically performed on the fly (during the program execution). So all alternative paths in the supervisor are preserved and the system behaviour under supervision is ensured to be nonblocking.

All enabled controllable events which are likely to occur in the plant are generated in the PS and the states are updated in the subsystems and supervisors models. Finally, these events are mapped to PLC outputs and another scan cycle begins.

Notice that the coherence of control actions is guaranteed because before defining the set of disabled events and generating the controllable events in the PS, the states of the subsystems and supervisors are all updated (this means that the supervisors know in which state they are in and which events are enabled).

4.1 Solving the avalanche effect: Damming

To avoid the avalanche effect the methodology indicates to use two auxiliary memories for each uncontrollable event: the first one is used to store the events generated by the plant and; the second is used to enable PS and supervisors to transit states. Every time the second memory is used, it is reset (deleted).

In this way, an event is used only once and multiple transitions are hold up. In any case the initial state of the event is required, the first memory can be used. As long as the PS is composed of asynchronous subsystems, once it is updated due to a given uncontrollable event, this information is not used any more until this same event is generated by the plant.

However, the information that a given uncontrollable event is active can be used somewhere in de program, especially to update supervisors states, once an uncontrollable event can be part of many supervisors simultaneously. Therefore, it should be possible to recover all information regarding uncontrollable events generated by the plant before updating supervisors. That's the reason for the second auxiliary memory.

4.2 Solving the choice problem: Random choice

To deal with the choice problem, one should analyse each supervisor at a time, to look after states where the problem may occur and which events are involved in each case. The simpler case is when there are only two controllable events to choose from, but situations involving a handful of choices are not rare in real applications. Figure 4 presents a flowchart for dealing with the choice problem. It helps identifying corresponding states and events.

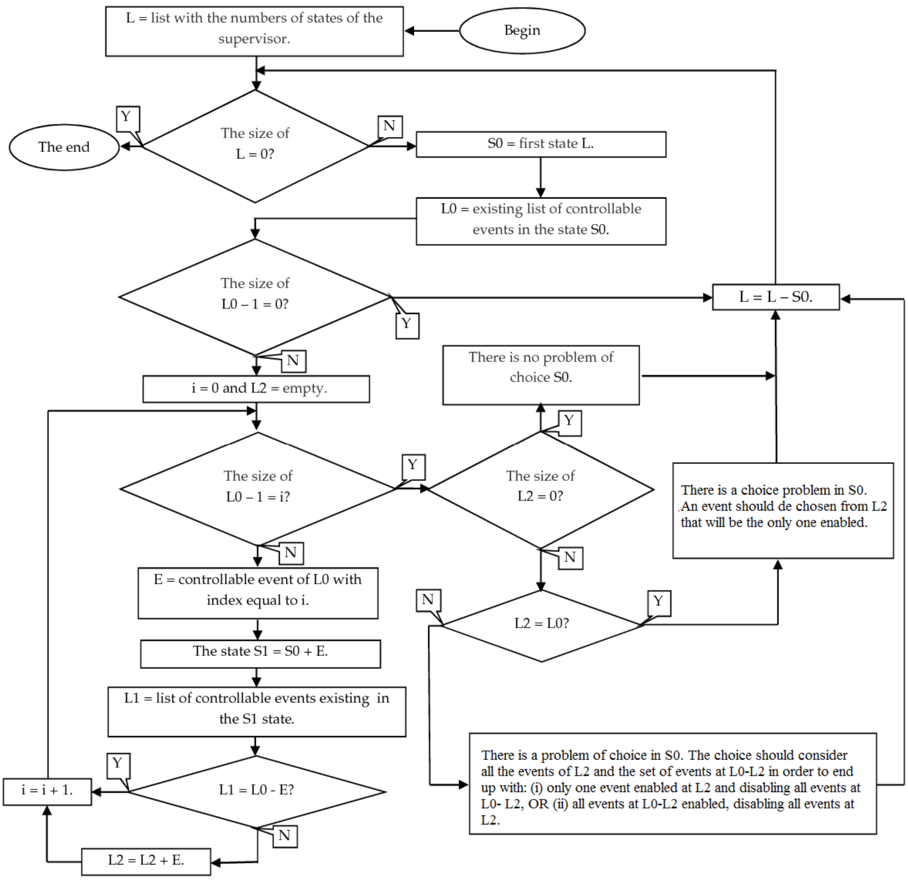


Fig. 4. A flowchart to identify a possible choice problem

To solve the choice problem, states that present multiple control alternatives need to be identified. But, if just after the disabling routine more than one controllable event is enabled at an active state, a routine is called that performs a random selection between pairs of these events. Randomness happens because an auxiliary memory is used to help perform the selection. This memory changes state at every scan cycle and, because there is no deterministic way to predict the number of scan cycles before the choice problem occurs, it acts as if it is random.

4.3 Solving simultaneity and inexact synchronization: Hardware interruption

For the simultaneity and inexact synchronization problems the solution adopted is the use of hardware interruption. Then, the uncontrollable events that may cause these problems must be associated with that type of PLC input. Thus, when a change in one of these PLC inputs occurs, the program is interrupted and the event is registered at the moment of its occurrence. It is important to notice that many PLCs do not have this kind of input and, in that case, there is no particular solution in DECON9.

It should be pointed out that these are not problems that arise exclusively while implementing SCTs into PLCs; they could happen in any given conventional approach that did not even use SCTs. Nevertheless, it can be said that an advantage of using SCTs is that these problems can be identified and we could be aware of them at the very beginning stage of designing the control system. On conventional approaches however, they can only be identified, if ever, after an extensive trial-and-error validation experiments.

On the other hand, not all systems' models will present this kind of problem. Thus, for the local modular supervisory control structure to be implemented without such problems the model must abide to some properties:

- To be sure that no problem regarding the "inability to identify event's order" problem will happen, it should be ensured that all automata that model every supervisor show the "interleave insensitivity" property (Fabian & Hellgren, 1998);
- Complementary, to be sure that no problem regarding "inexact synchronization" will happen, it should be ensured that the resulting language from every supervisor's automata, and their corresponding supervisors, show the "delay insensibility" (Balemi & Brunner, 1992).

5. DECON9 methodology step-by-step

This section will detail all nine steps of DECON9 methodology. DECON9 organizes the resulting program as a set of subroutine calls for the sake of better understanding, code reuse and reduction and, easy maintenance. Subroutine calls is a common feature available in almost every PLC.

Ten subroutines are created to fulfil all steps of DECON9 and they must follow a specific order. Figure 5 presents a complete flowchart where one can see all subroutine calls on the left and all nine steps on the right. It should be noticed that the third and fourth steps deal with uncontrollable events and the fifth to eighth steps deal with controllable ones. Also, "Make *Mx.0* = *Mx.1*" routine is called twice for every scan cycle.

First step is to be performed at the very first scan cycle only because it sets initial states of all auxiliary memories that store the initial state of all supervisors and PS subsystems. The remaining states are reset.

Second step reads all PLC inputs and identify events coming from the plant according to signal changes that corresponds to uncontrollable events.

Third step promotes the state transitions for the whole PS altogether with all just identified uncontrollable events. For this end, only transitions belonging to PS that involves uncontrollable events are transited. At this point it should be reminded that each event transition performed produces the information on the event to be erased in order to avoid the avalanche effect.

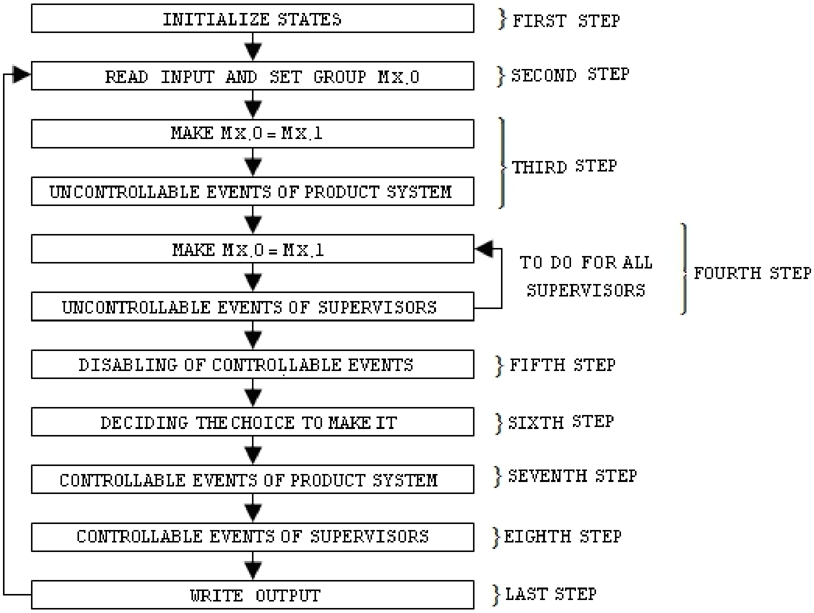


Fig. 5. Complete flowchart of the main routine

During the **fourth** step all supervisors must perform their state transitions considering uncontrollable events (and only these). The structure of the PLC program to be implemented for supervisors is the same as previously used for the PS system.

Because the information that a given event was active was erased during the transitions of the PS system, before updating the supervisors, the information of which uncontrollable events were generated by the plant must be recovered. For this end, all uncontrollable events use a pair of auxiliary memories: one of them to store which events were generated by the plant and the other is used for state transition and is discarded immediately after.

As a consequence, the methodology gives priority to uncontrollable events but do not neglect the synchronization of PS and supervisor states with the physical plant, therefore avoiding the avalanche effect.

The **fifth** step do not differ from what was proposed by Queiroz & Cury (2002) where, from the current state of each supervisor, the events disabled by any supervisor is disabled by the whole set of supervisors.

Sixth step starts off taking into account the status of all supervisors and a list of all still enabled controllable events. From these, if any supervisor shows the choice problem, it is resolved at this step. The program structure to solve this problem depends on the number of choices available (as presented earlier) but if no supervisor presents this problem, the sixth step does not produce any code.

As for the **seventh** step, it relates to the generation of controllable events that were not disabled beforehand and could possibly occur on the plant. This event generation is done at the PS level and is followed by the PS state transitions update due to the just generated events. Therefore, step seven is responsible for all state transitions related to PS´s controllable events and completing the implementation of the PS in the PLC.

Eighth step updates all controllable events generated in preceding step in all supervisors. Therefore, the remaining transitions not dealt with at the third step are done here completing the implementation of all supervisors in the PLC.

Thus, as a result of the last two steps, even before a physical control at the PLC output is issued due to controllable events, the PS and supervisors will be anticipating the state of the physical plant.

It should be noticed that there is also the possibility of the avalanche effect problem for controllable events. However, DECON9 establishes no particular procedure to deal with them because it is understood that in a practical application this problem will not occur. In any case, if it ever happens, it can be treated the same way it was done for uncontrollable events.

Ninth, and last step, sends controllable events generated by the control logic to the physical plant. This is done by mapping events to specific drives at the PLC outputs. This action generates new events from the physical plant and another scan cycle begins.

6. Case study

In order to illustrate DECON9, a complete solution for supervisory control problem is presented. The case study covers the plant and specifications modelling, the synthesis of supervisors, up to the coding of the control logic in ladder, ready to be implemented in a PLC.

6.1 Description of the physical system

The problem to be studied consists of a transfer line with six industrial machines M_X (where $X = 1,..., 6$) connected by four buffers B_Y (where $Y = A, B, C, D$), capable of storing only one part, as shown in Figure 6. This problem was studied by Queiroz & Cury (2000a) and was chosen because it produces simple automata, is easy to understand, is fairly possible to occur as part of bigger layouts and presents some of the problems DECON9 deals with.

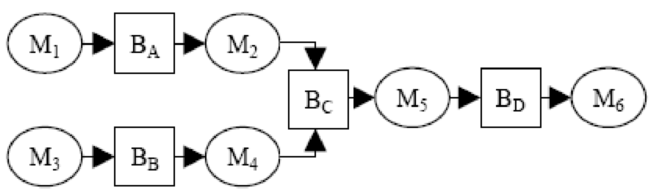


Fig. 6. Industrial transfer line case study

Start operation of the machines are controllable events, and the end operation are uncontrollable events. The transfer line should work in order to transport the parts through the machines, but a machine can't start operation if there is no part in its input buffer. Since the machines M_1 and M_3 have no input buffer, it should be considered that there will always be parts available for these machines. Similarly, M_6 can release as many parts as it is capable of producing.

6.2 Plant modelling

Table 1 presents a list of events associated with the operation of each machine as well as the type of event according to its controllability, the description of the event and, which PLC input (I) or output (Q) it is associated with.

DEVICE	EVENT	EVENT TYPE	DESCRIPTION	I/O
Machine 1	A_1	Controllable	Machine 1 start operation	Q0.0
	B_1	Uncontrollable	Machine 1 stop operation	I0.0
Machine 2	A_2	Controllable	Machine 2 start operation	Q0.1
	B_2	Uncontrollable	Machine 2 stop operation	I0.1
Machine 3	A_3	Controllable	Machine 3 start operation	Q0.2
	B_3	Uncontrollable	Machine 3 stop operation	I0.2
Machine 4	A_4	Controllable	Machine 4 start operation	Q0.3
	B_4	Uncontrollable	Machine 4 stop operation	I0.3
Machine 5	A_5	Controllable	Machine 5 start operation	Q0.4
	B_5	Uncontrollable	Machine 5 stop operation	I0.4
Machine 6	A_6	Controllable	Machine 6 start operation	Q0.5
	B_6	Uncontrollable	Machine 6 stop operation	I0.5

Table 1. Machine-related events for the case study

The behaviour of each Mx (where x = 1, .., 6) machine is represented by a Gx automaton shown in Figure 7. Each machine has only two states: in the state 0 the machine is stopped, waiting to work, and the state 1 the machine is working on a part. According to Table 1, the start operation is a controllable event Ax, and the stop operation is uncontrollable event Bx.

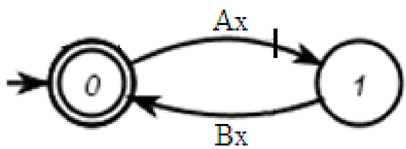


Fig. 7. G_X automaton for each machine

It is important to observe that passive devices need not be modelled, i.e., devices that don't have proper events, such as the buffers in the transfer line, for instance.

6.3 Control specifications modelling

Control specifications are models that describe the desired behaviour for the closed loop system.

The automaton presented at the left-side of Figure 8, shows the control specification of the buffers to avoid their overflow and underflow. It represents the working specification of B_A if x = 1, B_B if x = 3 and, B_D if x = 5. For all buffers, state 0 represents an empty buffer while state 1 represents a full buffer. The specification represented by the *E* automaton at the right-side of Figure 8 prevents the B_C buffer overflow and underflow. B_C will be full (state 1) if either B_2 or B_4 events occur and will be emptied (state 0) if an A_5 event happen. Therefore, machine M_5 will only be able to start operation after machine M_2 or M_4 produce a part in their output buffers. Once a part is deposited on a buffer, another part can only be deposited after a subsequent machine start operation (which signals that it collected a part from the buffer). Note that randomness must be guaranteed to prevent the machine M_5 from working with parts coming from only one of M_2 or M_4 machines.

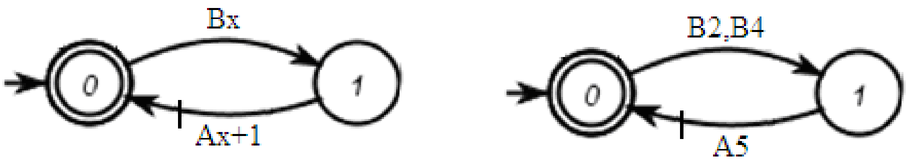


Fig. 8. "E" specifications for B_A, B_B and B_D, buffers (on the left) and *B*c (on the right)

6.4 Synthesis of local modular supervisors

From the devices (G_x) and operating specifications (E_Y) models, a synchronous composition between these models must be performed (as required by Queiroz & Cury, 2000b).

Firstly, you must determine the Product System (PS). To do this, the synchronous composition of all sub-plants that present common events should be performed. It should be looked for the biggest amount of asynchronous subsystems possible. For the present case study no common events between any sub-plants exist, therefore the models previously presented are the set of subsystems of PS.

Then the set of local plants must be determined. To do this, a synchronous product between the subsystems that are affected directly or indirectly by a particular specification must be done.

The most practical way to identify common events is through a table, like Table 2, so the local plants (those that share specifications) are: $G_{locA} = G_1 \| G_2$, $G_{locB} = G_3 \| G_4$, $G_{locC} = G_2 \| G_4 \| G_5$ and $G_{locD} = G_5 \| G_6$. From the common events between specifications and sub-plants analysis, it should be verified if some specification can be grouped together. For the present case study this grouping does not occur.

Following, a synchronous composition of local plants with the specifications should be performed to generate local specifications: $K_{locA} = G_{locA} \| E_A$, $K_{locB} = G_{locB} \| E_B$, $K_{locC} = G_{locC} \| E_C$, and $K_{locD} = G_{locD} \| E_D$.

Finally, the maximum controllable sublanguage of K_{locY} is calculate, which is denoted $SupC(K_{locY} \| G_{locY})$, where Y = {A, B, C, D}. The results are the local supervisors: S_{locA}, S_{locB}, S_{locC} and S_{locD}, that are presented in Figure 9 where the left side shows the supervisors S_{locA} if z=1, S_{locB} if z=3 and S_{locD} if z=5 and; at the right side the S_{locC} supervisor is shown.

	M_1		M_2		M_3		M_4		M_5		M_6	
	A_1	B_1	A_2	B_2	A_3	B_3	A_4	B_4	A_5	B_5	A_6	B_6
E_A		X	X									
E_B						X	X					
E_C				X				X	X			
E_D										X	X	

Table 2. Common events between sub-plants and specifications

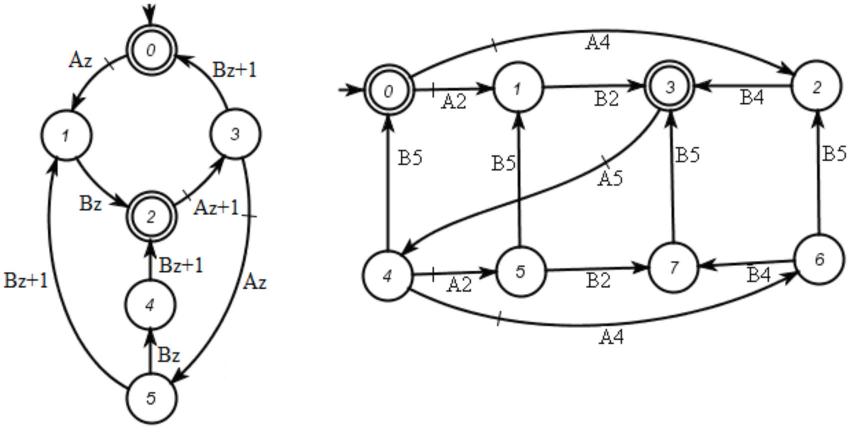


Fig. 9. Local modular supervisors

However, it is necessary to ensure the local modularity of local supervisors altogether so the joint action of all supervisors is nonblocking, as demonstrated by Queiroz & Cury (2000b). To verify local modularity, the synchronous composition of all local supervisors must be performed, as follows: $S = S_{locA} \| S_{locB} \| S_{locC} \| S_{locD}$. The resulting automaton from this composition should be checked for blocking states. If no blocking states can be found, then it can be said that local supervisors are modular to each other.

In order to reduce the amount of memory used in the implementation of these supervisors some tools to reduce the supervisors, these tools keep the control action that disable controllable events, but the supervisors lose information about the plant (Su & Wonham, 2004). However, as the product system will be implemented together with supervisors in the PLC, the information that was lost by reducing the supervisors will be preserved in the product system. Figure 10 shows the same supervisors of Figure 9, but in reduced form where the left-side show the S_{locA} (if z=1), S_{locB} (if z=3) and S_{locD} (if z=5) supervisors, and the S_{locC} supervisor is presented at the right-side.

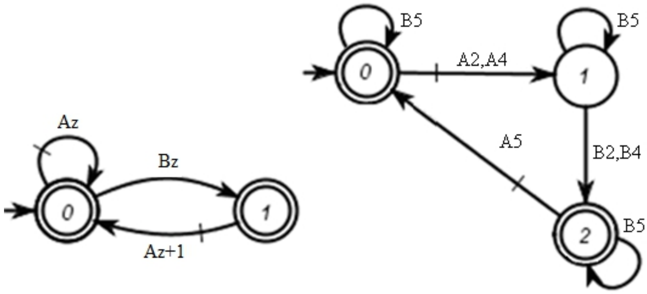


Fig. 10. Reduced supervisors automata

6.5 Following DECON9´s methodology

a. Main Routine

To easy understand, the PLC program is organized as a main routine that calls subroutines for every single block in the flowchart of Figure 5. Figure 11 shows DECON9's main routine. The sequence of calls should be followed in such a way that this main routine works like a template for all systems. Therefore, there will be 10 (ten) subroutines that will be detailed following. Notice that some abbreviations were considered in order to simplify the PLC code. Thus, S_C is the abbreviation for S_{locC} and Sc.0 means state 0 of Supervisor S_C. Moreover, dAx is used to indicate the disabling of the Ax event. Thus, in Figure 11 dA2 and dA4 are used to indicate the disabling of A2 and A4, respectively.

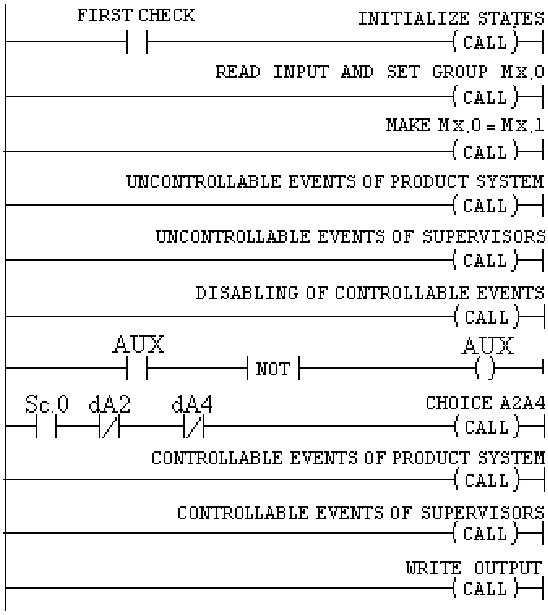


Fig. 11. DECON9 main routine

Figure 12 shows the subroutines in the order they will be called by the main routine. In order to facilitate understanding, the code for each subroutine is shown just below the line that promotes the corresponding call. Each of the subroutines is presented in sequence.

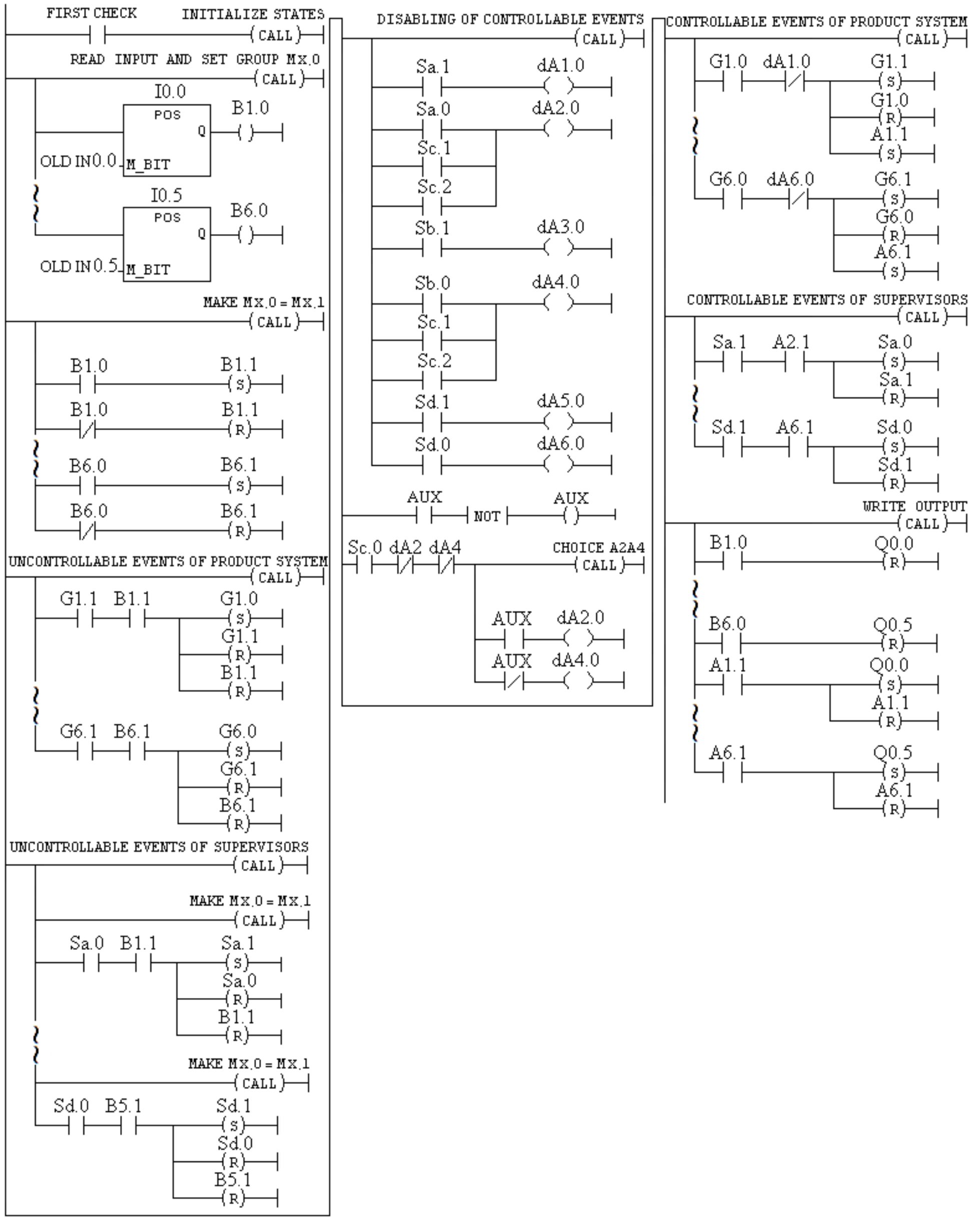


Fig. 12. PLC implementation for the case study

b. State Initialization

The first subroutine initializes all states of the Product System (PS) and supervisors. Thus, the memory that corresponds to the initial state of all automata is set to 1. Remaining memories that represents all other states are set to 0. This should be done only on the very first scan cycle (that's why a memory flag called "first check" is used alongside it) so the automata do not lose its evolutionary feature during a sequence of scan cycles.

c. Reading Inputs

Second subroutine reads PLC inputs and identifies controllable and uncontrollable events that came from the plant. This subroutine is called at the beginning of every scan cycle to verify if there is any positive transition at any PLC input. If so, there is an uncontrollable event being generated that corresponds to that input. The correspondence of inputs and uncontrollable events for the case study is in Table 1. It should be noticed that the "*POS*" PLC function (see Figure 12) ensures that the uncontrollable event will be identified only at the scan cycle immediately after the corresponding input signal changes (positive edge) and that this function is available to the RockWell PLC family.

d. Rescuing Uncontrollable Events

Every uncontrollable event uses two memories, *Mx.0* and *Mx.1*. The first group of memories, *Mx.0*, is responsible for storing the information of all events that actually have been produced by the plant. Therefore, there is a subroutine that updates the second set of memories (*Mx.1*) with the information stored at the other set (*Mx.0*) so the second group is used to promote the state transitions at PS as well as at supervisors.

As long as the information of each event that have been issued by the plant is stored in *Mx.0*, *Mx.1* can eventually lose its information because it can always be recovered from *Mx.0* by calling "Make *Mx.0* = *Mx.1*" subroutine, as shown in Figure 12.

e. Updating Product System with Uncontrollable Events

Next subroutine deals with PS uncontrollable events and is responsible for performing PS state transitions due to these events. It can be interpreted as an "automata player". There is no restriction on the number of events issued by the plant that this automata player is able to deal with. Therefore, at each scan cycle, PS automata can transit states regardless the number of uncontrollable events coming from the plant.

The current state of all subsystems is updated. This can be seen by observing for instance that when G_1 sub-plant is in state 1 and B_1 event happens, the state transition to 0 will occur and, to avoid the avalanche effect, *B1.1* memory is reset to 0 (see Figure 12). It should be noticed that if any other sub-plant is able to promote state transitions, it will be possible to promote it as well.

Because PS is composed of asynchronous subsystem, an event that is dealt with in one subsystem will not occur in another. Thus, there is no problem of erasing its information when its state transition occurs.

f. Updating Supervisors with Uncontrollable Events

Another automata player is implemented here but only to promote transitions for uncontrollable events of the supervisors. For these supervisors, that are not necessarily asynchronous, the same event can produce state transitions in more than one supervisor. Therefore, once the information on every event that promotes a transition is erased to avoid the avalanche effect, this information should be recovered before executing the automata player for each supervisor. That's why "Make *Mx.0* = *Mx.1*" subroutine is called once before each supervisor in the case study, as illustrated in Figure 12.

Note that the program structure used to update the supervisors is the same used for the PS but the supervisors' states are considered instead.

g. Disabling Controllable Events

According to the current state of all supervisors, all controllable events that should be disabled are determined.

Once PS and supervisors states are updated with the transitions promoted by the uncontrollable states issued by the plant, it can be said that all automata implemented into the PLC are in synchrony with the plant, i. e., they are all at the same states as the physical plant.

Therefore, it is possible to identify events that need to be disabled by the conjunction of the supervisors. It is possible that a single event became disabled by the action of many supervisors.

h. The Choice Problem

This subroutine should be called only if necessary and, depending on the state of a given supervisor and on the events involved in the choice problem. For each choice problem that appears, a different subroutine must be created to deal with it.

It is possible that two or more controllable events became disabled by the supervisors. If they belong to the same supervisor a choice problem may occur. In the case study at hand, M_2 and M_4 machines cannot start operation at the same time because they share the same output buffer and thus, once one machine issue a part to that buffer, the other cannot issue another one. In other words, when supervisor S_C is in state 0, A_2 e A_4 events are enabled but cannot be issued at the same time (neither at the same scan cycle) because issuing one means disabling the other. This is a clear choice problem whereby the Product System must decide which one to issue.

According to the flowchart shown in Figure 4 that ensures a solution to the choice problem at the same time that it avoids rendering a blocking system, a "Choice A2A4" subroutine is called (see Figure 12). This subroutine randomly enables only one event at a time, either A_2 or A_4 for the supervisor S_C when it is in its 0 state, for the present case study. An auxiliary memory, called "*AUX*", is used which changes its state (from 0 to 1, and vice-versa) at every scan cycle. Therefore, when *AUX* holds 1, A_2 event is disabled and, when *AUX* holds 0, A_4 is disabled.

i. Issuing Controllable Events from PS

Another automata player is implemented but only controllable events of the Product System are dealt with. Thus, each controllable event that has not been disabled and ready to occur would make PS to transit states and an event to be issued from PLC which means that a controllable event occur at the physical plant.

It is important to observe that the choice problem happens among non-disabled controllable events at a particular state of some supervisor and not among events of different supervisors. Thus, DECON9 allow that many controllable events can be issued at the same scan cycle. For instance, at the present case study, M_1 and M_3 start operation can happen at the same time and, as a consequence, A_1 e A_3 events can also be issued at the same scan cycle.

It should be noticed that every state transition that occur at PS corresponds to signalling a particular event that must be issued.

j. Issuing Controllable Events from Supervisors

As can be noticed in Figure 12, controllable events of PS might promote state transitions on the supervisors. Therefore, another automata player is implemented here but only for supervisors' controllable events.

k. Writing Outputs

Finally, at the end of the scan cycle, PLC outputs are updated. It should be noticed that all output reset conditions are implemented first and just afterwards, output signals are issued according to controllable events.

7. Conclusions

Supervisory Control Theory (SCT) of Discrete Event Systems (DES) has become a major player in the next step manufacturing system automation once it brings formality, predictability and a higher abstract level of specification to the analysis of complex layouts. Some of the advantages of using SCT include: plant and supervisors are high level models; testing of resulting control program is not required once it is produced from a sound theoretical background; equipment or plant behaviour models can be easily reused and; better control programs can be achieved by engineers focusing on the modelling instead of the intricacies of implementing it.

But widespread use of SCT has been hold up by the fact that Programmable Logic Controllers (PLCs) are the basic devices that can be found in the shop-floor. Implementing SCT in PLCs is not a trivial task because many problems and constraints arise while attempting to do it. Many researches have dealt with producing PCL programs from TCS. Some attempts did not propose a methodology but focused on solving particular situations which is far from a generic approach. Most of the existing proposals are based on the monolithic approach for the supervisors' synthesis, and the implementation is performed in ladder language. In some works the synthesis of supervisors is performed according to the local modular approach, which reduces the computational complexity of the synthesis process and the size of supervisors by exploiting specifications modularity and the decentralized structure of composite plants. But almost all of them can only tackle one event per PLC scan cycle, which represents a problem when handling large scale plants. Moreover, in this way the supervisor's update rate and actions will be lower than that obtained via traditional solution, without the use of the SCT. Finally, just a few of them proposed solutions for the broad spectrum of problems that arise when implementing supervisory control in a PLC-based control system.

In this chapter a nine step methodology, named DECON9, was presented. DECON9 is a methodology to implement SCT into PLCs in standardized, efficient and robust ways, closer to real size plants. It is a standardized approach because represents a complete methodology for the whole process, and is divided into simple sub-routines. It can deal with large scale plants because it uses the local modular approach for the supervisors' synthesis. It is an efficient solution because can tackle more than one event per scan cycle. It is robust because can predict problems and solves some of the most common ones. It turns PLCs into a state-machine where supervisors and plant events are explicitly represented and their control reasoning depends on their controllability.

This chapter also reviewed the basics of DES and the problems of implementing SCT into a PLC, presented detailed functioning and implementation of DECON9 and gave an example on how to apply it.

The local modular approach was used for the synthesis of supervisors and their implementation in PLC was programmed using the well-known ladder language. DECON9 use is exemplified by the implementation of the supervisory control of an industrial transfer line case study found in the literature. Using this case study, it is demonstrated that DECON9's advantages include: (*i*) it allows the control logic to deal with many events at each scan cycle, which improves existing approaches that are constrained to only one event at a time; (*ii*) a nonblocking property is achieved thanks to the random selection of controllable events approach that solves the choice problem; (*iii*) there is no fear of an avalanche effect thanks to the use of auxiliary memories, and; (*iv*) uncontrollable events are prioritized.

A computational tool for automatic generation of PLC programs obtained through the Supervisory Control Theory (SCT) is under development. This tool will comply with DECON9. With this tool, the gap between theory and practice will be reduced even further thanks to the automatic procedure based on a sound methodology.

8. Acknowledgment

The authors would like to thank *Pollux Automation Company* and *Santa Catarina State University* (*UDESC*) for their support to pursue this work.

9. References

Afzalian, A.; Noorbakhsh, M. & Navabi, A. (2008). PLC implementation of decentralized supervisory control for dynamic flow controller. *Proceedings of the 17th IEEE International Conference on Control Applications (CCA'08)*, pp. 522-527, San Antonio, Texas (USA), September, 2008.

Afzalian, A. A.; Noorbakhsh, S. M. & Wonham, W. M. (2010). Discrete-Event Supervisory Control for Under-Load Tap-changing Transformers (ULTC): from synthesis to PLC implementation, In: *Discrete Event Simulations*, Aitor Goti (Ed.), pp. 285-310, InTech, ISBN 978-953-307-115-2, Retrieved from <http://www.intechopen.com/download/pdf/pdfs_id/11550>

Ariñez, J.F.; Benhabib, B.; Smith, K.C. & Brandin, B.A. (1993). Design of a PLC-Based Supervisory-Control System for a Manufacturing Workcell, *The Canadian High Technology Show and Conference*, Toronto, 1993.

Balemi, S. (1992). *Control of Discrete Event Systems: Theory and Application*, Ph.D. thesis, Swiss Federal Institute of Technology, Zürich, Switzerland.

Balemi, S. & Brunner, U. A. (1992). Supervision of discrete event systems with communication delays, *Proceedings of the American Control Conference*, pp. 2794-2798, Chicago, IL, USA, June, 1992.

Cassandras, C. G. & Lafortune, S. (2008). *Introduction to Discrete Event Systems. (2nd edition)*, Springer, ISBN: 978-0-387-33332-8, New York, USA.

Chandra, V.; Huang, Z. & Kumar, R. (2003) Automated Control Synthesis for an Assembly Line Using Discrete Event System Control Theory. *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, Vol. 33, No. 2, (may 2003), pp. 284-289, ISSN 1094-6977.

Dietrich, P.; Malik, R.; Wonham, W. M. & Brandin, B. A. (2002). Implementation considerations in supervisory control. In: *Synthesis and Control of Discrete Event Systems*, Caillaud, B.; Darondeau, P.;Lavagno, L.; Xie, X. (Eds), pp. 185-201, Kluwer Academic Publishers.

Fabian, M. & Hellgren, A. (1998). PLC-based implementation of supervisory control for discrete systems, *Proceedings of the 37th IEEE Conference on Decision and Control*, Vol. 3, pp. 3305-3310.

Feng, L. & Wonham, W. M. (2006). TCT: a computation tool for supervisory control synthesis, *Proceedings of the 8th International Workshop on Discrete Event Systems – WODES*, pp. 388-389, Ann Arbor, Michigan, USA, July 2006.

Hasdemir T., Kurtulan, S., & Gören, L. (2004). Implementation of local modular supervisory control for a pneumatic system using PLC. *Proceedings of the 7th International Workshop on Discrete Event Systems (WODES'04)*, pp. 27-31, Reims, France.

Hasdemir, T.; Kurtulan, S.; & Gören, L. (2008). An Implementation Methodology for Supervisory Control Theory. *International Journal of Advanced Manufacturing Technology*, Vol. 36, No. 3, (March 2008), pp. 373-385. ISSN 0268-3768.

Hellgren, A.; Lennartson, B. & Fabian, M. (2002). Modelling and PLC-based implementation of modular supervisory control, *Proceedings of the 6th International Workshop on Discrete Event Systems (WODES'02)*, pp. 1-6. ISBN: 0-7695-1683-1. Zaragoza, Spain, October 2002.

Lauzon, S. C. (1995). *An implementation methodology for the supervisory control of flexible-manufacturing workcells*, M.A. Sc. Thesis, Mechanical Engineering Dpt. University of Toronto, Canada.

Lauzon, S. C.; Mills, J. K.; & Benhabib, B. (1997). An Implementation Methodology for the Supervisory Control of Flexible Manufacturing Workcells, *Journal of Manufacturing Systems*, Vol. 16, No. 2, pp. 91-101.

Leal, A. B.; Cruz, D.L.L.; Hounsell, M.S. (2009). Supervisory Control Implementation into Programmable Logic Controllers. In: *Proceedings of the 14th IEEE International Conference on Emerging Technologies and Factory Automation - ETFA*, pp. 899-905, Palma de Mallorca, 2009.

Leduc, R. J. (1996). *PLC Implementation of a DES supervisor for a manufacturing testbeb: an implementation perspective*, M.A.Sc. Thesis, Dept. of Electrical and Computer Engineering, Univ. of Toronto, Canada.

Leduc, R .J. & Wonham W. M. (1995). PLC implementation of a DES supervisor for a manufacturing test bed. *Proceeding of Thirty-Third Annual Allerton Conference on Communication, Control and Computing*, pp. 519-528, University of Illinois.

Liu, J. & Darabi, H. (2002). Ladder logic implementation of Ramadge-Wonham supervisory controller, *Proceedings of the 6th International Workshop on Discrete Event System (WODES'02),* pp. 383-389. ISBN: 0-7695-1683-1. Zaragoza, Spain, October 2002.

Malik, P. (2002). Generating Controllers from Discrete-Event Models. In: F. Cassez, C. Jard, F. Laroussinie, M. D. Ryan (Eds.), *Proceedings of the Summer school in MOdelling and VErification of Parallel processes (MOVEP),* pp. 337-342.

Manesis, S. & Akantziotis, K. (2005). Automated synthesis of ladder automation circuits based on state-diagrams. *Advances in Engineering Software*, 36, pp .225–233.

Morgenstern, A. & Schneider, K. (2007). Synthesizing Deterministic Controllers in Supervisory Control In: *Informatics in Control, Automation and Robotics II*. Filipe, J.; Ferrier, J-L.; Cetto, J.A.; Carvalho, M. (Eds), pp. 95-102, Springer, ISBN 978-1-4020-5626-0, Netherlands.

Noorbakhsh, M. & Afzalian, A. (2007a). Design and PLC Based Implementation of Supervisory Controller for Under-load Tap-Changer. *Proc. of the 2007 IEEE Int. Conf. on Control, Automation and Systems (ICCAS'07),* pp. 901-906, Seoul, Korea.

Noorbakhsh, M. & Afzalian, A. (2007b). Implementation of supervisory control of DES using PLC. *15th Iranian Conf. on Electrical Engineering (ICEE'07),* (in Farsi), Tehran, Iran.

Noorbakhsh, M. (2008). *DES Supervisory Control for Coordination of Under-Load Tap-Changing Transformer (ULTC) and a Static VAR Compensator (SVC)*. M.A.Sc Thesis, Dept. of Electrical & Computer. Eng., Shahid Abbaspour University of Technology, (in Farsi), Tehran, 2008.

Noorbakhsh, M. & Afzalian, A. (2009). Modeling and synthesis of DES supervisory control for coordinating ULTC and SVC. *Proceedings of the 2009 American Control Conference (ACC' 09)*, pp. 4759-4764, Saint Louis, Missouri USA, June 10-12, 2009.

Possan, M. C.; Leal, A. B. (2009). Implementation of Supervisory Control Systems Based on State Machines. In: *Proceedings of the 14th IEEE International Conference on Emerging Technologies and Factory Automation – ETFA*, pp. 819-826, Palma de Mallorca, 2009.

Queiroz, M. H. de & Cury, J. E. R. (2000a). Modular supervisory control of large scale discrete event systems, In: *Discrete Event Systems: Analysis and Control*. 1st Ed. Massachusetts: Kluwer Academic Publishers, pp. 103-110. Proc. WODES 2000.

Queiroz, M. H. de & Cury, J. E. R. (2000b). Modular control of composed systems. In: *Proc. of the American Control Conference*, pp. 4051-4055, Chicago, USA, 2000.

Queiroz, M. H. de & Cury, J. E. R. (2002). Synthesis and implementation of local modular supervisory control for a manufacturing cell, *Proceedings of the 6th International Workshop on Discrete Event Systems (WODES)*, pp. 1-6. ISBN: 0-7695-1683-1. Zaragoza, Spain, October 2002.

Qiu, R. G. & Joshi, S. B. (1996). Rapid prototyping of control software for automated manufacturing systems using supervisory control theory. ASME, *Manufacturing Engineering Division*, 4, pp. 95-101.

Ramadge, P. J. & Wonham, W. M. (1987). Supervisory control of a class of discrete event processes, *SIAM Journal on Control and Optimization*, Vol. 25, No. 1, pp. 206 - 230.

Ramadge, P.J. & Wonham, W.M. (1989). The control of discrete event systems, *Proceedings of the IEEE*, Vol. 77, No. 1, pp. 81-98.

Reiser, C.; Cunha, A. E. C. da & Cury, J. E. R. (2006). The Environment Grail for Supervisory Control of Discrete Event Systems, *Proceedings of the 8th International Workshop on Discrete Event Systems* – WODES, pp. 390-391, Ann Arbor, Michigan, USA, July 2006.

Rudie, K. (2006). The Integrated Discrete-Event Systems Tool, *Proceedings of the 8th International Workshop on Discrete Event Systems – WODES'06*, pp. 394-395, Ann Arbor, Michigan, USA, July 2006.

Silva, D. B.; Santos, E. A. P.; Vieira, A. D. & Paula, M. A. B. (2008). Application of the supervisory control theory in the project of a robot-centered, variable routed system controller, *Proceedings of the 13th IEEE International Conference on Emerging Technologies and Factory Automation – ETFA'08*, pp. 1-6.

Su, R. & Wonham, W. M. (2004). Supervisor reduction for discrete-event systems, *Discrete Event Dynamic Systems*, Vol. 14, No. 1, pp. 31-53.

Uzam, M.; Gelen, G. & Dalci, R. (2009). A new approach for the ladder logic implementation of Ramadge-Wonham supervisors, *Proceeding of the XXII International Symposium on Information, Communication and Automation Technologies (ICAT'09)*, pp. 1-7. ISBN: 978-1-4244-4220-1, Bosnia 2009.

Vieira, A. D.; Cury, J. E. R. & Queiroz, M. (2006). A Model for PLC Implementation of Supervisory Control of Discrete Event Systems. *Proceedings of the IEEE Conference on Emerging Technologies and Factory Automation - ETFA'06*, pp. 225–232. Czech Republic, September 2006.

Wonham, W. M. (2011). *Supervisory Control of Discrete-Event Systems*, The University of Toronto, available from: http://www.control.utoronto.ca/DES.

9

Platform for Intelligent Manufacturing Systems with Elements of Knowledge Discovery

Tomasz Mączka and Tomasz Żabiński
Rzeszów University of Technology
Poland

1. Introduction

Numerous and significant challenges are currently being faced by manufacturing companies. Product customization demands are constantly growing, customers are expecting shorter delivery times, lower prices, smaller production batches and higher quality. These factors result in significant increase in complexity of production processes and the necessity for continuous optimization. In order to fulfil market demands, managing the production processes require effective support from computer systems and continuous monitoring of manufacturing resources, e.g. machines and employees. In order to provide reliable and accurate data for factory management personnel the computer systems should be integrated with production resources located on the factory floor.

Currently, most production systems are characterized by centralized solutions in organizational and software fields. These systems are no longer appropriate, as they are adapted to high volume, low variety and low flexibility production processes. In order to fulfil current demands, enterprises should reduce batch sizes, delivery times, and product life-cycles and increase product variety. In traditional manufacturing systems this would create an unacceptable decrease in efficiency due to high replacements costs, for example. (Christo & Cardeira, 2007)

Modern computer systems devoted to manufacturing must be scalable, reconfigurable, expandable and open in the structure. The systems should enable an on-line monitoring, control and maximization of the total use of manufacturing resources as well as support human interactions with the system, especially on the factory floor. Due to vast amounts of data collected by the systems, they should automatically process data about the manufacturing processes, human operators, equipment and material requirements as well as discover valuable knowledge for the factory's management personnel. The new generation of manufacturing systems which utilizes artificial intelligence techniques for data analyses is referred to as Intelligent Manufacturing Systems (IMS). IMS industrial implementation requires computer and factory automation systems characterized by a distributed structure, direct communication with manufacturing resources and the application of sophisticated embedded devices on a factory floor. (Oztemel, 2010)

Many concepts in the field of organizational structures for manufacturing have been proposed in recent years to make IMS a reality. It seems that the most promising concepts are: holonic (HMS or Holonic Manufacturing System), fractal, and bionic manufacturing. Further references can be found in (Christo & Cardeira, 2007). In general, it could be stated that a promising organizational structure is a conglomerate of distributed and autonomous units which operate as a set of cooperating entities. It would be impossible to successfully implement the new organizational concepts in the manufacturing industry without suitable distributed control and monitoring hardware and software.

In publications (Leitão, 2008), (Colombo et al., 2006) an agent-based software is designated as technology for industrial IMS realization, regardless of the chosen organizational paradigm. The integration of HMS and multi-agent software technology is currently presented as the most promising foundation for industrial implementations of IMS. The HMS paradigm is based on concepts originally developed by Arthur Koestler in 1969 with reference to social organizations and living organisms. The term holon describes a basic unit of organization and has two important characteristics: autonomy and co-operation. In a manufacturing system, a holon can represent a physical or logical activity, e.g. a machine, robot, order, machine section, flexible manufacturing system and even a human operator. The holon possesses the knowledge about itself and about the environment and has an ability to cooperate with other holons. From this viewpoint, a manufacturing system is a holarchy, which is defined as a system of holons organized in a dynamical hierarchical structure. Manufacturing system goals are achieved by cooperation between holons. Due to conceptual similarities of HMS and agent-based software, it seems clear that their combination should create a promising platform for an industrial IMS implementation.

On the other hand, humans still play an important role in manufacturing systems and in spite of predictions from the seventies, which suggested that human operators would no longer be needed in fully automated production, they even play a more important role nowadays than they did in the past (Oborski, 2004). Proper cooperation between humans and machines or humans and manufacturing control systems can significantly improve overall production effectiveness, so human system interface design is still an active research area. Human System Interfaces (HSI) are responsible for efficient cooperation between operators and computer systems. (Gong, 2009) It seems clear that convenient and reliable human system interaction is a key factor for successful industrial IMS implementation.

In this chapter, results of the project devoted to the development of a hardware and software platform for IMS are to be described. The platform based on Programmable Automation Controllers (PAC) has already been successfully tested by being included in everyday production processes in four small and medium Polish metal component production companies. The platform was employed in order to monitor production resources in real-time and to conduct communication between computer systems, machines and operators. On the basis of the tests results, it has been experimentally proven that the main development-related barrier for real deployment of IMS (Leitão, 2008), i.e. absence of industrial controllers with appropriate capabilities, is out-of-date. Within 28 months of the system operation it has been proven that modern PAC are capable of running data processing, communication and graphical user interface modules directly on the PAC controller in parallel with PLC programs. The hardware and software system which has been created constitutes of a platform for future complete implementation of the IMS.

The project has been created by the Department of Computer and Control Engineering, Rzeszów University of Technology, in cooperation with Bernacki Industrial Services company and the students from the Automation and Robotics scientific circle called ROBO (ROBO, 2011). The project has been made under the auspices of Green Forge Innovation Cluster.

The long-term goal of this project is the development and industrial implementation of full IMS concept for small and medium production companies. The goal determines two main assumptions for the selection of hardware and software elements. The first assumption is a possibility to include in the system newer machines with advanced control equipment as well as older ones without controllers. The second one is a reasonable cost of the system. In the paper (Żabiński et al., 2009) the results of the first stage of the project were presented. During the first stage, the hardware was selected and the prototype and functionality limited testbed for one machine was installed in a screw manufacturing factory. The system was tested in a real production process. The first stage of the project was to provide valuable benefits for the factory management board in the field of monitoring availability, performance and quality of operation of the machines and operators. Additionally, control tasks in a PLC layer of the system were done using ST (Structured Text) programming language for the machines which had not been previously equipped with a controller. The success of the first stage resulted in the project continuation and the findings of the current stage are presented in this paper and in (Mączka & Czech, 2010), (Żabiński & Mączka, 2011), (Mączka & Żabiński 2011), (Mączka et al., 2010).

The main goal of the current project stage is to develop and test in a real production environment a system structure with PACs integrated with touchable panels for each machine. Some improvements of HSI for the factory floor are being made, according to suggestions acquired during the first stage, and new functionalities for different organizational units, i.e. maintenance, tool and material departments. The important part of current and future works is to employ artificial intelligence and data mining technology to give factory management personnel reliable and long-term knowledge about the production processes.

2. Systems testbed structures

Up to now, four system industrial testbeds have been constructed. The first one was installed in a screw manufacturing company which is a member of the Green Forge Innovation Cluster. The Green Forge Innovation Cluster is an association of metal production companies and scientific institutions from southern Poland, which aims at innovative solutions development for metal components production. The second one was installed at the WSK PZL-Rzeszów company, in the department which produces major rotating parts for the aviation industry. The first testbed consists of two machine sections formed by cold forging press machines. The first section includes machines without PLC controllers but the second one consists of 12 modern machines equipped with PLCs and advanced cold forge process monitoring devices. The second testbed includes one production line with four CNC vertical turning lathes and two CNC machining centres. The machines in the testbeds have been operated by experienced operators who interact with the system using various peripheral devices, such as barcode readers, electronic calipers and industrial touch panels. (Żabiński & Mączka, 2011) There is also a third testbed installed in a different department at “WSK PZL-Rzeszów”, which monitors one machine.

During the project, a mobile testbed with GSM communication was also constructed. The purpose of a mobile testbed is to allow production companies to test the system without bearing the costs of communication infrastructure installation. Thanks to such a testbed, companies can better define system functionalities better which are very important for them, taking into account the production profile and organizational structure. Currently, this testbed has been installed in a metal component manufacturing company, where one machine has been monitored. (Mączka & Żabiński, 2011)

In the hardware and software part of the system, three main layers can be distinguished: a factory floor hardware and software, a data server layer and WWW (World Wide Web) client stations.

2.1 Factory floor layer

Due to the scalable, reconfigurable, expandable and open structure of the platform, the industrial implementations differ in functionality as well as in hardware and software elements installed on a factory floor.

2.1.1 Testbed with PACs

In the first implementation type, Programmable Automation Controllers (PACs), also known as embedded PCs, are used on a factory floor. PACs are equipped with operating systems and meet the demands of modern manufacturing systems as they combine features of traditional Programmable Logic Controllers (PLCs) and personal computers (PCs). The main feature of PACs is the ability to use the same device for various tasks simultaneously, e.g. data acquiring, processing and collection; process and motion control; communication with databases or information systems; Graphical User Interface (GUI) implementation, etc. There are two kinds of Windows system available for the controllers, i.e. Windows CE and Windows XP Embedded. Windows CE is equipped with .Net Compact Framework, Windows XP Embedded is equipped with .Net Framework. There are benefits of using the XP Embedded platform, e.g. homogeneity of the software platform for controllers and PC stations as well as availability of network and virus protection software. Due to the financial reasons, Ethernet network for communication and controllers with Windows CE were chosen for the two testbeds.

In the system, PACs acquire data form machines using distributed EtherCAT (Granados, 2006) communication devices equipped with digital or analog inputs. Each PAC is equipped with Windows CE 6.0 operating system, real-time PLC subsystem, UPS (Uninterruptible Power Supply), Ethernet as well as RS-232/485 interfaces for communication and DVI (Digital Visual Interface)/USB (Universal Serial Bus) interfaces for touchable monitors connection. One controller with peripherals, i.e. an industrial 15" touch panel, RFID (Radio-frequency Identification) cards reader and barcode reader, is installed in each machine section or production line. The hardware system structure for a factory floor is shown in Fig. 1.

The software for PACs consist of two layers. The first layer is PLC software written in ST (Structured Text) language, which is mainly responsible for reading and writing physical inputs and outputs. The second layer constitutes the application written in C# language for .NET Compact Framework (CF), which runs under Windows CE. (Microsoft Developer Network, 2011)

The second layer consists of the following modules:

- the module for communication between the PLC program and other system parts,
- the module for communication with database using web services technology,
- the operator's GUI.

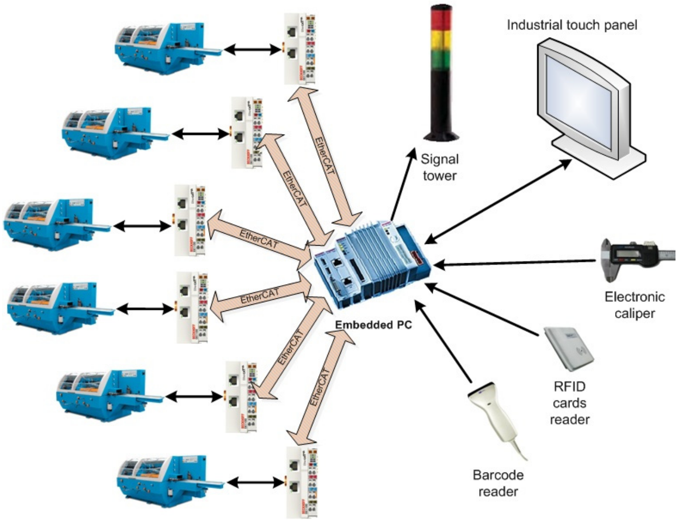


Fig. 1. Hardware structure of the platform for the first implementation type

PLC control programs run in the PLC layer on the same device simultaneously with GUI, data processing and database communication modules which run in Windows CE layer (Fig. 2). The ADS (Automation Device Specification) protocol enables C# programs to read and write data directly from and to PLC programs via names of PLC variables. It significantly simplifies the communication between PLC and C# applications.

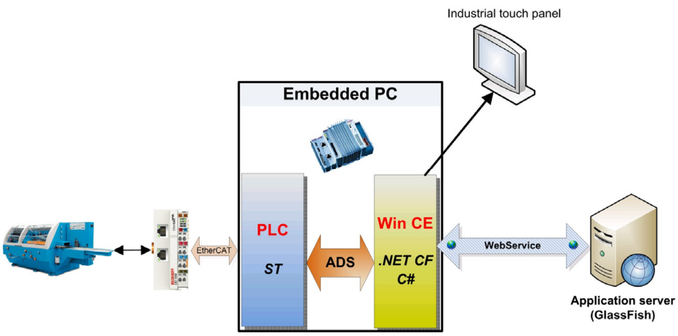


Fig. 2. PAC software structure

The PLC and Windows software for embedded PC controller was designed and implemented in order to control up to 6 machines. It gives flexibility in the system structure, e.g. for more demanding PLC or CNC control tasks there is a configuration of one controller for one possible machine. For simple machines or machines already equipped with controllers, it is possible to create a configuration with one embedded PC controller and up to six machines. This configuration can be used, for instance, to incorporate machine sections into the system.

Currently, the new implementation type supported by EU funds, is under construction. It concerns installation of PACs integrated with touchable panels for each machine (Fig. 3). The panel allows operators to interact with the system and input basic data regarding corresponding machines, e.g. reason for production stoppage and references for orders, materials and tools. In this case, the computational power of each PAC is devoted to one machine and is used for data gathering, intelligent data analysis, intelligent condition monitoring, alarms detection, etc. The PACs models were selected in order to provide sufficient performance for the future multi-agent system structure version, with separate machine agents running on each PAC.

In this implementation type, apart from separate PAC for each machine, personal computers with HSI for operators can be flexibly connected to the system. PCs are treated as additional system "access points", which allow interaction with the system on the same level as PACs. Additionally, it can e.g. provide technical documentation for a particular production process in a more convenient way than on the PAC with a small touchpanel.

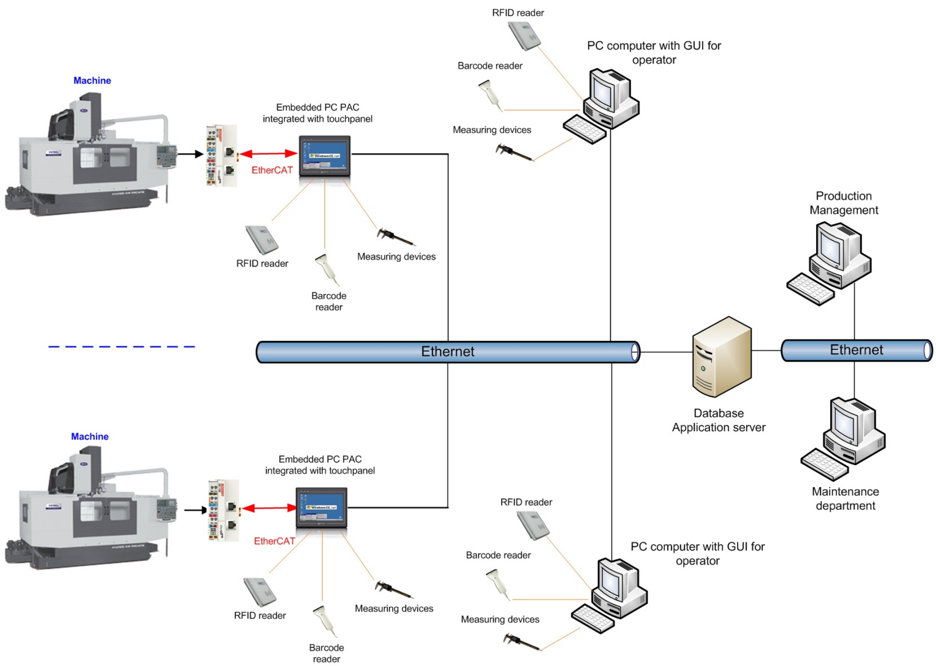


Fig. 3. Hardware structure of the platform with PAC for single machine

2.1.2 Testbed with industrial PC

System structure in the "WSK PZL-Rzeszów" testbed is going to be modified, in order to obtain more flexible architecture. The new structure should also simplify the system implementation in companies with various production and data resources. The new software structure should enable its easy adaptation to the needs of other production sectors. During the development of the testbed, additional diagnostics and process monitoring equipment will be included in the system, e.g. quality measurement devices, current and force sensors, etc. In concern to the WSK testbed, it is planned to include an additional 64 machines in the system. Due to customer demands, one industrial PC computer will be installed for each production line. The industrial PC's task is monitoring states of the machines included in particular production line on base of digital (machine work mode, machine engine state) and analog (spindle load) signals from machines' control systems. Windows XP Embedded with real-time and soft PLC TwinCAT subsystem is the operating environment for industrial PC; connection with input/output modules is performed via EtherCAT bus using star network topology. Chosen topology minimizes communication problems if some part of the EtherCAT network infrastructure will fail, e.g. in case of network cable break.

In this type of implementation, machine operators interact with the system using PCs workstations, placed near monitored machines. The basic scope of data, which can be viewed and input carried out by operators, is similar like in PAC implementation. Detailed description will be given in section 3.1.

Additionally, the system is going to be integrated with an SAP business software and will be used for delivering electronic versions of technical and quality control documents directly to operators' workstations. The system should support production management as well as support the maintenance department by the usage of advanced and intelligent machine condition monitoring software tools. The new testbed structure is shown in Fig. 4. (Żabiński & Mączka 2011)

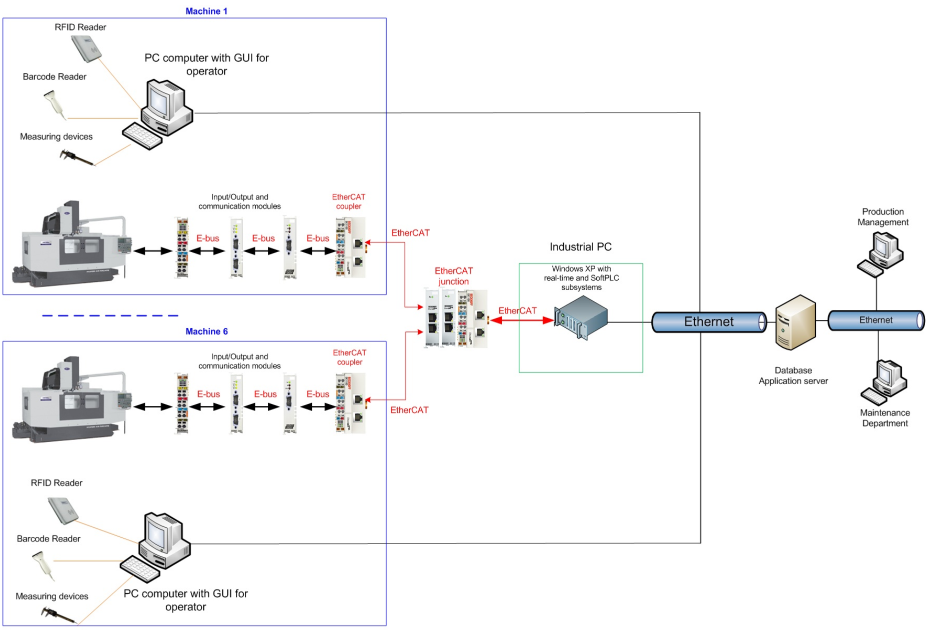


Fig. 4. Hardware structure with industrial PC

2.2 Data server and end clients layer

The data and application server layer is common for all the structures. It includes the PostgreSQL database server and the GlashFish application server. GlassFish is an open-source application server compatible with Java Platform Enterprise Edition (J2EE). (GlassFish Community, 2011) PostgreSQL is open source object-relational database system which conforms to the ANSI-SQL (Structured Query Language) 2008 standard. (PostgreSQL, 2011) The application server hosts communication and data processing modules with web services written in Java. The WWW client layer utilizes websites written in JSF (Java Server Faces), JSP (Java Server Pages), AJAX (Asynchronous Javascript and XML) and also works under GlassFish server control. Communication between PACs or PCs and the database and between the presentation layer and the database is performed using web services or Enterprise Java Beans (EJB) technology.

3. Human system interfaces

Human System Interface, which was developed for the IMS project, consists of two main layers, i.e. a factory floor layer and a WWW layer. The first layer is a GUI application which runs on an embedded PC installed on the factory floor. In this layer the communication between an operator and the system is done via a 15″ touchable monitor. The second layer is a web page accessed through a web browser from the factory intranet or the Internet. The Polish language is used in HSI, as the system has been installed in Polish factories. Due to this reason, the GUI language presented in this section is Polish.

3.1 Factory floor layer

The HSI for factory floor layer has two main operation modes, i.e. locked and unlocked. In the locked mode an operator can only observe information presented on the screen. In the unlocked mode an operator can interact with the system. An operator can change the HSI mode using his RFID card. Thanks to the RFID operator's badges a security policy was implemented on the factory floor. In the locked mode visual information of machines operation mode, production plan and plan realization and the necessity of an operator interaction with the system is presented. The necessity of an operator interaction with the system is indicated by the blinking of a panel associated with the machine which needs intervention. The locked mode screen for a machine section with six machines is presented in Fig. 5.

In the unlocked HSI mode an operator can perform various tasks connected with the system, e.g. login, logout, taking up shift, order selection or confirmation, stop reason and quality control data input etc. The unlocked mode screen for a machine section with six machines is presented in Fig. 6.

The HSI consists of two main sections, i.e. the system section (it allows login, logout, etc.) and the machines section (it allows data input for particular machines). Small rectangles with letters O, SR, QC, S, associated with each machine panel (a large rectangle with machine name, e.g. Tłocznia T-19), indicates the action which should be performed for the particular machine, i.e. O – order selection or confirmation, SR – stop reason input, QC – quality control data input and S – service. When the machine panel is blinking, an operator can quickly determine the operation which should be performed, the color of the letters O, QC, SR or S becomes red for the active action. An example of an order input screen is shown in Fig. 7. Operators can input order code using on-screen keyboard or via barcode reader. (Żabiński & Mączka, 2011)

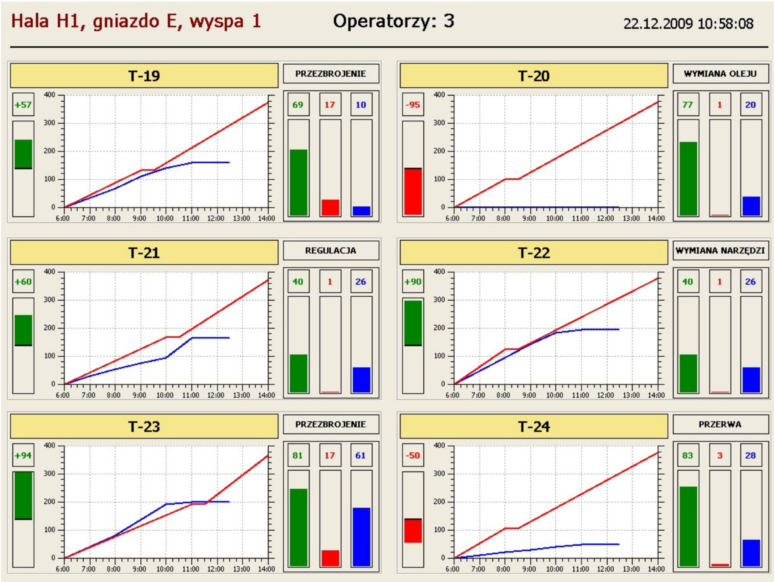


Fig. 5. The factory floor HSI locked mode screen for six machines

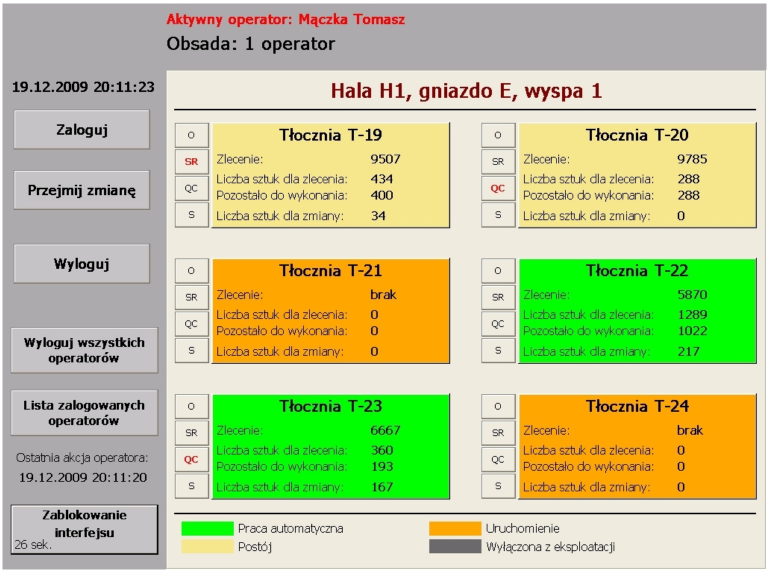


Fig. 6. The factory floor HSI unlocked mode screen for six machines

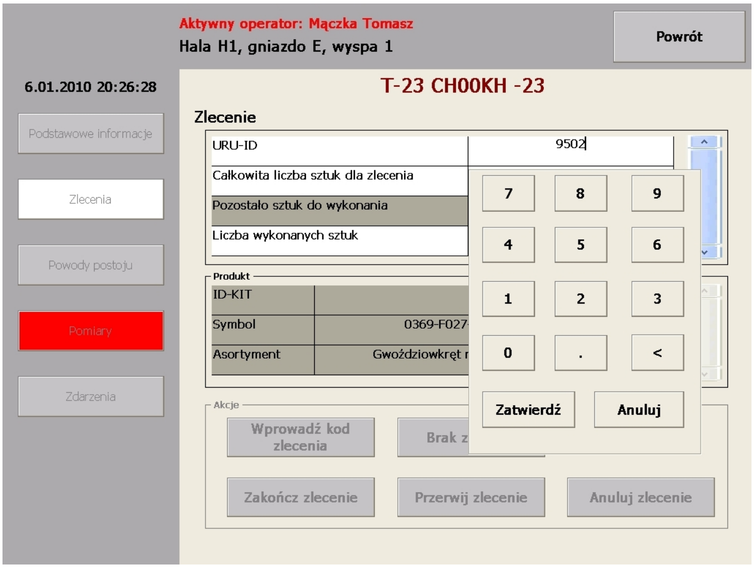


Fig. 7. Order inputting screen for particular machine

3.2 WWW layer

The WWW layer includes two main modules, i.e. an on-line view and statistics. The on-line view enables on-line monitoring of machines operation mode, e.g. production, stoppage, lack of operator and also other information like: operator identifier, order identifier, shift production quantity, daily machine operation structure or detailed history of events. The on-line view for a production hall is presented in Fig. 8.

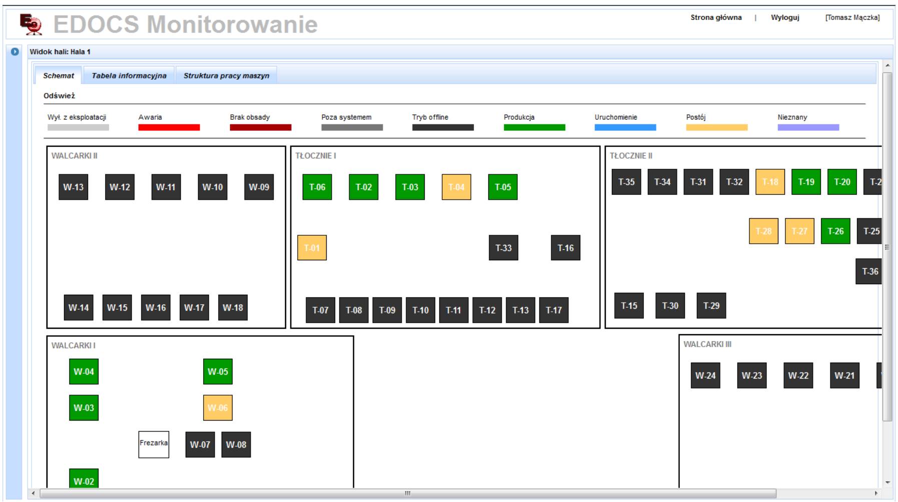


Fig. 8. Production state on-line view for a production hall

The statistics module enables computing some statistical factors concerning: machines work time, production quantity, failures, orders, operators work time etc. It enables users to configure statistics parameters, i.e. analysis time interval, statistics elements, type of presented data (for instance daily or weekly analysis type) and chart type (e.g. bar graph, line graph). An example of a statistics screen for machine production quantity with its configuration options is shown in Fig. 9. (Żabiński & Mączka, 2011)

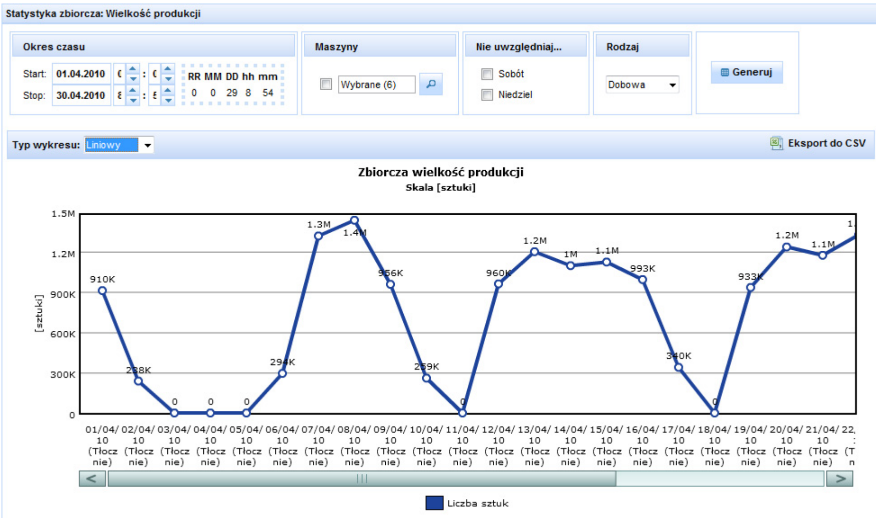


Fig. 9. Machines production quantity statistic with its configuration options

4. System operation results

Currently, four system testbeds are installed in real production environments in small and medium factories and have been used in daily production processes. All of the testbeds currently utilize the first structure described in section 2.1.1, with embedded PCs on the factory floors. The first testbed has been installed in a screw manufacturing factory since 16 May, 2009. Eighteen machines for cold forging are currently monitored. The second testbed was installed at the WSK PZL-Rzeszów company, in the department which produces major rotating parts for the aviation industry and it has been in operation since 21 September, 2010 with six CNC machines included in the system. The new system implementation with the second type structure for this testbed, described in section 2.1.2, is under construction. There is also one testbed where a single machine is monitored using a mobile (all-in-one) system testbed with GSM communication. The next separated testbed with one machine was installed in April 2011 in a different department at "WSK PZL-Rzeszów"of the aviation parts producing company.

Currently, the system is responsible for collecting data concerning machine operation and operators' work. The PLC layer is responsible for detecting and registering events which occurred in the machines, for instance the oil pump and the main motor start/stop, failure and emergency signals, the machine operational mode (manual or automatic), signals from diagnostics modules (process monitoring devices), etc. The PLC program is also responsible for registering the quantity produced. Information about events, including timestamps, machine and operator identifiers and other additional parameters, is stored in the database. Two mechanisms are used to store data in the database, such as: an asynchronous event driven method and a synchronous one with a 10 second time period for diagnostics purposes. The system is also responsible for detecting and storing information on breakdowns, setup and adjustments, minor stoppages, reduced speed etc. Every production stoppage must be assigned with an appropriate reason. Tool failure, for example, are automatically detected, while others have to be manually chosen by operators via the HSI.

On the server side, there are software modules used for calculating different KPIs (Key Performance Indicators), e.g. production efficiency, equipment and operators efficiency etc. The real production quantity report as a function of day, calculated for time interval from 1-03-2011 to 31-03-2011 for 6 machines, is presented in Fig. 10. During this period, the planned production time for the machines was 24 hours per day (3 shifts). As shown in Fig. 10, there were some fluctuations of production quantity.

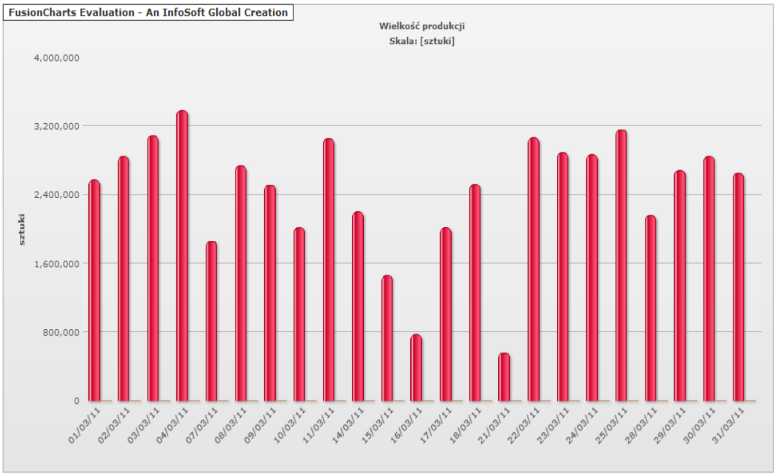


Fig. 10. Production quantity report – number of produced items as a function of days

A machine operation time structure is analyzed and can be shown as a horizontal graph (Fig. 11). At the moment there is a possibility of analyzing data from three points of view: a general view, a view with stop reasons and a detailed view. The general view divides machine operational time into three categories, i.e. the operator's absence, the automatic production and the stoppage. In the analysis for the view with stop reasons, each stoppage time period is associated with the appropriate stop reason. In the detailed view, periods of the manual machine operation are distinguished in each stoppage time. Different colors are designated for appropriate time intervals (Fig. 11), e.g. the stoppage – light brown, the automatic production – green, the manual operation – dark green, the start-up time – blue, the electrical breakdown – red, etc.

During the long term test, when the system was included in the regular daily production, it was proven that the selected hardware and software platform is suitable for industrial implementation of the IMS. The software modules (HSI, communication, web services, data acquisition) for Windows CE have been running successfully on the embedded PC controller in parallel to PLC programs. (Mączka & Czech, 2010)

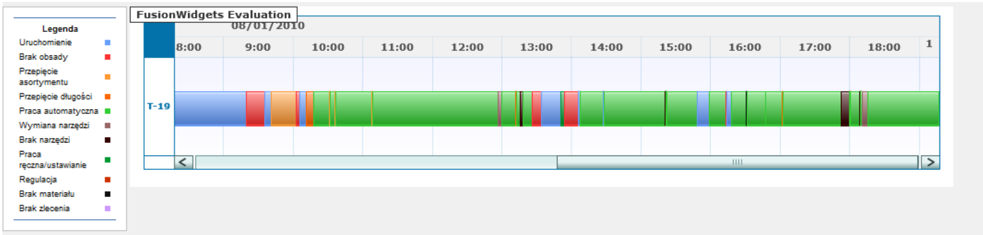


Fig. 11. Machine operation time structure with stop reasons

Within 704 days of system operation in the screws production testbed, the whole number of events registered in the testbed was 3,883,374. The system covers: production monitoring, quality control, material and tools management and also fundamental support for the maintenance department. In the company which produces aviation parts, 57,230 events were registered within 59 days of system operation.

5. Association rules application for knowledge discovery

Taking into account the limited number of machines included in the currently working system testbeds, it can be stated that the number of data collected is considerable. A long-term data analysis to make reasoning and generalized conclusions about production processes would be a demanding task for human analysts. Therefore, there is a need to employ artificial intelligence and data mining technology to give factory management personnel reliable knowledge of the production processes. In this section, the results of the initial tests in the areas of applying data mining and artificial intelligence techniques to discover knowledge about the production processes, are to be described.

It is expected that continuously discovered knowledge will support everyday production process management and control, thus providing the answers to numerous questions, e.g. what are the bottlenecks in the production systems etc. Moreover, it is envisaged that the system will be able to automatically identify relationships in the production systems, discover possibilities for more effective usage of production resources and use Statistical Process Control (SPC) with artificial intelligence support for early detection of possible problems in production systems.

Initial work in this area concerns the creation of tools for automatic rules (patterns) generation, which will describe relations between values in the database. The discovered patterns could be used for detecting operators' improper actions, which could have an influence on machine operation, e.g. increasing downtime duration and number of breakdowns. The rules are referred to as association rules.

5.1 Introduction to the experiment

The goal of association rules is to detect relationships between specific values of categorical variables in large data sets.

The formal definition of the problem is as follows: Let D = {t_1; t_2; …; t_m} be a set of *m* transactions, called data set or database. Let I = {i_1; i_2; …; i_n} be a set of possible *n* binary attributes for transaction, called items. Single transaction T is a set of items such that $T \subseteq I$.

Assume that X is a set of some items from I, so $X \subseteq I$. A transaction T contains X if the transaction contains all items from X, so $X \subseteq T$.

An association rule is an implication of the form X => Y, where $X \subset I$, $Y \subset I$, and $X \cap Y = \emptyset$. The rule X => Y holds in the transaction set D with confidence c if c% of transactions in D that contain X also contain Y. The rule X => Y has support *s* in the transaction set D if s% of transactions in D contain $X \cup Y$. Given a set of transactions D, the problem of mining association rules is to generate all association rules that have support and confidence greater than the user-specified minimum support (minsup) and minimum confidence (minconf) respectively. (Agrawal & Srikant, 1994).

The experiment of finding association rules in production data has been performed. The first step of the experiment was to choose the subset of data to analyze, i.e. time period and attributes, and to extract raw data from screw manufacturing company database to CSV (Comma Separated Values) text format. CSV format was chosen because of the possibility of loading data directly into data mining software, i.e. Statistica or Weka. 127232 events registered on 12 machines from 4.01.2011 to 16.07.2011, concerning machine operational state, were extracted using SQL query and pgAdmin database management tool. The structure of events is shown in Table 1, the table contains only a subset of registered events.

plc_time	**event_type_id**	**machine_id**
2011-01-04 07:59:25.784	2000	41
2011-01-04 08:26:27.565	2001	41
2011-01-04 08:28:49.845	2000	41
2011-01-04 08:29:42.705	2001	41
2011-06-14 21:49:38.032	2000	44
2011-06-14 23:29:17.732	2001	44
2011-07-11 06:59:49.812	2000	68
2011-07-11 07:35:15.332	2001	68
2011-07-11 07:41:44.872	2000	68
2011-07-11 08:21:07.812	2001	68

Table 1. Structure of raw events extracted from production database

Attributes of a single event are:

- plc_time– time of event registration in the PLC layer
- event_type_id– type of registered event, 2000 is production start, 2001 is production stop
- machine_id– identifier of a machine for which the event was registered

After consultation with the company production management personnel, an assumption has been made that length of times of continuous machine state, i.e. length of production and length of stoppage will be important factor to analyze. It seems to be clear that if continuous production time of a particular machine lasts longer than the others, this machine works more efficiently, without the need for operator action. It is worthwhile noticing that lower number of stoppages should positively affect machine lifetime and save energy.

5.2 Data preparation

The format for raw data presentation in Table 1 is not useful for discovering associations concerning machines continuous interval length, as raw events do not directly reflect particular machine state during a particular period of time. Because of this, data needs to be pre-processed to the list of production and stoppage intervals for each machine. Each record should contain interval start date, interval end date and interval length. Pre-processing task was done using Python script, which analyzes events list and produces stoppage intervals, if the current analyzed event is 2001 (production stop) and next analyzed event is 2000 (production start). In the opposite situation, production interval is inserted to the result list. Data structure after the pre-processing phase is shown in Table 2.

start	End	type	length_min	machine_id
2011-01-04 07:59:25.784	2011-01-04 08:26:27.565	W	27.03	41
2011-01-04 08:26:27.565	2011-01-04 08:28:49.845	S	2.37	41
2011-01-04 08:28:49.845	2011-01-04 08:29:42.705	W	0.88	41
…				
2011-06-14 21:49:38.032	2011-06-14 23:29:17.732	W	99.66	44
…				
2011-07-11 06:59:49.812	2011-07-11 07:35:15.332	W	35.43	68
2011-07-11 07:35:15.332	2011-07-11 07:41:44.872	S	6.49	68
2011-07-11 07:41:44.872	2011-07-11 08:21:07.812	W	39.38	68

Table 2. Data structure after the pre-processing phase

Attributes of single interval:

- start – timestamp of interval begin,
- end – timestamp of interval end,
- type – interval type, W is work, S is stoppage,
- len_min – interval time length in minutes,
- machine_id – identifier of machine associated with interval.

Start and end timestamps has only informational role and they are omitted in the process of finding associated rules. However, data presented in Table 2 are not ready for application of associated rules finding algorithms. It results from fact, that known association rules discovering algorithms deals with data, whose attributes have discrete or categorical values. In above case, attributes *type* and *machine_id* are categorical, but *len_min* has continuous values. The solution of this problem is mapping attributes with continuous values to categorical attributes, referred in (Agrawal & Srikant, 1996) as partitioning quantitative attributes.

This transformation was, like the previous, performed by Python script. Number of categories and length of each category's interval were chosen based on minimum and maximum values of *len_min*, in order to obtain regular distribution of data in generated categories. 46 categories were generated, starting from [0-0.1m] (interval length greater than 0 to 0.1 minutes, or 10 seconds) to [700-+Inf m] (interval length greater of equal 700 minutes). Example values of processed items are contained in table 3. Letter 'M' was added before machines identifiers to indicate that this is categorical, not numerical attribute.

start	End	type	length_category	machine_id
2011-01-04 07:59:25.784	2011-01-04 08:26:27.565	W	[20.1-30.1m]	M41
2011-01-04 08:26:27.565	2011-01-04 08:28:49.845	S	[2.1-3.1m]	M41
2011-01-04 08:28:49.845	2011-01-04 08:29:42.705	W	[0.8-0.9m]	M41
…				
2011-06-14 21:49:38.032	2011-06-14 23:29:17.732	W	[90.1-100.1m]	M44
…				
2011-07-11 06:59:49.812	2011-07-11 07:35:15.332	W	[30.1-40.1m]	M68
2011-07-11 07:35:15.332	2011-07-11 07:41:44.872	S	[6.1-7.1m]	M68
2011-07-11 07:41:44.872	2011-07-11 08:21:07.812	W	[30.1-40.1m]	M68

Table 3. Data after partitioning quantitive attribute *len_min* to categorical attribute *length_category*

5.3 Finding association rules using Weka

For finding associations in previously prepared data, Weka (Waikato Environment for Knowledge Analysis) software was used. This is a popular suite of a machine learning software written in Java that was developed at the University of Waikato, which is available under the GNU General Public License. Weka is a collection of machine learning algorithms for data mining tasks. Weka contains tools for not only for association rules, but also for classification, regression, clustering and visualization. Its architecture is well-suited for developing new machine learning schemes. (Hall et al., 2009)

In the first step, Weka knowledge explorer was run and preprocessed mining dataset was loaded from CSV file. Weka displays list of attributes in dataset and its basic information like categories, number of items in each category etc. (Fig. 12).

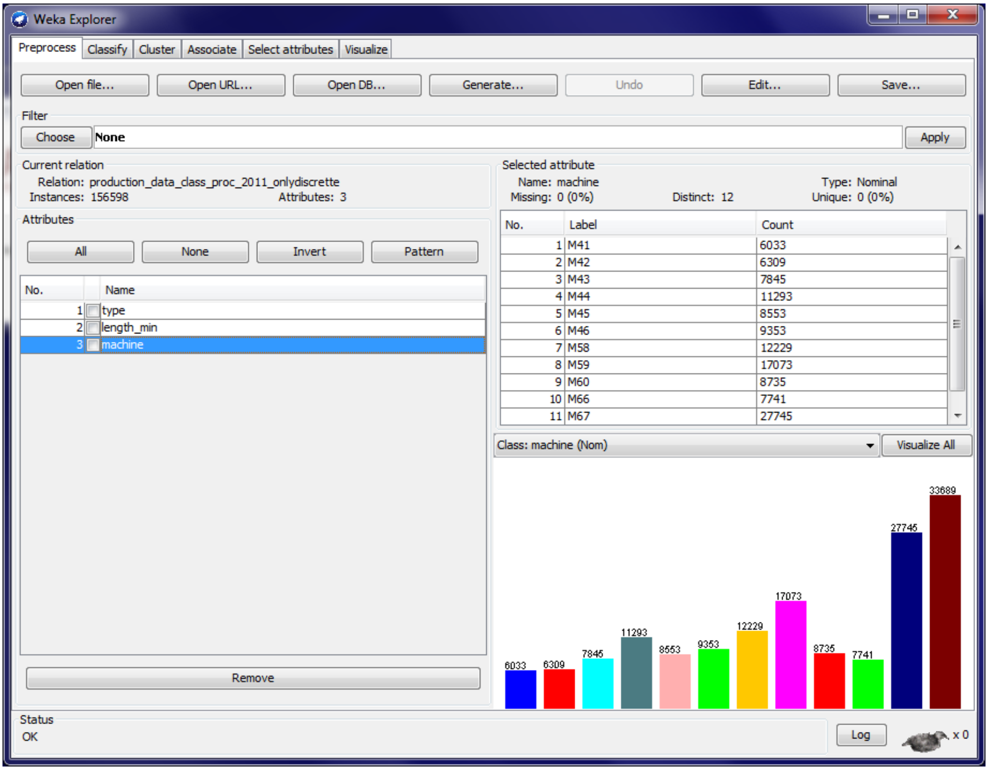


Fig. 12. Weka initial screen after data loading

The next step was running module for association rules generation, choosing algorithm for mining rules and configuring algorithm's parameters. Weka provides few rules discovering algorithms, including Apriori, Predictive Apriori, Generalized Sequential Patterns. In the described experiment, most popular Apriori algorithm was selected with parameters:

support 0.1 (10%), confidence 0.3 (30%). The small value of support was choosen experimentally to not omit rules for particular machines, and purpose of 30% confidence was to ignore irrelevant rules, which applies only to relative small part of data set. Configuration parameters are shown in Fig. 13.

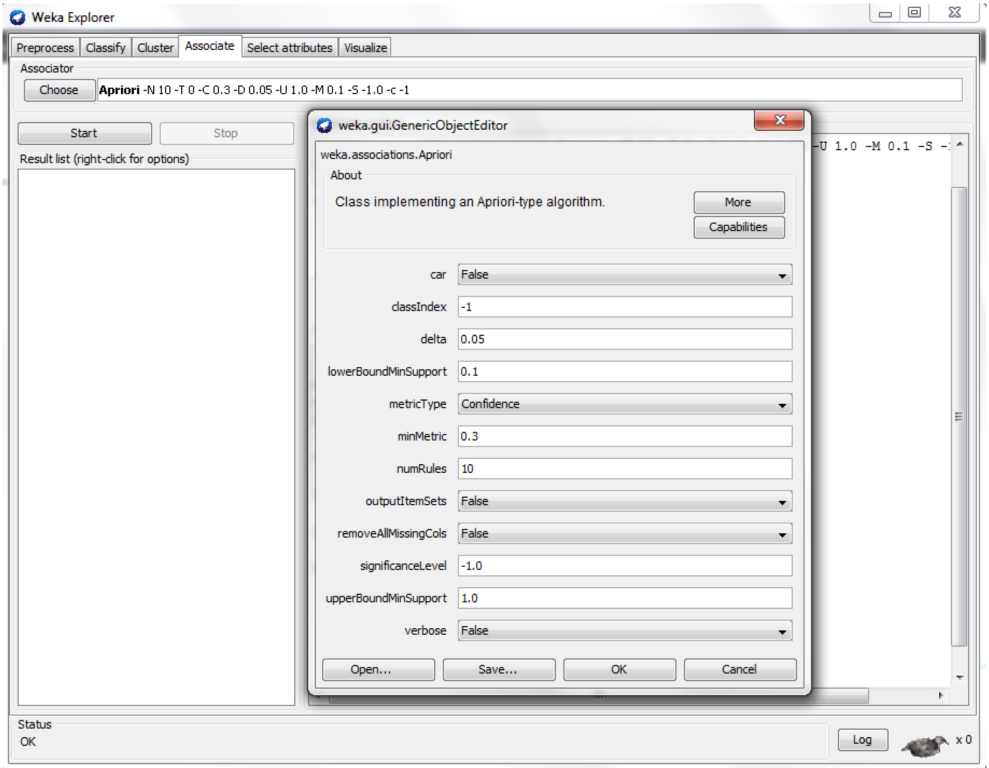


Fig. 13. Apriori algorithm configuration

The algorithm found 10 rules listed below:

1. machine=M67 [27745] ==> length_min=0-0.1m [20099] conf:(0.72)
2. machine=M68 [33689] ==> length_min=0-0.1m [23825] conf:(0.71)
3. length_min=0-0.1m [64730] ==> type=W [37527] conf:(0.58)
4. machine=M68 [33689] ==> type=W [16845] conf:(0.5)
5. machine=M68 [33689] ==> type=S [16844] conf:(0.5)
6. type=W [78305] ==> length_min=0-0.1m [37527] conf:(0.48)
7. length_min=0-0.1m [64730] ==> type=S [27203] conf:(0.42)
8. length_min=0-0.1m [64730] ==> machine=M68 [23825] conf:(0.37)
9. type=S [78293] ==> length_min=0-0.1m [27203] conf:(0.35)
10. length_min=0-0.1m [64730] ==> machine=M67 [20099] conf:(0.31)

Number of records which contain attributes values listed in predecessor and consequent are present in bracket squares. Some potentially interesting rules are underlined, their interpretation may be as follows.

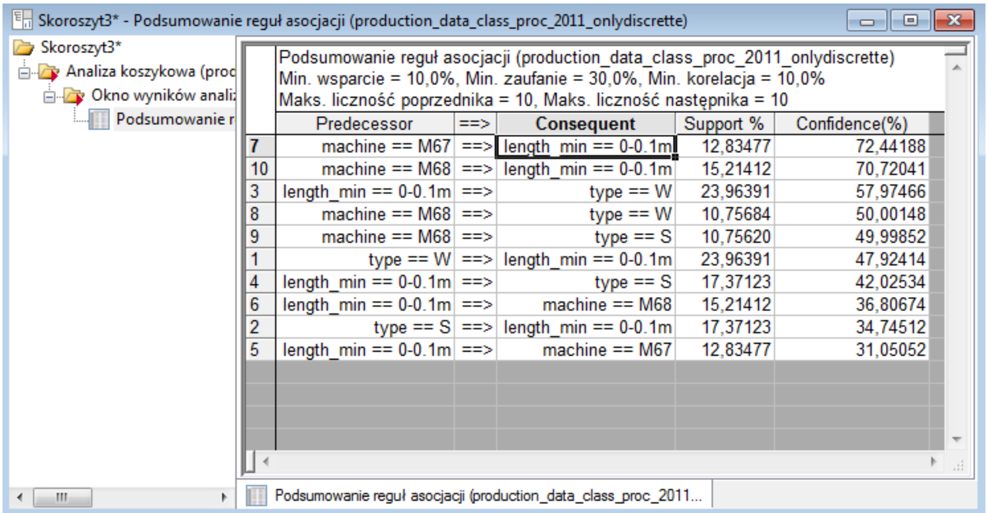
Rules 1 and 2 indicates that for machine M67 and M68, production or stoppage interval lasts usually relatively short – from above 0 to 10 seconds. It suggests that listed machines may have some troubles with stable work, which can lead to low efficiency. Potential reasons of this situation are mechanical problems, improper material, improper operator's actions etc. This situation is also covered by rules 8 and 10, where predecessor and consequent are reversed, these rules are more general.

On the other hand, 48% of work intervals and 35% of stoppage intervals lasts up to 10 seconds, so short time intervals can be, in some level, screw production profile and pushing process specific. However, such potential problem appears to be interesting for deeper analyze by experts from production manager personnel.

5.4 Finding association rules with Statistica

The same experiment like in 5.3 was performed in Statistica 10 Data Miner. This commercial software includes modules for neural network, clusterization, classification trees etc. (StatSoft, 2011) For the association rules mining, module called basket analysis is available. It includes only one rules finding algorithm, the same like used in previous experiment, i.e. Apriori.

Rules found by Statistica, shown in Fig. 14, are the same like in previous experiment, where non-commercial Weka software was used.



Podsumowanie reguł asocjacji (production_data_class_proc_2011_onlydiscrette)
Min. wsparcie = 10,0%, Min. zaufanie = 30,0%, Min. korelacja = 10,0%
Maks. liczność poprzednika = 10, Maks. liczność następnika = 10

	Predecessor	==>	Consequent	Support %	Confidence(%)
7	machine == M67	==>	length_min == 0-0.1m	12,83477	72,44188
10	machine == M68	==>	length_min == 0-0.1m	15,21412	70,72041
3	length_min == 0-0.1m	==>	type == W	23,96391	57,97466
8	machine == M68	==>	type == W	10,75684	50,00148
9	machine == M68	==>	type == S	10,75620	49,99852
1	type == W	==>	length_min == 0-0.1m	23,96391	47,92414
4	length_min == 0-0.1m	==>	type == S	17,37123	42,02534
6	length_min == 0-0.1m	==>	machine == M68	15,21412	36,80674
2	type == S	==>	length_min == 0-0.1m	17,37123	34,74512
5	length_min == 0-0.1m	==>	machine == M67	12,83477	31,05052

Fig. 14. Association rules found by Statistica's basket analysis module

6. Conclusion

So far the project devoted to preparing platform for industrial implementation of Intelligent Manufacturing System, the hardware and software has been selected and tested in a real production process using the preliminary testbeds. Currently, the platform is mainly used for data collection concerning the production processes, machine operation and operators' work. It also provides HSI for machines' operators and system end-users, i.e. factory management board.

The hardware and software layer, which has already been installed in the factory, has created the need and the basis for the employment of advanced data mining and artificial intelligence techniques and multi-agent software structure in the system. Such techniques could be used for detecting operators' improper actions which could have an influence on machines operation, e.g. increasing downtime duration and number of breakdowns. The analysis of the rules could probably give an answer to the following questions, e.g. what factors and in what manner influence production processes, under what circumstances problems occur, how operators react to diagnostic messages, etc. So far, experiment of finding association rules from the data gathered by production monitoring system was performed for the fixed range of events concerning 12 machines work. Some potentially interesting rules, which can help the factories management personnel to detect and eliminate bottlenecks in the production processes, were discovered. Process of finding association rules is not currently automated, but thanks to open source data mining software Weka, it can be easily integrated with the existing system infrastructure. Remaining work to be done in this area includes extending events attributes list, by adding for each item in data set product type identifier and machine's operators identifier. Experiments with different types of rules discovering algorithms, i.e. Apriori Predictive and Generalized Sequence Patterns, will be also performed. Eventually, the system should automatically generate rules, which help to detect the possibility of the problem occurrence on the basis of historical data. Additionally it should suggest the best solution to the problem, therefore the probability of stoppages will be reduced.

New system structures are being developed in order to simplify the system implementation in companies with various production and data resources. The goal is to achieve easy adaptation to the needs of many production sectors. Some of these structures can decrease the cost of system deployment, as popular devices like personal computers can be used on condition that the factory floor is appropriately adaptable, e.g. if there is no oil mist.

However, it should be clearly stated that for industrial implementation of IMS, the structure with separate PAC installed for each machine seems to be the most promising. PAC computational power allows running separate machine agents for each machine, so one of new organizational structure, e.g. holonic, can be implemented.

Tests of new system versions, especially those with multi-agent solutions, are going to be performed with the usage of the laboratory Flexible Manufacturing System (FMS) testbed, presented in Fig. 15. The FMS testbed consists of an integrated CNC milling machine, robot and vision system.

Current results of the project give promising perspectives for an advanced Intelligent Manufacturing System development and implementation with PACs as hardware platform in a real factory for next project stages.

Fig. 15. Laboratory FMS testbed

7. Acknowledgment

FMS testbed was bought as a part of the project No POPW.01.03.00-18-012/09 from the Structural Funds, The Development of Eastern Poland Operational Programme co-financed by the European Union, the European Regional Development Fund.

8. References

Agrawal R., Srikant R. (1994). Fast Algorithms for Mining Association Rules in Large Databases. *VLDB'94, Proceedings of 20th International Conference on Very Large Data Bases*, pp. 487-499, ISBN 1-55860-153-8, Morgan Kaufmann, Santiago de Chile, Chile, 1994.

Agrawal R., Srikant R. (1996). Mining quantitative association rules in large relational tables. *Proceedings of the 1996 ACM SIGMOD international conference on Management of data*, pp. 1 - 12, ISBN 0-89791-794-4, ACM New York, NY, USA, 1996.

Christo C., Cardeira C. (2007). Trends in Intelligent Manufacturing, *Proceedings of IEEE International Symposium on Industrial Electronics*, pp. 3209-3214, ISBN 978-1-4244-0754-5, Vigo, Spain, June 4-7, 2007.

Colombo A., Schoop R., Neubert R. (2006). *An Agent-Based Intelligent Control Platform for Industrial Holonic Manufacturing Systems*, IEEE Trans. Ind. Elect., vol. 53, no. 1, pp. 322-337, ISSN: 0278-0046, Seligenstadt, Germany, February 2006.

GlassFish Community (2011). *GlassFish Server Open Source Edition*, 23.11.2011, Available from: http://glassfish.java.net/.

Gong C. (2009). *Human-Machine Interface: Design Principles of Visual Information*. Human-Machine Interface Design. In: Proc IEEE Conference on Intelligent Human-Machine Systems and Cybernetics, pp 262–265, San Antonio Texas, USA, 2009.

Granados F. (2006). *Analysis: Industrial Ethernet - Driving the growth*, Computing & Control Engineering Journal, vol.17, no.6, pp.14-15, Dec.-Jan. 2006.

Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. (2009). *The WEKA Data Mining Software: An Update,* pp.10-18, SIGKDD Explorations, Volume 11, Issue 1, ACM New York, NY, USA, 2009.

Leitão P. (2008). Agent-based distributed manufacturing control: A state-of-the-art survey. *Engineering Applications of Artificial Intelligence,* No. 22 (7), pp. 979-991, ISBN 0952-1976, ELSEVIER, 2008.

Microsoft Developer Network (2011). *C# Language Specification*, 23.11.2011, Available from http://msdn.microsoft.com/en-us/library/aa645596(v=vs.71).aspx.

Mączka T., Czech T. (2010). Manufacturing Control and Monitoring System – Concept and Implementation. *Proceedings of IEEE International Symposium on Industrial Electronics*, t.1, p.3900-3905, Bari, Italy, July 4-7 2010.

Mączka T., Żabiński T. (2011). System for remote machines and operators monitoring - selected elements (in Polish). *Pomiary Automatyka Robotyka*, p. 62-65, No. 3/2011.

Mączka T., Czech T., Żabiński T (2010). Innovative production control and monitoring system as element of factory of future (in Polish). *Pomiary Automatyka Robotyka*, p.22-25, No. 2/2010.

Oborski P. (2004). Man-machine interactions in advanced manufacturing systems. *The International Journal of Advanced Manufacturing Technology*, Vol. 23. No. 3-4, pp 227-232, ISSN: 1433-3015, Springer-Link, 2004.

Oztemel E. (2010). Intelligent manufacturing systems. *L. Benyoucef, B. Grabot, (Ed.) Artificial Intelligence Techniques for Networked Manufacturing Enterprises Management*, pp. 1-41, ISBN 978-1-84996-118-9, Springer-Verlag, London, 2010.

PostgreSQL (2011). *About*, 23.11.2011, Available from: http://www.postgresql.org/about/.

ROBO (2011). *Student Automation and Robotics scientific circle ROBO (in Polish)*, 23.11.2011, Available from http://www.robo.kia.prz.edu.pl/.

StatSoft (2011). *STATISTICA Product Overview*, 23.11.2011, Available from http://www.statsoft.com/products/.

Żabiński T., Mączka T. (2011). Implementation of Human-System Interface for Manufacturing Organizations. *Human-Computer Systems Interaction. Backgrounds and Applications 2*, Advances in Soft Computing, Springer-Verlag Co., 2011.

Żabiński T., Mączka T., Jędrzejec B. (2009). Control and Monitoring System for Intelligent Manufacturing – Hardware and Communication Software Structure. *Proceedings of Computer Methods and Systems,* p. 135-140, Kraków, Poland, November 26-27 2009.

10

Integrated Cellular Manufacturing System Design and Layout Using Group Genetic Algorithms

Michael Mutingi[1] and Godfrey C. Onwubolu[2]
[1]*National University of Singapore, Electrical & Computer Engineering*
[2]*School of Applied Technology, HITAL, Toronto*
[1] *Singapore*
[2]*Canada*

1. Introduction

Cellular Manufacturing System (CMS), an application of group technology philosophy, is a recent technological innovation that can be used to improve both productivity and flexibility in modern manufacturing environments (Signh, 93; Sarker and Xu, 1998). In practice, the essence of CMS is to decompose a manufacturing system into manageable autonomous subsystems (called manufacturing cells) so as to enhance shop-floor control, material handling, tooling, and scheduling. The decomposition process involves identification of part families with similar processes or design features and machine cells so that each family can possibly be processed in a single cell. In addition to this, machine layout within each cell is considered essential in order to improve efficiency and effectiveness of the overall production system. Consequently, setup times, work-in-process inventories, and throughput times are reduced significantly. The overall process of designing CMS involves the following four generic phases:

1. *Cell formation*: involves grouping of machines which can operate on a product family with little or no inter-cell movement of the products.
2. *Group layout*: includes layout of machines within each cell (intra-cell layout), and layout of cells with respect to one another (inter-cell layout).
3. *Group scheduling*: involves scheduling of parts for production
4. *Resource allocation*: assignment of tools, manpower, materials, and other resources

In general, the design of CMS includes three critical decisions, namely, cell formation, group layout, and scheduling. In the most ideal case, these criteria should be addressed simultaneously so as to obtain the best possible results (Kaebernick and Bazargan-Lari, 1996; Mahdavi and Mahadevan, 2008). However, due to the complex nature of the decision problem coupled with the limitations of conventional approaches, most of the cell formation studies have focused on these decisions independently or sequentially (Selim, 1998; Onwubolu and Mutingi, 2001). Most cell formation approaches proposed in literature use flow patterns of parts (sequence data) for cell design issues only. On the other hand, the layout designers did not consider the cell formation problem. Due to the fact that the sequential approach addresses the cell formation and the cell layout problem in a disjointed fashion, the quality of the final solution is often compromised. In this chapter, an integrated approach to cell formation and layout design is presented, based on available sequence data. The GGA-based approach utilizes sequence data to identify machine cells as well as machine layout within each cell. In this view, the major objectives for this chapter are as follows:

- to develop a GGA based methodology for solving the integrated CMS design and layout problem using sequence data, or flow patterns.
- to develop relevant performance metrics to address the integrated cell formation and layout problem.
- to make a comparative analysis between the GGA approach and other well known algorithms found in literature.

The next section describes the cell formation and cell layout problem. Section 3 briefly explains the general GA framework. A GGA approach is presented in section 4. Section 5 provides the results and discussion. Finally, section 6 concludes this chapter.

2. Cell formation and layout problem

The cell formation problem (CFP) in CMS involves grouping of machines which can operate on a product family with similar manufacturing processes and features such that little or no inter-cell movement of products is involved. The overall objective of cell formation is to gain the advantages inherent in the philosophy of group technology. In assessing the quality of solutions, various objectives are considered. These objective functions include the following;

i. Minimization of inter-cell movements;
ii. Minimization of intra-cell movements;
iii. Maximization of utilization;
iv. Minimization of material handling costs, and
v. Minimizing cell work-load imbalances

The cell layout problem involves layout machine within each cell and layout of cells with respect to one another. Recently, researchers have made efforts to utilize interval data and ordinal data, consisting of process sequence data which identifies the order in which jobs are processed (Nair and Narendran, 1998; Won and Lee, 2001; Jayaswal and Adil, 2004). The application of sequence data in CMS has received little attention in the research community and in industry. Sequence data provides useful information on flow patterns of jobs in a manufacturing system. As such, sequence data is useful not only in identifying part family and machine groups but also the actual layout of machines within each cell, based on flow patterns. Earlier studies focused on the use of zero-one machine-component incidence matrix as the input data for the cell formation problem. However, the joint CFP and the layout problem are often treated independently in literature. In an attempt to jointly address the CF and the layout problem, solution methods from various researchers and practitioners often utilize a sequential approach. In this approach, cells are formed first, followed by intra-cell layout construction. Since the final solution is largely dependent on the initial cell formation, the quality of the final solution is often compromised.

The joint cell formation and layout problem is a new approach that seeks to identify manufacturing cells and the layout (sequence) of machines in the cells in an integrated manner. The whole aim of the approach is to avoid compromising the quality of solutions with respect to cell formation and cell layout objectives. Therefore, this approach to the joint layout problem is of practical value. The basic cell formation problem is NP-complete, meaning that it has no known polynomial time algorithm due to its combinatorial nature (Kumar *et al*, 1986). It follows that the integrated cell formation and cell layout problem is highly computationally intractable. In this respect, the use of heuristic approaches such as simulated annealing, tabu search, and genetic algorithms, is quite appropriate. Simulated annealing is a probabilistic meta-heuristic method proposed in Kirkpatrick, Gelatt and Vecchi (1983) and in Cern (1985) for finding the global minimum of a cost function that may possess several minima. It works by emulating the physical process whereby a solid is slowly cooled down so that when its structure eventually frozen, this occurs at a minimum energy configuration. Tabu Search is a meta-heuristic local search algorithm created in Glover and McMillan (1986) for solving combinatorial optimization problems. It uses the concept of a local or neighbourhood search to iteratively move from one potential solution x to an improved solution x' in the neighbourhood of x, until some stopping criterion has been satisfied, usually an attempt limit or a score threshold (Glover, 1989; Glover, 1990).

3. Genetic algorithms

Genetic algorithm (GA), originated by Holland (1975), is a meta-heuristic approach based on evolutionary principles of natural selection and survival of the fittest. The GA methodology has been applied extensively in a wide range of combinatorial problems in engineering, business, manufacturing, agriculture, telecommunications and sciences (Gen and Cheng; Goldberg, 1989; Man *et al*, 1999). The method integrates the elements of stochastic and direct search to obtain optimal (or near-optimal) solutions within reasonable computation time. GA attempts to evolve a population of candidate solutions by giving preference of survival to quality solutions, whilst allowing some low quality solutions to survive in order to maintain a level of diversity in the population. This process enables GA to provide good solutions so as to avoid premature convergence. Each candidate is coded into a string of digits, called chromosomes. New offspring are obtained from probabilistic operators, mainly crossover and mutation. Comparison of new and old (parent) candidates is done based on a given objective or fitness function so as to retain the best performing candidates into the next population. In this process, characteristics of candidate solutions are passed from generation to generation through probabilistic selection, crossover, and mutation actuated in the population of candidate solutions.

The general GA framework can be represented as follows:

BEGIN
    Initialize *population* with random candidate solutions;
    Evaluate each candidate;
REPEAT
    Select parent chromosomes;
    Recombine pairs of parents;
    Mutate the resulting offspring;
    Evaluate new candidates;
    Select individuals for next generation
UNTIL (Termination condition is satisfied**)**

Figure 1 shows the general flow of the genetic algorithm.

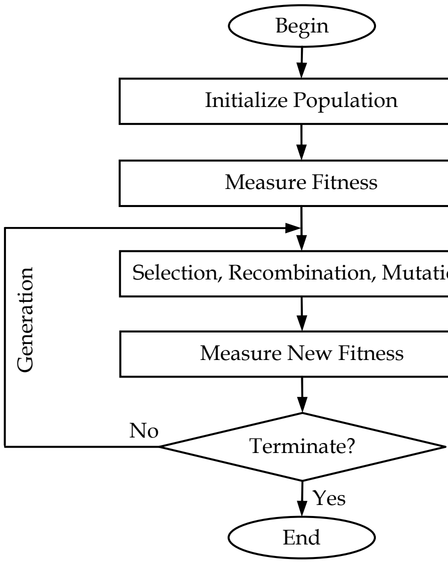


Fig. 1. Genetic algorithm framework

Genetic algorithm offers unique advantages over other stochastic searches, population-based search, including implicit parallelism, independence from gradient information, and flexibility to hybridization with other heuristics. Early applications of the GA approach to the cell formation problem include the work by Venugopal and Narendran (1992) based on minimization of cell load variation and inter-cell moves. Other applications were done by Gravel et al. (1998), and Hsu and Su (1998). However, Falkenauer (1992) realized several significant shortcomings of using classical GAs for grouping problems. Falkenauer (1998) pointed out that though attempts have been made to minimize the drawbacks associated with applying GAs to grouping problems by use of specialized genetic operators, this still result in various shortcomings. In this view, Falkenauer (1992) introduced a grouping genetic algorithm, designed to handle the special structure of grouping problems. Group genetic algorithm (GGA) is a modification of conventional GA designed specifically for clustering/grouping problems. In the next section, an enhanced GGA approach is proposed for the machine cell formation and layout problem.

4. A group genetic algorithm approach

Grouping genetic algorithm (GGA) combines specifically designed operators for grouping problems with the power of local search in order to refine new chromosomes generated. Therefore, GGA is a preferable approach over other heuristic and conventional approaches. The design of the proposed GGA for the joint cell design and layout problem is presented, based on its six main building blocks, namely:

i. Fitness/objective function
ii. Chromosome coding scheme
iii. Initial population generation
iv. Selection and recombination
v. Group genetic operators: crossover, mutation and inversion
vi. Genetic parameters

The next sections elaborate on these building blocks.

4.1 Objective/fitness function formulations

From the CMS design perspective, the existence of voids and exceptions should be minimized. In layout design, adjacency of machines in a cell is a key factor as it can reduce material handling costs significantly (Mahdavi and Mahadevan, 2008). From production planning perspective, the sequence in which machines are placed in cells may create unwanted reverse flows and skipping of workstations. For instance, a cell with machines 1 and 2 has two possible sequences (layouts), i.e., [1, 2] or [2, 1]. From Table 1, it can be seen that the cell layout [2, 1] has only one consecutive forward flow, while layout [1, 2] has four. From this analysis, layout [1, 2] is preferred.

	Parts										
	2	1	4	5	3	7	6	8	9	11	10
Machines											
2	2	2	3	2	2	2					
1	1	1	1	3	1	1					
5							1	1	1	2	1
3			2	1			3	2	2	3	2
4							2	3	3	1	3

Table 1. A typical solution for a cell formation problem

Ideally, a good objective function should be able to capture and evaluate the effects of the sequence of machines within each cell. A simplified way of evaluating the fitness of a cell layout is to express the objective function in terms of the number of consecutive forward flows. In this connection, Mahdavi and Mahadevan (2008) defined the cell flow index (CFI) and the overall flow index (OFI) for evaluating the performance of cell design and cell layout solutions.

The following notation is used in this model.

n number of parts in the system
m number of machines in the system
n_c number of parts in cell c
m_c number of machines in cell c
v_c number of voids in cell c
N_{fc} number of consecutive forward flows within cell c
S_{jk} machine-component matrix $[s_{jk}]$; $s_{jk} = 1$ if part k visits machine k, and 0 otherwise

In order to determine the average flow and overall flow performance measures, the total number of operations and the consecutive flows between a pair of machines are calculated. The total number of flows N_{flow} is:

$$N_{flow} = \sum_{k} \max_{j} s_{jk} - n \tag{1}$$

The total number of flows in each cell *c* is determined as follows:

$$N_{tc} = (n_c m_c) - v_c - n_c \tag{2}$$

4.1.1 Cell flow index (CFI)

The cell flow index for cell *c*, CFI_c is the ratio of the number of consecutive forward flows to the total number of flows within the cell. The cell average flow index is the weighted average of CFIs. This is further explained in the following expressions;

$$\text{CFI}_c = \frac{N_{fc}}{N_{tc}} \tag{3}$$

Therefore, the average cell flow index, ACFI is

$$\text{ACFI} = \left(\frac{1}{n}\right) \cdot \sum_{c} n_c \text{CFI}_c \tag{4}$$

It is clear from the above analysis that as the number of voids in the cell decreases and as the number of consecutive forward flows increases, the CFI measure increases. This indicates that the CFI represents the solution quality with respect to the number of voids and the intra-cell moves. Therefore, a combination of these performance measures ensures that the cell formation and layout are addressed jointly.

4.1.2 Overall cell flow index (OFI)

The OFI performance measure defines the ratio of the sum of consecutive forward flows in all the cells to the total number of the flows required to process all the parts. This can be expressed as follows;

$$\text{OFI} = \left(\frac{1}{N_{nflow}}\right) \cdot \sum_{c} N_{fc} \tag{5}$$

Expression (5) shows that the overall cell flow index defines the extent of inter-cell moves (exceptions); increasing values of OFI can be obtained by decreasing values of inter-cell moves. While the OFI points to the inter-cell movements, the ACFI addresses the intra-cell movements.

4.2 Solution encoding – chromosome representation

The GGA's performance strongly depends on the type of the coding scheme, that is, the chromosome (string) representations used. Effective coding schemes can improve the search efficiency and quality. Most of the coding schemes in literature used strings of integer numbers to where the position of the number represents the machine and the value of the number identifies the cell number. For example, a typical chromosome (2 3 1 1 2 3 1 1 2) containing 9 machines represents a manufacturing system with 3 cells. Machines 1, 5, 6 and 8 are in cell 1, machines 2 and 3 are in cell 2, and machines 4, 7, 9 occupy cell 3.

Machine position:	1 2 3 4 5 6 7 8 9
Chromosome :	2 3 1 1 2 3 1 1 2

The proposed GGA algorithm has an improved coding scheme, similar to the one proposed Filho and Tibert (2006). The coding scheme improves the utilization of the group structure by using a group structure for each feasible string based on three code schemes as shown in Figure 2. The first, code 1, is a string of size *m*, where *m* represents the total number of machines in the system. The second is a group structure upon which the genetic operators act, while the third represents the positions of the last nodes of each group.

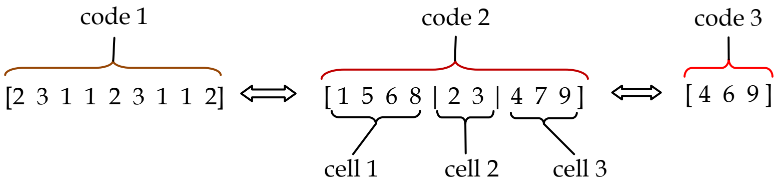


Fig. 2. Chromosome representation

It can be seen from code 3 in Figure 2 that cell 1 consists of the first four genes in code 2. Similarly, cell 2 is made up of the next two genes. Finally, cell 3 is composed of the last three genes in code2. Several features are enhanced in the implementation of the GGA structure, such as in formulation of objective/fitness functions, the genetic operators, chromosome repair and other genetic strategies.

4.3 Initial population

An initial population of the desired size, *popsize*, is randomly generated from the solution space. Consider a typical problem consisting of *m* machines and a predetermined number of cells, *v*. Assume that each cell comprises at least two machines. Then, the initial population is created according to the following procedure:

Repeat

1. For each cell *j* (*j*=1,...,*v*), randomly select two machines from the set of machines.
2. For the remaining (*n*-2*j*) unassigned machines, randomly assign a machine to a cell, until all machines are assigned.
3. Encode the chromosome using code 1 and add to the initial population.

Until (population size *popsize* is achieved).

In GGA application, the goal is to minimize some cost function which is usually mapped to a score function which is used to evaluate the generated chromosomes. A mapping procedure initially suggested by Goldberg (1989) is applied as follows;

$$f^{i}(t)=\begin{cases} f^{i}_{\max}-g^{i}(t) & \text{if } g_{i}(t)<f^{i}_{\max} \\ 0 & \text{if otherwise} \end{cases} \tag{6}$$

where, $g(t)$ is the objective function of a chromosome and $f_{\max}$ is the largest objective function in the current population.

4.4 Selection strategy

Several selection strategies have been suggested by Goldberg (1989), such as deterministic sampling, remainder stochastic sampling with/without replacement, stochastic tournament, and stochastic sampling with/without replacement. The remainder stochastic sampling without replacement has been found to be the most effective and is applied in this work (Goldberg, 1989). In this strategy, each chromosome i is selected and stored in the mating pool according to the expected count e_i calculated as,

$$e_{i}=\frac{f_{i}}{\left(1/popsize\right)\sum_{i=1}^{s}f_{i}} \tag{7}$$

Where, *popsize* is the desired population size and f_i is the score function value of the i^{th} chromosome.

Each chromosome receives copies equal to the integer part of e_i, that is, $[e_i]$, while the fractional part is treated as success probability of obtaining additional copies of the same chromosome into the mating pool.

4.5 Genetic operators

In this section, design issues relating to the development of the proposed GGA approach for the manufacturing cell design problem are defined. Unique crossover, mutation and inversion strategies are developed for the GGA algorithm.

4.5.1 Crossover

Crossover is a probabilistic evolutionary mechanism which seeks to mate chromosomes, chosen by the selection strategy, in order to produce a pool of new offsprings, called *selection pool*. It allows the algorithm to generate new solutions and to explore unvisited regions in the solution space. The proposed crossover, called group crossover operator, exchanges groups of genes of selected chromosomes. The crossover operation occurs with probability *prcoss* until the desired pool size, $poolsize = popsize \cdot pcross$, is obtained. The procedure for the group crossover operator is as follows:

Repeat

1. Generate a random integer number between 1 and (v-1), where v is number of cells. This number defines the crossover point.

2. Swap the groups to the right of the crossover point to generate two offspring.
3. Repair the offspring by eliminating any duplicated machines and introducing missing machines.

Until (selection *poolsize* is achieved).

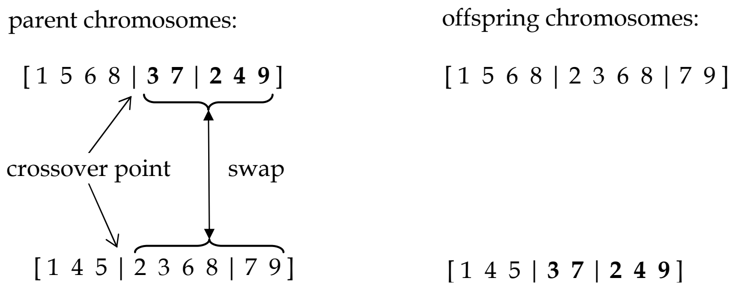


Fig. 3. Crossover operator

In the crossover process, some machines may appear in more than one cell, and some may be missing. Such offspring should be repaired. The repair procedure identifies duplicated machines and eliminates those to the left of the crossover point. Missing machines are inserted on the cell with the least number of nodes. Thus, the group representation scheme enhances the crossover operator by taking advantage of the group structure.

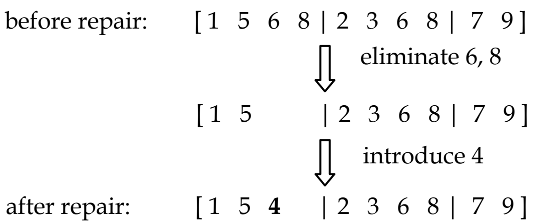


Fig. 4. Chromosome repair procedure

4.5.2 Mutation

The mutation operator is applied to every new chromosome in order to maintain diversity of the population and avoid premature convergence. Two mutation operators are proposed, namely *swap mutation* and *shift mutation*. The swap mutation operates by swapping genes between two randomly chosen groups in a chromosome (see Figure 5.). Its general procedure can be summarized as follows:

1. Randomly select two integer numbers from the set {1,2, ..., v}, where v is the number of cells or groups.
2. Randomly choose a gene from each group

3. Swap the selected genes, exchanging their values

		swap
offspring chromosome	:	[1 5 4 | 2 3 6 8 | 7 9]
select group or cell	:	1 and 3
select genes or machines	:	genes **4** and **7**
mutated offspring	:	[1 5 **7** | 2 3 6 8 | **4** 9]

Fig. 5. Swap mutation

The *shift mutation operator* works by shifting the frontier between two adjacent groups by one step either to the right or to the left, as shown in Figure 6. Essentially, the number of nodes is increased in one group and simultaneously decreased in the other. The procedure for the mutation operator is summarized thus;

1. Generate a random integer number between 1 and (v-1). Let this number represent the chosen frontier.
2. Randomly choose the direction of shift: *right* or *left*.
3. Shift the frontier in the selected direction, thereby moving one node between adjacent groups.

offspring chromosome	:	[1 5 4 **|** 2 3 6 8 | 7 9]
select frontier, rand (1,2)	:	1 → shift frontier
select direction	:	right
mutated offspring	:	[1 5 4 **2** **|** 3 6 8 | 7 9]

Fig. 6. Shift mutation operator

4.5.3 Inversion operator

In order to curb premature convergence and control diversity level of the population, an inversion operator is designed. The inversion operator is applied, at a very low probability, on chromosomes selected by the selection strategy prior to crossover operation. Basically, the inversion strategy operates by rearranging the groups in the reverse order, for instance, the order of cells [1, 2, 3] is transformed to a new [3, 2, 1]. This procedure is further illustrated in Figure 7.

chromosome before inversion	:	[1 5 6 8 | 3 7 | 2 4 9]
chromosome after inversion	:	[2 4 9 | 3 7 | 1 5 6 8]

Fig. 7. Inversion operator

4.5.4 Diversification

In GGA application it is observed that as the iterations proceed, the solution space (population) converges to a particular solution. However, rapid loss of diversity and premature convergence may occur before an optimal solution is obtained; a problem called genetic drift. To track the diversity of the solution space, Grefenstette (1987) proposed an entropic measure H_i in a population of candidates. For each machine i, the H_i can be defined for GGA in this form;

$$H_i = \sum_{j=1}^{m} \frac{(n_{ij}/p) \cdot \log(n_{ij}/p)}{\log(m)} \tag{8}$$

Where n_{ij} is the number of strings in which machine i is assigned position denoted by j in the current population, p is the solution space size, and m is the number of machines. Divergence H is calculated as;

$$H = \sum_{i=1}^{m} H_i / m \tag{9}$$

As the iterations proceed, the divergence parameter H approaches zero. Thus, the diversity of the solution space can be monitored and controlled by applying the inversion operator till diversity improves to a preset value. In order to prevent loss of good solutions, a fraction (e.g., 0.2) of best performing solutions from the undiversified population is preserved. Performing candidate solutions from the diversified population are compared with those from the diversified population, preferring those that fair better. Thus, the best performing candidates are taken into the next generation.

4.6 The group genetic algorithm implementation

The structure of the proposed GGA for solving the integrated cellular manufacturing system problem was developed incorporating the group operators described in previous sections. A multi-objective approach is adopted in this application based on the two performance measures, ACFI and OFI. The overall GGA structure is now summarised as follows:

Step 1. *Input*: initial data input:

i. Select the typical initial GGA parameter values (see Table 2)
ii. Input the manufacturing data, with sequence data

Step 2. *Initial population*: create randomly, two initial populations, called old populations, *oldpop*1 *and oldpop*2.

Step 3. *Selection and recombination*: Select chromosomes using stochastic sampling without replacement

i. Evaluate strings by objective function, fitness function and expected count
ii. create two temporal population, *temppop*1 and *temppop*2 using the integer parts of expected count and fractional parts as success probabilities

Step 4. *Crossover/recombination*: Apply the group crossover to *temppop*1 and *temppop*2 to create a two selection pool populations, *spool*1 and *spool*2.

i. Select two candidates for crossover using remainder selection without replacement, one from *temppop*1 and another from *temppop*2.
ii. Apply crossover operator to the two strings
iii. If crossover is successful, apply inversion operator, otherwise go to step 5
iv. Apply repair mechanism if necessary

Step 5. *Mutation*: apply mutation operators to the two offspring and move them to new population

Step 6. *Replacement strategy*: Replace old populations with corresponding new populations

i. Compare corresponding chromosomes successively in each selection pool and old population
ii. Take the one that fares better in each comparison
iii. For the rest of the offspring, selection with probability 0.555

Step 7. *Diversification*: Diversify population by applying the mutation operator if mutation falls below a predetermined minimum

i. Calculate diversity H, of the population
ii. If the acceptable diversity H_a is such that $H<H_a$ then diversify until diversity is acceptable.
iii. re-evaluate chromosomes in terms of fitness functions, defined by ACFI and OFI

Step 8. *New population*: Check the current generation count, *gen* against maximum generation count *maxgen*.

i. If $gen < maxgen$ then go to Step 3, otherwise stop
ii. Return the best solutions

GGA Parameter	Variable	Value
Number of generations	*maxgen*	Variable
Population size	*popsize*	10-40
Crossover probability	*crossprob*	0.4 – 0.7
Mutation probability	*mutprob*	0.02 – 0.3
Inversion probability	*invprob*	0.04 – 0.2
Chromosome size	*chrom*	Number of machines

Table 2. Typical GGA genetic parameters

Part families are identified based on the number of operations required by a part in a cell. Therefore, a part is assigned to a cell where it requires the maximum number of operations (or machines) for its processing.

5. Results and discussion

The proposed GGA approach was implemented in Java SE 7. An illustration of the GGA execution is first given. A comparative analysis on of the performance of the proposed approach with other algorithms is then presented based on computational analysis on known published data sets.

5.1 GGA computational analysis

This section first provides a numerical illustration obtained when executing the GGA algorithm on well known problem data sets in literature. The set of input data used in this illustration is found in Nair and Narendran (1998). The design and layout problem consists of 25 machines and 40 parts (a 25 x 40 problem). Figure 8 shows an illustration of the intermediate stages arrived at as the algorithm solves design and layout problem. The objective function represents the ACFI and the OFI objective values. The input number of cells used for the simulation run was four. The results of the simulation run show that the ACFI values increased from 20% to 68% after 40 iterations, while the OFI values rose from 21% to 42% after 25 iterations.

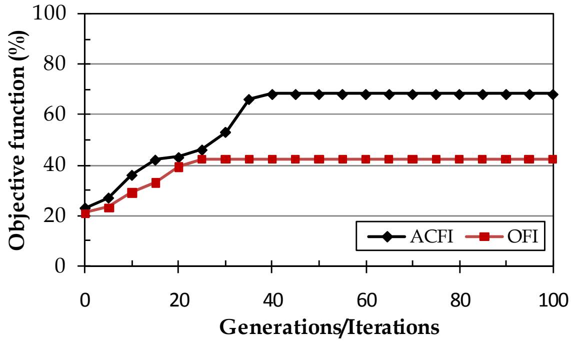


Fig. 8. GGA objective function for Nair & Narendran (1998) 25 x 40 problem

Further numerical experiments were carried out based on an 8 x 20 problem obtained from Nair and Narendran (1998), as shown in Table 3. With a typical set of input data for genetic parameters, the final solution from the GGA simulation run is an improved version of the Nair and Narendran (1998) problem. Table 4 provides the improved solution to the problem. Furthermore, a summary of the final improved version of the solution is provided in Table 5.

	Parts																			
	2	8	9	11	13	14	16	17	19	3	4	6	7	18	20	1	5	10	12	15
Machines																				
3	2	2	3	2	2	3	2	1	2											
1	1	1	1	3	1	1	1	3	1	2										
4										5	2	2	2	1	1			2		
7				1						3	3	3	3	4	4				2	
8										4	4	4	1	3	5					
2						2				1	1	1	4	2	2					
5								2				5				2	2	3	1	1
6			2												3	1	1	1	3	2

Table 3. Solution from Nair and Narendran (1998) – 8 x 20 problem

The machine cells obtained by the GGA approach are the same as those obtained from the CASE algorithm in Nair and Narendran (1998) and those obtained from the CLASS algorithm in Mahdavi and Mahadevan (2008). Similar to the results from the CLASS algorithm, GGA obtained an improved sequence of machines based on the use of sequence data, showing a remarkable improvement in the layout of machines within cells. In this respect, the GGA approach is effective when compared to well known algorithms in literature. Thus, the algorithm is able to simultaneously address the cell formation and the cell layout problems effectively within a reasonable computation time.

	Parts																			
	2	8	9	11	13	14	16	17	19	3	4	6	7	18	20	15	1	5	10	12
Machines																				
1	1	1	1	3	1	1	1	3	1	2										
3	2	2	3	2	2	3	2	1	2											
2						2				1	1	1	4	2	2					
4										5	2	2	2	1	1				2	
7				1						3	3	3	3	4	4					2
8										4	4	4	1	3	5					
6			2												3	2	1	1	1	3
5								2				5				1	2	2	3	1

Table 4. New solution of Nair and Narendran (1998) using GGA – 8 x 20 problem

Cell	Machines	Parts
C1	1, 3	2, 8, 9, 11, 13, 14, 16, 17, 19
C2	2, 4, 7, 8	3, 4, 6, 7, 18, 20
C3	6, 5	1, 5, 10, 12, 15

Table 5. Final improved solution from Nair and Narendran (1998) problem 8 x 20 problem

Cell No.	CASE Solution				CLASS Solution				GGA Solution			
	n_c	N_{fc}	N_{tc}	CFI%	n_c	N_{fc}	N_{tc}	CFI%	n_c	N_{fc}	N_{tc}	CFI%
1	9	1	9	11.1	9	5	9	55.6	9	5	9	55.6
2	6	7	18	38.9	6	9	18	50	6	9	18	50
3	5	1	5	20.0	5	2	5	40.0	5	2	5	40.0
N_{flow} = 41												
ACFI (%)				21.0				50.0				50.0
OFI (%)				22.0				39.0				39.0

Table 6. Comparative study of GGA, CASE and CLASS algorithms - 8 x 20 problem

In order to demonstrate the utility of the proposed GGA algorithm, a comparative study was done with GGA, CASE and CLASS algorithms. Table 6 provides the results of the comparative analysis. It can be seen from this analysis that though machine groups and part families are the same for the three algorithms, the ACFI and OFI differ with the CASE solution. However, the ACFI and OFI values of GGA are similar to those obtained from CLASS. This shows a remarkable improvement of the solution to the joint cell formation and layout problem.

The next section provides a comparative analysis of the performance of GGA approach and other algorithms found in literature.

5.2 Comparison of GGA with other algorithms

In order to gain more understanding on the effectiveness of the GGA, further comparative experiments were carried out based on data sets reported in literature including Tam (1988), Harhalakis et al. (1990), and Nair & Narendra (1998). Park and Suresh (2003) made a comparative study of known algorithms on sequence data. Algorithms such as fuzzy ART neural network and conventional clustering methods were compared. In addition to these algorithms, other approaches such as CASE designed by Nair and Narendran (1998) and CLASS originated by Mahdavi and Mahadevan (2008) are included in the comparative study. Therefore, the performance of GGA can sufficiently be analyzed based on these known data sets and algorithms. The results obtained in this comparative study are shown in Table 7.

Data set	Size	CLASS			Fuzzy Art			Hierarchical			GGA		
		Cells	ACFI	OFI	Cells	ACFI	OFI	Cells	ACFI	OFI	Cells	ACFI	OFI
1.	12 X 19	2	65%	50%	2	49%	36%	2	48%	45%	2	65%	50%
2.	20 x 20	4	65%	41%	4	42%	34%	4	42%	34%	4	69%	43%
3.	25 X 40	4	52%	34%	7	38%	27%	8	37%	22%	4	68%	42%
4.	08 x 20	3	50%	39%							3	50%	39%

Key: 1. Tam (1988); 2. Harhalakis et al. (1990); 3. Nair & Narendra (1998); 4. Nair & Narendra (1998);

Table 7. A comparison of GGA with other approaches

In all cases, the ACFI and OFI values obtained by GGA are much more preferable than those obtained from other algorithms. From this analysis, it can be seen that the utilization of sequence data in joint cell design and layout is important.

6. Conclusions

Integrated cellular manufacturing system design and layout is an important but hard and complex decision process that involves two main problems; cell formation and machine layout within each cell. Sequence data provides additional information on the dominant flow patterns in cells, which forms the basis for solving the integrated layout problem. However, sequence data has not been fully utilised in manufacturing cell design. The main challenge, therefore, is the extension of the application of sequence data and the development of a robust meta-heuristic algorithm for solving the joint design and layout problem.

In this chapter, a GGA meta-heuristic approach was proposed to solve the integrated layout design problem based on sequence data. The proposed GGA meta-heuristic has unique enhanced features, including a group chromosome scheme, a group crossover operator, a group mutation operator, and a chromosome repair mechanism. The group operators enable the algorithm to reveal the group structure inherent in a data set, producing comparably good quality solutions. While crossover operator enhances exploration of unvisited points in the potential solution space, the mutation exploitation of the best solution in the near-optimal space. Although increasing the number of cells and/or machines may demand more iterations/generations before the algorithm converges to a good solution, the number of parts has no effect on the solution space when grouping machines. Moreover, the parallel mechanism of the approach gives the algorithm robustness and effectiveness over a variety of ill-structured input matrices. Thus, the algorithm is quite preferable in problem situations with a large number of parts.

Comparison with known algorithms in literature was done using known data sets. Apart from well-known performance measures, the average cell flow index was included as a performance parameter, which is a measure of the average magnitude of consecutive forward flows. This measure enabled the GGA approach to evaluate and solve the cell formation and layout design problem in an integrated fashion. The computational results in this study show the utility of the enhanced GGA approach.

Prospects for further research and application of the proposed GGA approach may be interesting. For instance, the group genetic algorithm can be extended to similar clustering problem domains, scheduling problems, as well as network design problems. Further research in these areas is worth exploring.

7. References

Cern, V. (1985). A thermodynamic approach to the travelling salesman problem: An efficient simulation. *Journal of Optimization Theory and Applications*, Vol. 45, 41-51.

Falkenauer E. (1992). The grouping genetic algorithms - widening the scope of the GAs. *Belgian Journal of Operations Research, Statistics and Computer Science*, Vol. 33, 79-102.

Falkenauer, E., 1998, *Genetic Algorithms for Grouping Problems,* New York, Wiley.

Filho, E.V.G., and Tibert, A.J. 2006. A group genetic algorithm for the machine cell formation problem. *International Journal of Production Economics.*

Gen, M. and Cheng, R. (1997). *Genetic Algorithms and Engineering Design*, Wiley Interscience Publication, MA, 1997.

Glover, F. (1989). Tabu Search – Part 1. *ORSA Journal of Computing*, Vol. 1 (2), 190-206.

Glover, F. (1990). Tabu Search – Part 2. *ORSA Journal of Computing*, Vol. 2, No.1, 4-32.

Glover, F., McMillan, C. (1986). The general employee scheduling problem: an integration of MS and AI. *Computers and Operations Research*, Vol. 13, No. 5, 563-573.

Goldberg, D. E. (1989). *Genetic Algorithms: In Search, Optimization & Machine Learning*, Addison-Wesley, Inc., MA, 1989.

Gravel, M., Nsakanda, A.L., Price, W., 1998. Efficient solutions to the cell-formation problem with multiple routings via a double-loop genetic algorithm. *European Journal of Operations Research*, Vol. 109, 286-298.

Grefenstette, J.J. (1987). Incorporating problem specific knowledge into genetic algorithms. In L. Davis, *Genetic and Simulated Annealing*, London: Pitman.

Holland, J. H. (1975). *Adaptation in Natural and Artificial System*, University of Michigan Press, Ann Arbor, MI, 1975.

Hsu, C.M., and Su, C.T. (1998). Multi-objective machine-component grouping in cellular manufacturing: a genetic algorithm approach. *Production Planning & Control*, Vol. 9, No. 2, 155-166.

Jayaswal, S. and Adil, G. K. (2004). Efficient algorithm for cell formation with sequence data, machine replications and alternative process routings. *International Journal of Production Research*, Vol. 42, 2419-2433.

Kaebernick H. and Bazargan-Lari, M. (1996). An integrated Approach to the Design of Cellular Manufacturing. *Annals of the CIRP*, Vol. 45, No. 1, 421-425.

Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P., (1983). Optimisation by simulated annealing. *Science*, Vol. 220, 621-630.

Kumar, K. R., Kusiak, A. and Vaneli, A. (1986). Grouping parts and components in FMS. *European Journal of Operational Research*, Vol. 24387-24400.

Mahdavi, I. and Mahadevan, B. (2008). CLASS: An algorithm for cellular manufacturing system and layout design using sequence data. *Robotics and Computer-Integrated Manufacturing*, Vol. 24, 488-497.

Man, K. F., Tang, K. S. and Kwong, S. (1999). *Genetic Algorithms: Concepts and Design*, Springer, London, 1999.

Nair, G. J. and Narendran, T. T. (1998). CASE: a clustering algorithm for cell formation with sequence data. *International Journal of Production Research*, Vol. 36, 157-179.

Onwubolu, G.C. and Mutingi, M. (2001). A genetic algorithm approach to cellular manufacturing systems. *Computers & Industrial Engineering*, Vol. 39, 125-144.

Sarker B. R., and Xu, Y. (1998). Operation sequences-based cell formation methods: a critical survey. *Production Planning & Control*, Vol. 9, 771--783.

Selim, H. M., Askin, R. G., and Vakharia, A. J. (1998). Cell formation in group technology: Evaluation and directions for future research. *Computers & Industrial Engineering*, Vol. 34, No. 1, 3--20.

Singh N. (1993). Design of cellular manufacturing systems: an invited review. *European Journal of Operational Research*, Vol. 69, 284--291.

Suresh NC, Slomp J, KAparth S., 1999. Sequence-dependent clustering of parts and machines: a Fuzzy ART neural network approach. *International Journal of Production Research*, Vol. 37, 2793-816

Venugopal, V., Narendran, T.T., 1992. A genetic algorithm approach to the machine-component grouping problem with multiple objectives. *Computers and Industrial Engineering*, Vol. 22, No. 4, 469-480.

Won, Y. and Lee, K. C. (2001). Group technology cell formation considering operation sequences and production volumes. *International Journal of Production Research*, Vol. 39, 2755-2768.

11

Hybrid Manufacturing System Design and Development

Jacquelyn K. S. Nagel[1] and Frank W. Liou[2]
[1]*James Madison University*
[2]*Missouri University of Science and Technology*
USA

1. Introduction

Reliable and economical fabrication of metallic parts with complicated geometries is of considerable interest for the aerospace, medical, automotive, tooling, and consumer products industries. In an effort to shorten the time-to-market, decrease the manufacturing process chain, and cut production costs of products produced by these industries, research has focused on the integration of multiple unit manufacturing processes into one machine. The end goal is to reduce production space, time, and manpower requirements. Integrated systems are increasingly being recognized as a means to meet these goals. Many factors are accelerating the push toward integrated systems. These include the need for reduced equipment and process cost, shorter processing times, reduced inspection time, and reduced handling. On the other hand, integrated systems require a higher level of synthesis than does a single process. Therefore, development of integrated processes will generally be more complex than that of individual unit manufacturing processes, but it could provide simplified, lower-cost manufacturing.

Integrated systems in this research area have the ability to produce parts directly from a CAD representation, fabricate complex internal geometries, and form novel material combinations not otherwise possible with traditional subtractive processes. Laser metal deposition (LMD) is an important class of additive manufacturing processes as it provides the necessary functionality and flexibility to produce a wide range of metallic parts (Hopkinson et al., 2006; Liou & Kinsella 2009; Venuvinod & Ma, 2004). Current commercial systems that rely on LMD to produce tooling inserts, prototype parts, and end products are limited by a standard range of material options, building space, and a required post-processing phase to obtain the desired surface finish and tolerance. To address the needs of industry and expand the applications of a metal deposition process, a hybrid manufacturing system that combines LMD with the subtractive process of machining was developed achieving a fully integrated manufacturing system.

Our research into hybrid manufacturing system design and development has lead to the integration of additive and subtractive processes within a single machine footprint such that both processes are leveraged during fabrication. The laser aided manufacturing process (LAMP) system provides a rapid prototyping and rapid manufacturing infrastructure for research and education. The LAMP system creates fully dense, metallic parts and provides all the advantages of the commercial LMD systems. Capabilities beyond complex geometries and good surface finish include: (1) functional gradient material metallic parts where different materials are added from one layer to the next or even from one section to another, (2) seamlessly embedded sensors, (3) part repair to reduce scrap and extend product service life, and (4) thin-walled parts due to the extremely low processing forces (Hopkinson et al., 2006; Liou et al., 2007; Ren et al., 2008). This hybrid system is a very competitive and economical approach to fabricating metallic structures. Hybrid manufacturing systems facilitate a sustainable and intelligent production model and offer flexibility of infrastructure to adapt with emergent technology, customization, and changing market needs (Westkämper, 2007). Consequently, the design strategies, system architecture, and knowledge required to construct hybrid manufacturing systems are vaguely described or are not mentioned at all in literature.

The goal of this chapter is to summarize the key research findings related to the design, development, and integration of a hybrid manufacturing process that utilizes LMD to produce fully dense, finished metallic parts. Automation, integration, and control strategies along with the associated issues and solutions are presented as design guidelines to provide future designers with the insight needed to successfully construct a hybrid system. Following an engineering design perspective, the functional and process knowledge of the hybrid system design is explored before physical components are involved. Key results are the system architecture, qualitative modeling, and quantitative modeling and simulation of a hybrid manufacturing process.

In summary, this chapter provides an interdisciplinary approach to the design and development of a hybrid manufacturing system to produce metal parts that are not only functional, but also processed to the final desired surface-finished and tolerance. The approach and strategies utilized in this research coalesce to facilitate the design of a sustainable and intelligent production system that offers infrastructure flexibility adaptable with emergent technologies and customizable to changing market needs. Furthermore, the approach to hybrid system design and development can assist in general with integrated manufacturing systems. Applying the strategies to design a new system or retrofit older equipment can lead to increased productivity and system capability.

2. Related work

Any process that results in a solid physical part produced directly from a 3D CAD model can be labeled a rapid prototyping process (Kalpakjian & Schmid, 2003; Venuvinod & Ma, 2004). Equally, a process that converts raw materials, layer-by-layer into a product is a rapid prototyping process, but is typically referred to as additive manufacturing or layered manufacturing. Subtractive manufacturing is the process of incrementally removing raw material until the desired dimensions are met. Where additive processes start from the ground up, subtractive processes start from the top down. The combination of manufacturing processes from different processing categories establishes a hybrid manufacturing system. Herein, a hybrid manufacturing system refers to a manufacturing system that is comprised of an additive and subtractive manufacturing process.

Both additive and subtractive manufacturing cover a wide range of fabrication processes. For example, additive manufacturing can involve powder-based (e.g., selective laser sintering), liquid-based (e.g., stereolithography) or solid-based (e.g., fused deposition modeling) processes, each using a wide range of materials (Gebhardt, 2003; Kai & Fai, 1997; Venuvinod & Ma, 2004). While traditional subtractive manufacturing is typically reserved for metals, advanced or non-conventional subtractive processes have emerged to handle a greater variety of materials which include electric discharge machining, water jet cutting, electrochemical machining and laser cutting (Kalpakjian & Schmid, 2003). The physical integration of additive and subtractive manufacturing processes, such as laser metal deposition and machining, is the key to leveraging the advantages of each process. The vast domains of additive and subtractive manufacturing have provoked many to test boundaries and try a new concept, in an attempt to discover the next best system that will play a key role in advancing manufacturing technologies. Academic and industry researchers alike have been developing novel, hybrid manufacturing systems, however, the design and integration strategies were not published. On the other hand, a few approaches taken to develop reliable hybrid systems that deliver consistent results, with the majority based on consolidation processes, have published a modest guide to their system design. In following paragraphs, a number of hybrid manufacturing systems are reviewed to give an idea of what has been successful.

Beam-directed technologies, such as laser cladding, are very easy to integrate with other processes. Most have been integrated with computer numerically controlled (CNC) milling machines by simply mounting the cladding head to the z-axis of the milling machine. Kerschbaumer and Ernst retrofitted a Röders RFM 600 DS 5-axis milling machine with an Nd:YAG laser cladding head and powder feeding unit, which are all controlled by extended CNC-control (Kerschbaumer & Ernst, 2004). Similarly, a Direct Laser Deposition (DLD) process utilizing an Nd:YAG laser, coaxial powder nozzle and digitizing system as described by (Nowotny et al., 2003) was integrated into a 3-axis Fadal milling machine. Laser-Based Additive Manufacturing (LBAM) as researched at Southern Methodist University, is a technique that combines an Nd:YAG laser and powder feeder with a custom built motion system that is outfitted with an infrared imaging system (Hu et al., 2002). This process yields high precision metallic parts with consistent process quality. These four systems perform all deposition steps first, and then machine the part to the desired finish, consistent with conventional additive fabrication.

Two powder-based manufacturing processes that exhibit excellent material usage and in most cases produced components do not require finishing are Direct Metal Laser Sintering (DMLS) and Laser Consolidation (LC). Using layered manufacturing technology, a DMLS system such as the EOS EOSINT M270 Xtended system, can achieve an acceptable component finish using a fine 20 micron thick metal powder material evenly spread over the build area in 20 micron thick layers (3axis, 2010). Laser Consolidation developed by NRC Canada is a net-shape process that may not require tooling or secondary processing (except interfaces) (Xue, 2006, 2008). Parts produced using these processes exhibit net-shape dimensional accuracy and surface finish as well as excellent part strength and material properties.

Non-conventional additive processes demonstrate advanced features, alternate additive and subtractive steps, filling shell casts, etc. A hybrid RP process proposed by (Hur et al., 2002) combines a 6-axis machining center with any type of additive process that is machinable, a sheet reverse module, and an advanced process planning software package. What differentiates this process is how the software decomposes the CAD model into machining and deposition feature segments, which maximize the CNC milling machine advantages, and significantly reduces build time while increasing shape accuracy. Laser welding, another hybrid approach, involves a wire feeder, CO_2 laser, 5-axis milling center, and a custom PC-NC based control unit that has been used to produce molds for injection molding (Choi et al., 2001). Hybrid-Layered Manufacturing (HLM) as researched by (Akula & Karunakaran, 2006) integrates a TransPulse Synergic MIG/MAG welding process with a conventional milling machine to produce near-net shape tools and dies. This is direct rapid tooling. Welding and face milling operations are alternated to achieve desired layer height and to produce very accurate, dense metal parts. A comparable process was developed at Fraunhofer IPT named Controlled Metal Build-up (CMB), in which, after each deposited layer the surface is milled smooth (Kloche, 2002). However, CMB utilizes a laser integrated into a conventional milling machine.

Song and Park have developed a hybrid deposition process, named 3D welding and milling because a wire-based gas metal arc welding (GMAW) apparatus has been integrated with a CNC machining center (Song & Park, 2006). This process uses gas metal arc welding to deposit faster and more economically. Uniquely, 3D welding and milling can deposit two materials simultaneously with two welding guns or fill deposited shells quickly by pouring molten metal into them. The mold Shape Deposition Manufacturing (SDM) system at Stanford also uses multiple materials to deposit a finished part, however, for a different purpose (Cooper, 1999). A substrate is placed in the CNC mill and sturdy material such as UV-curable resin or wax is deposited to form the walls of a mold, which then is filled with an easily dissolvable material. The top of the mold is deposited over the dissolvable material to finish the mold; once the mold has cooled down the dissolvable material is removed, and replaced with the desired part material. Finally, the sturdy mold is removed to reveal the final part, which can be machined if necessary. Contrary to the typical design sequence (Jeng & Lin, 2001) constructed their own motion and control system for a Selective Laser Cladding (SLC) system and integrated the milling head, which evens out the deposition surface after every two layers. Clearly, each system has its advantages and contributes differently to the RM industry.

Although using a CNC milling machine for a motion system is the most common approach to constructing a hybrid system, a robot arm can easily be substituted. This is the case with SDM created at Stanford University (Fessler et al., 1999). The robot arm was fitted with an Nd:YAG laser cladding head which can be positioned accurately, allowing for selective depositing of the material and greatly reducing machining time. Integration of a handling robot can reduce positioning errors and time between operations if the additive and subtractive processes are not physically integrated.

Most of the aforementioned systems have been built with versatility in mind and could be set-up to utilize multiple materials or adapted to perform another operation. However, an innovative hybrid system that has very specific operations and capabilities is the variable lamination manufacturing (VLM-ST) and multi-functional hotwire cutting (MHC) system (Yang et al., 2005). The VLM-ST system specializes in large sized objects, up to 3 ft. x 5 ft., by converting polystyrene foam blocks into 3D objects utilizing the turntable of the 4-axis MHC system during cutting; if the object is bigger still, multiple pieces are cut and put together.

The design strategy behind several of the reviewed hybrid systems was not emphasized and documented. Thus, key pieces of information for the design and development of hybrid systems are missing which prevents researchers and designers from easily designing and constructing a hybrid system of their own. The information contained within this chapter aims to provide a comprehensive overview of the design, development, and integration of a hybrid manufacturing system such that others can use as a guideline for creating a hybrid system that meets their unique needs.

3. Research approach

As previously mentioned, the design strategies, system architecture, and knowledge required to construct a hybrid manufacturing system is vaguely described if mentioned at all in the literature. Consequently, our research approach is mainly empirical. Although our approach relies heavily on observation and experimental data, it has allowed us to identify opportunities for applying theory through modeling and simulation.

A major challenge to hybrid manufacturing system design is accurately controlling the physical dimension and material properties of the fabricated part. Therefore, understanding the interaction of all process parameters is key. Layout of the preliminary system architecture provides a basis for qualitative modeling. Independent and dependent process parameters are identified through qualitative modeling, which defines the parameters that require a quantitative understanding for accurate control of the process output. Qualitative models of the hybrid manufacturing process are developed and analyzed to understand both process and functional integration within the hybrid system. This allowed lost, competing or redundant system functionality to be identified and used to inform design decisions. Modeling how the material and information flows through the hybrid system facilitates the development of the automation, integration, and control strategies.

Quantitative modeling and simulation of our hybrid manufacturing system concentrates on process control and process planning. Process control modeling is used to predict the layer thickness via an empirical model based on the direct 3D layer deposition, the particle concentration of the powder flow, the nozzle geometry, the carrier gas settings, and the powder-laser interaction effects on the melt pool. Process planning models are used to automate part orientation, building direction, and the tool path. These models assist with resolving the challenges of the laser deposition process including building overhang structures, producing precision surfaces, and making parts with complex structures.

Revisiting the preliminary system architecture design with the knowledge gained from qualitative and quantitative modeling has resulted in a system architecture that enables accurate and efficient fabrication of 3D structures. Decomposition of the system architecture allows for direct mapping of customer needs and requirements to the overall system architecture.

4. Hybrid manufacturing system

The laser aided manufacturing process (LAMP) lab at Missouri University of Science and Technology (formerly University of Missouri-Rolla) houses a 5-axis hybrid manufacturing system, which was established by Dr. Liou and other faculty in the late 1990s. This system entails additive-subtractive integration, as shown in Fig. 1, to build a rapid prototyping/ manufacturing infrastructure for research and education at Missouri S&T. Integration of this kind was planned specifically to gain sturdy thin wall structures, good surface finish, and complex internal features, which are not possible by a LMD or machining system alone. Overall, the system design provides greater build capability, better accuracy, and better surface finish of structures with minimal post-processing while supporting automated control. Applications of the system include repairing damaged parts (Liou et al., 2007), creating functionally gradient materials, fabrication of overhang parts without support structures, and embedding sensors, and cooling channels into specialty parts.

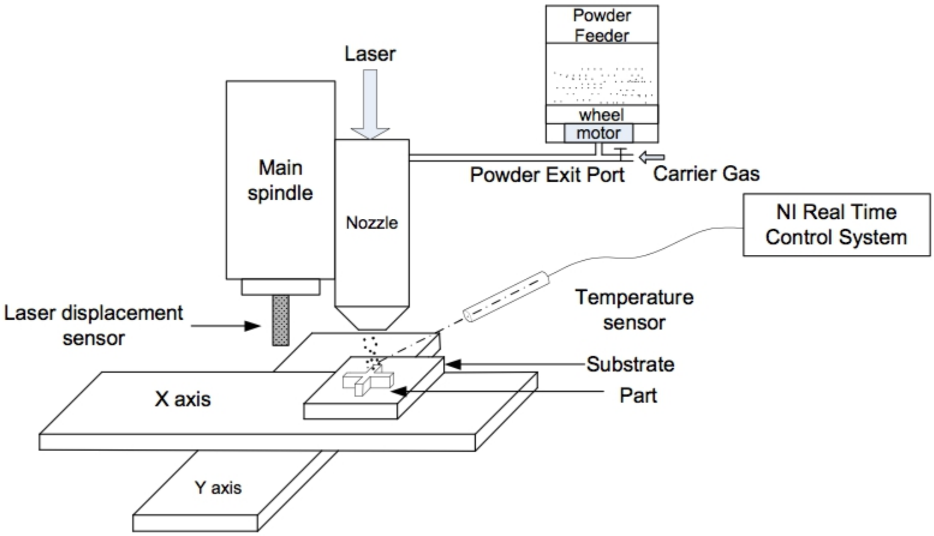


Fig. 1. Five-axis Hybrid Manufacturing Process (Adapted from Tang et al., 2007)

The LAMP hybrid system is comprised of five subsystems or integration elements: process planning, control system, motion system, manufacturing process, and a finishing system. Equipment associated with subsystem is described in the following paragraphs and summarized in Table 1.

The LAMP process planning system is a in-house layered manufacturing or slicing software that imports STL models from a commercial CAD package to generate a description that specifies melt pool length (mm), melt pool peak temperature, clad height (mm) and sequences of operations. The objective of the process planning software is to integrate the five-axis motion and deposition-machining hybrid processes. The results consist of the subpart information and the build/machining sequence (Ren et al., 2010; Ruan et al., 2005). To generate an accurate machine tool path a part skeleton, which calculates distance and offset edges or boundaries, is created of the CAD model. Distance, gradient, and tracing functions were modified to allow more complicated and unconnected known environments for successful implementation with the LAMP hybrid manufacturing system. Basic planning steps involve determining the base face, extracting the skeleton, decomposing a part into subparts, determining build sequence and direction for subparts, checking the feasibility of the build sequence and direction for the machining process, and optimization of the deposition and machining.

Hybrid Manufacturing Subsystems	LAMP Hybrid System Equipment
Process Planning	Commercial and in-house CAD software
Motion	Fadal 3016L 5-axis VMC
Manufacturing Process	Nuvonyx 1kW diode laser, Bay State Surface Technologies 1200 powder feeder
Control	NI RT PXI chassis & LabVIEW, Mikron temperature sensor, Omron laser displacement sensor, Fastcom machine vision system
Finishing	Fadal 3016L 5-axis VMC

Table 1. LAMP Hybrid System Equipment

True 3D additive manufacturing processes can be achieved with a 5-axis machining center without additional support structures (Ruan et al., 2005), as opposed to 2.5D that is afforded by a 3-axis machine. Therefore the motion subsystem for the LAMP hybrid manufacturing system is a 5-axis Fadal 3016L VMC, which also constitutes the finishing subsystem. Servo motors control the motion along the axes as compared with crank wheels and shafts in conventional machine tools. The Fadal VMC is controlled via G and M codes either entered at the control panel or remotely fed through an RS-232 connection.

The main manufacturing process of the hybrid system is laser metal deposition, the additive manufacturing process. Metal powder is melted using a 1kW diode laser while the motion system traverses in response to the tool path generated by the process planning software, thereby creating molten tracks in a layer-by-layer fashion on a metal substrate. Layers are deposited with a minimum thickness of 10μm. The melt pool temperature is between 1000°C and 1800°C, depending on the material (e.g. H13 tool steel, Titanium alloy), but is less than 2000°C. A commercial powder delivery system, designed for plasma-spraying processes carries the steel or titanium powder to the substrate via argon. The cladding head is mounted to the z-axis of the Fadal VMC to fully utilize the motion system and provide the opportunity to machine the fabricated part at any point in the deposition process by applying a translation algorithm. The beam focusing optics, beam splitter for out-coupling the process radiation from the laser beam path, water cooling connections, powder feeder connections, and various sensors (optional) are located within the cladding head. Built in to the cladding head are pathways for metal powder to travel through to the laser beam path in a concentric form, therefore, releasing metal powder in a uniform volume and rate. Quartz glass is used to focus the laser beam and water carried from the chiller to the cladding head by small plastic hoses reduces the wear on the focusing optics. Overall, the LMD subsystem includes equipment for lasing, cooling, and powder material delivery.

Control of the hybrid manufacturing subsystems require a versatile industrial controller and a range of sensors to acquire feedback. The National Instruments Real Time Control System (NI RT System) provides analog and digital I/O ports and channels, DAC, RS-232, and ADC for controlling all the subsystems of the hybrid system. The control system contains a PXI-

8170 Processor, 8211 Ethernet card, 8422 RS-232 card, 6527 Digital I/O card, 6711 Analog Output card, 6602 Timing I/O card, 6040E Multi-function card, and an SCXI Controller with 1304 card. PCI eXtensions for Instrumentation (PXI) is a PC-based platform for measurement and automation systems. PXI combines PCI electrical-bus features with the modular, Eurocard packaging of Compact PCI, and then adds specialized synchronization buses and key software features. Signal Conditioning Extension for Instrumentation (SCXI) is a front-end signal conditioning and switching system for various measurement devices, including plug-in data acquisition devices. Our control system offers modularity, expandability, and high bandwidth in a single, unified platform.

System feedback is acquired through temperature and laser displacement sensors. An Omron Z4M-W100 laser displacement sensor is used to digitally determine the cladding head height above the substrate. There are danger zones and safe zones that the nozzle can be with respect to the substrate. Output of the displacement sensor is -4 to +4 VDC which is converted into a minimum and maximum distance value, respectively. The temperature sensor is a Mikron MI-GA5-LO non- contact, fiber-optic, infrared temperature sensor. It was installed onto the Z-axis of the VMC with a custom, adjustable fixture. The set-up for data acquisition of the melt pool temperature, while deposition takes place is at an angle of 42°, 180 mm from the melt pool and sampling every 2 ms. There is also a machine vision system, a Fastcom iMVS-155 CMOS image sensor, to watch the melt pool in real-time. It has also been used to monitor melt pool geometry and assist with our empirical approach to fine tune process parameters.

5. Hybrid manufacturing system design and development

The critical success factors of an integrated system are quality, adaptability, productivity and flexibility (Garelle & Stark, 1988). Inclusion of additive fabrication technology in a traditional subtractive manufacturing system inherently addresses these four factors. Nevertheless, considering the four success factors during the initial design phase will ensure that the resultant manufacturing system will meet short and long term expectations, be reliable, and mitigate system obsolescence. In order for hybrid manufacturing systems to become a widespread option they must also be an economical solution. Dorf and Kusiakpoint out that the three flows within a manufacturing system i.e. material, information, and cost, which “should work effectively in close cooperation for efficient and economical manufacturing” (Dorf & Kusiak, 1994). This section reviews the qualitative and quantitative modeling efforts of material and information as well as the system architecture design that incorporates the knowledge gained through modeling. Cost modeling for the hybrid system has only been temporal, however, a cost benefit analysis as proposed in (Nagel & Liou, 2010) could be performed to quantify the savings.

5.1 System architecture

Initially, the LAMP system design was integrated only through the physical combination of the laser metal deposition process (additive manufacturing) and the machining center (subtractive manufacturing). Also, each subsystem housed a separate controller, including the LMD and VMC, which required manual control of the hybrid system. Reconfiguring the LAMP hybrid system to utilize a central control system, increased communication between the subsystems and eliminated the need for multiple people. Moreover, the process can be controlled and monitored from a remote location, increasing the safety of the manufacturing process. The hybrid manufacturing system architecture follows the modular, integration element structure as defined in (Nagel & Liou, 2010). Figure 2 shows the direct mapping of customer needs and requirements to the overall system architecture as well as the dependency relationships. Build geometry, surface finish, and material properties are the needs relating directly to the finished product. Efficient operation and flexibility are the system requirements to be competitive and relate directly to the system itself.

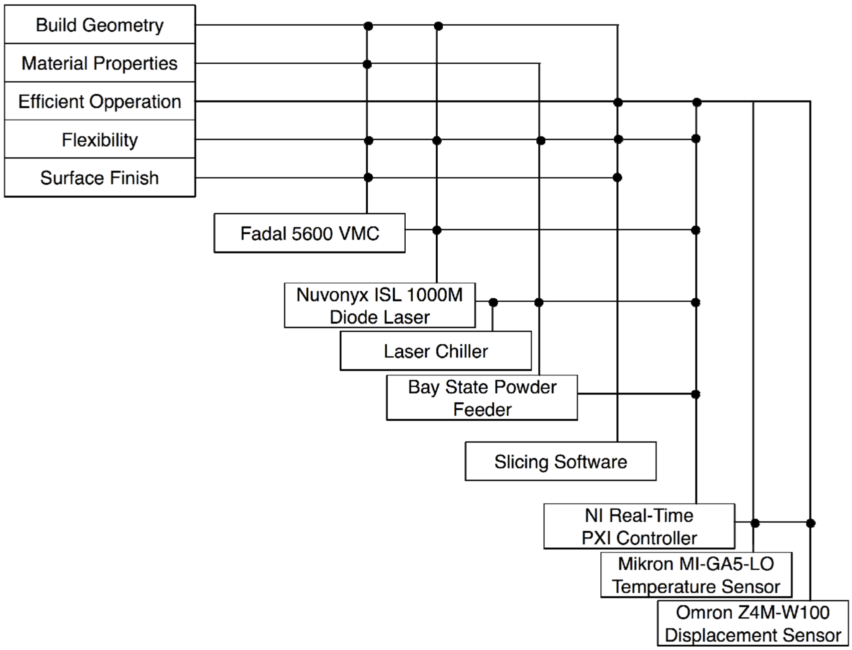


Fig. 2. LAMP Hybrid System Architecture

5.2 Qualitative modeling

Qualitative modeling efforts are focused on understanding process parameters and the flow of the process. Modeling the process parameter interactions uncovers the independent and dependent process parameters where as modeling the manufacturing process identifies opportunities for optimization. The following subsections summarize how qualitative modeling has been used to gain knowledge of the relationships among process parameters and resources utilized in each step of the hybrid manufacturing process.

5.2.1 Independent process parameters

The major independent process parameters for the hybrid manufacturing system include the following: laser beam power, process speed, powder feed rate, incident laser beam diameter, and laser beam path width (path overlap) as shown in Fig. 3 (Liou et al., 2001). Other parameters such as cladding head to surface distance (standoff distance), carrier gas flow rate, absorptivity, and depth of focus with respect to the substrate also play important roles.

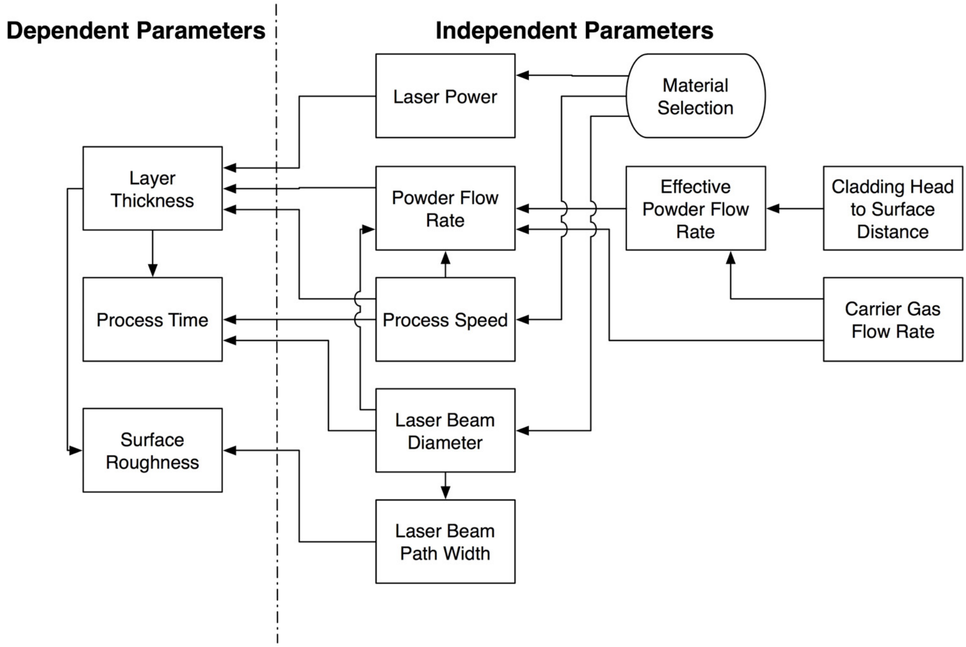


Fig. 3. LAMP Hybrid System Process Parameters (Adapted from Liou et al., 2001)

The layer thickness process parameter is directly related to the power density of the laser beam and is a function of incident beam power and beam diameter. Generally, for a constant beam diameter, the layer thickness increases with increasing beam power provided corresponding powder feed rate. It was also observed that the deposition rate increased with increasing laser power (Weerasinghe & Steen, 1983).

Powder mass flow rate is another important process parameter which directly affects layer thickness. However, effective powder flow rate, which includes powder efficiency during the LMD process, turned out to be a more important parameter (Lin & Steen, 1998; Mazumder et al., 1999). Also the factor that most significantly affected the percent powder utilization was laser power. The cladding head nozzle is set up to give a concentric supply of powder to the melt pool, and due to the nature of the set-up, the powder flow is hour glass-shaped. The powder flow initially is unfocused as it passes through the cladding head, but the nozzle guides the powder concentrically towards its center, and essentially "focuses" the beam of powder. The smallest diameter focus of the powder "beam" is dependent upon the design of the cladding head nozzle. Also, if the laser beam diameter becomes too small as compared to the powder beam diameter, e.g., 100μm, much of the supplied powder will not reach the melt pool. Thus, there will be unacceptably low powder utilization.

Process speed has a big impact on the process output. In general, decreasing process speed increases the layer thickness. There is a threshold to reduce process speed, however, as too much specific energy (as defined in Section 5.2.2) will cause tempering or secondary hardening of previous layers (Mazumder et al., 1997). Process speed should be well chosen since it has strong influence on microstructure.

The laser beam diameter parameter is one of the most important variables because it determines the power density. It can be difficult to accurately measure high power laser beams. This is partly due to the shape of the effective beam diameter (e.g., Gaussian, Top hat) and partly due to the definition of what is to be measured. Single isotherm contouring techniques such as charring paper and drilling acrylic or metal plates are well known but suffer from the fact that the particular isotherm they plot is both power and exposure time dependent. Multiple isotherm contouring techniques overcome these difficulties but are tedious to interpret.

Beam path width or beam width overlap has a strong influence on surface roughness. As the deposition pass overlap increases, the valley between passes is raised due to the overlap therefore reducing the surface roughness. Powder that has adhered to the surface, but has not melted will be processed in successive passes. In order to obtain the best surface quality, the percent pass overlap should be increased as much as possible. Conversely, to decrease the surface roughness, the deposition layers should be kept as thin as possible.

5.2.2 Dependent process parameters

The major dependent process parameters of the hybrid manufacturing system are: layer thickness, surface roughness, and process time (Fig. 3). Other dependent parameters such as hardness, microstructure, and mechanical properties should also be considered, but in this chapter we will focus only on the parameters related with physical dimension.

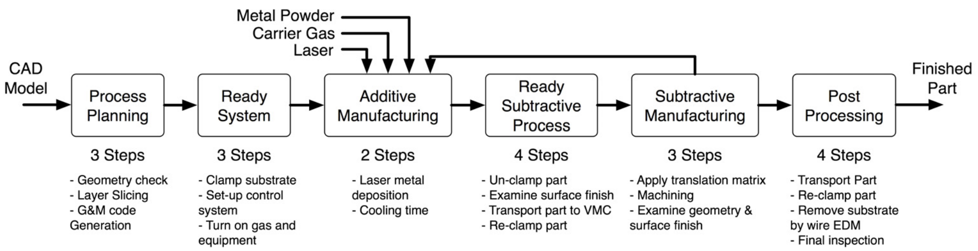
There is a large range of layer thicknesses as well as deposition rates that can be achieved using LMD. However, part quality consideration puts a limit on optimal deposition speeds. Both the layer thickness and the volume deposition rates are affected predominately by the specific energy and powder mass flow rate. Here, specific energy (SE) is defined as: $SE = p/(Dv)$, where p is the laser beam power, D is the laser beam diameter and v is the process speed. Also it has been well known that actual laser power absorbed in the melt pool is not the same as the nominal laser power measured from a laser power monitor due to reflectivity and other plasma related factors depending on the materials (Duley, 1983). The use of adjusted specific energy is thus preferable. Considering the factors, there is a positive linear relationship between the layer thickness and adjusted specific energy for a range of powder mass flow rates (Liou et al., 2001).

Surface roughness was found to be highly dependent on the direction of measurements with respect to the deposited metal (Liou et al., 2001; Mazumder et al., 1999). In checking the surface roughness, at least four directions should be tested from each sample; the length and width direction on the top surface, and the horizontal and vertical directions on thin walls. Since the largest roughness on each sample is of primary interest, measurements should be only taken perpendicular to the deposition direction on the top surface and in the vertical direction on the walls, based on our experiments.

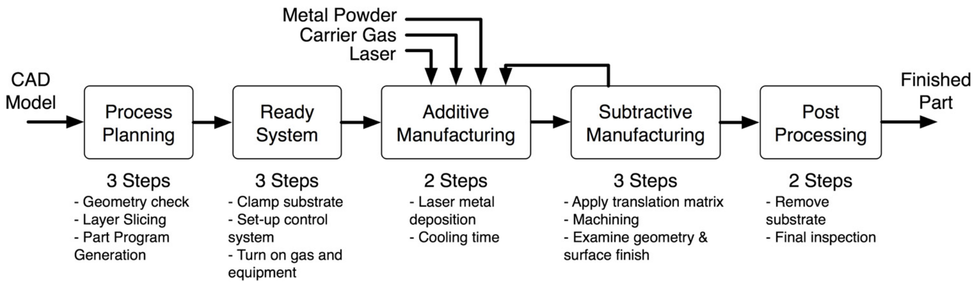
The overall deposition processing time is mainly dependent upon the layer thickness per slice, process speed, and laser beam diameter. The processing conditions need to be optimized prior to optimizing the processing time, since the processing time is directly influenced by the processing conditions. If the laser beam diameter is increased, the specific energy and power density will be decreased under the same process condition, that means, a lower deposition rate unless the laser power and powder mass flow rate are increased correspondingly. Similarly, when the process speed is increased the independent process parameters should be optimized accordingly.

5.2.3 Process modeling

Process modeling used to model the hybrid manufacturing system aims to optimize the sequence with which the material flows through the system (Shunk, 1992; Wang, 1997). Following the process modeling approach by (Nagel et al., 2009), process events and tasks within each event were identified. Part A of Fig. 4 shows the manually controlled hybrid manufacturing process. Decomposition of the system process aided with identification of integration points to reduce the number of steps and events within the process resulting in significant time savings. Part B of Fig. 4 shows the optimized process.



A)



B)

Fig. 4. LAMP Hybrid System Process Models, A) Before integration and optimization, B) After integration and optimization

Once the process is clearly laid out, the motion system and control system can be accurately defined. Reconfiguring the LAMP hybrid system elements to utilize a central control system increased communication between the subsystems and eliminated the need for multiple people. Moreover, the process can be controlled and monitored from a remote location, increasing the safety of the manufacturing process. Supplementary improvements were made to the process planning software, laser metal deposition subsystems, and the VMC. In efforts to eliminate the separate VMC computer, required only to upload machine code via direct numerical control, the RS-232 communication protocol utilized by Fadal was reverse engineered and implemented via LABVIEW. The laser, cooling, and powder material delivery subsystems of the laser metal deposition process are equipped with external control ports, but were not utilized in previous system configurations. Subsequently, all subsystems and modules were directly connected to the control system hardware so external control could be utilized. Initializing communications among the LAMP subsystems became the foundation for the control system software. Off-line, the in-house layered manufacturing software only converted CAD models into layer-by-layer slices of machine code to create the tool path. With the central control system now in place, the in-house layered manufacturing software was changed to generate machine code, laser power, and powder flow commands, which together comprise a part program and are distributed via the control system software. Overall, manufacturing process integration has resulted in modularity, easy maintenance, and process improvement. Thus, increasing system productivity and capability.

5.3 Quantitative modeling and simulation

Quantitative modeling and simulation provides a theoretical foundation for explaining the phenomena observed through empirical research. Additionally, detailed modeling assists with developing a quantitative understanding of the relationship between independent process parameters and dependent process parameters. Understanding the relationships among parameters affords accurate control of physical dimension and material properties of the part. While separate modeling efforts were undertaken, outputs of one model feed into another. The following subsections summarize how quantitative modeling has been used to develop a theoretical understanding of the LAMP hybrid manufacturing process.

5.3.1 Melt pool modeling and simulation

Melt pool geometry and thermal behavior control are essential in obtaining consistent building performances, such as geometrical accuracy, microstructure, and residual stress. A 3D model was developed to predict the thermal behavior and geometry of the melt pool in the laser material interaction process (Han et al., 2005). The evolution of the melt pool and effects of the process parameters were investigated through modeling and simulations with stationary and moving laser beam cases.

When the intense laser beam irradiates on the substrate surface, the melt pool will appear beneath the laser beam and it moves along with the motion of the laser beam. In order to interpret the interaction mechanisms between laser beam and substrate the model considers the melt pool and adjacent region. The governing equations for the conservation of mass, momentum and energy can be expressed in following form:

$$\frac{\partial}{\partial t}(\rho) + \nabla \cdot (\rho \mathbf{V}) = 0 \tag{1}$$

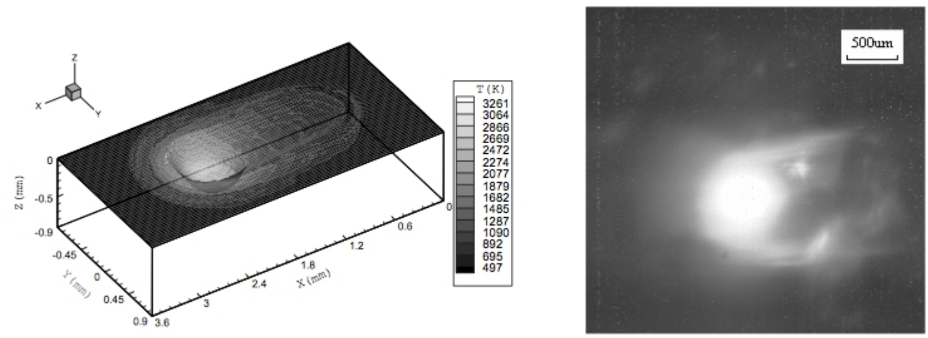
$$\frac{\partial}{\partial t}(\rho \mathbf{V}) + \nabla \cdot (\rho \mathbf{V}\mathbf{V}) = \nabla \cdot (\mu_l \frac{\rho}{\rho_l} \nabla \mathbf{V}) - \nabla p - \frac{\mu_l}{K}\frac{\rho}{\rho_l}(\mathbf{V} - \mathbf{V}_s) + \rho g \tag{2}$$

$$\frac{\partial}{\partial t}(\rho h) + \nabla \cdot (\rho \mathbf{V} h) = \nabla \cdot (k \nabla T) - \nabla \cdot (\rho (h_l - h)(\mathbf{V} - \mathbf{V}_s)) \tag{3}$$

where ρ, **V**, *p*, μ, *T*, *k*, and *h* are density, velocity vector, pressure, molten fluid dynamic viscosity, temperature, conductivity, and enthalpy, respectively. *K* is the permeability of mushy zone, **V***s* is moving velocity of substrate with respect to laser beam and subscripts of *s* and *l* represent solid and liquid phases. Since the solid and liquid phases may coexist in the same calculation cell at the mushy zone, mixed types of thermal physical properties are applied in the numerical implementation. The liquid/vapor interface is the most difficult boundary for numerical implementation in this model since many physical phenomena and interfacial forces are involved there. To solve those interfacial forces the level set method is employed to acquire the solution of the melt pool free surface (Han et al., 2005). To avoid numerical instability arising from the physical property jump at the liquid/vapor interface, the Heaviside function $H(\varphi)$ is introduced to define a transition region where the physical properties are mollified.

The energy balance between the input laser energy and heat loss induced by evaporation, convection and radiation determines surface temperature. Laser power, beam spot radius, distance from calculation cell to the beam center, and the absorptivity coefficient are used to calculate the laser heat influx. Heat loss at the liquid/vapor interface is computed in terms of convective heat loss, radiation heat loss and evaporation heat loss. The roles of the convection and surface deformation on the heat dissipation and melt pool geometry are revealed by dimensionless analysis. It was found that interfacial forces including thermo-capillary force, surface tension and recoil vapor pressure considerably affect the melt pool shape and fluid flow. Quantitative comparison of interfacial forces indicates that recoil vapor pressure is dominant under the melt pool center while thermo-capillary force and surface tension are more important at the periphery of the melt pool.

For verification, the intelligent vision system was utilized to acquire melt pool images in real time at different laser power levels and process speeds, and the melt pool geometries were measured by cross-sectioning the samples obtained at various process conditions (Han et al., 2005). Simulation predictions were compared to experimental results for both the stationary laser case and moving laser case at various process conditions. Model prediction results strongly correlate to experimental data. An example of melt pool shape comparison between simulation and experiment for the moving laser beam case is shown in Fig. 5.



A) B)

Fig. 5. Melt pool shape comparison, A) Simulation result of melt pool shape and surface temperature, B) Experimental result of melt pool shape (Adapted from Han et al., 2005)

5.3.2 Powder flow dynamics modeling and simulation

Analysis of metallic powder flow in the feeding system is of particular significance to researchers in order to optimize the LMD fabrication technique. Powder flow simulation holds a critical role in understanding flow phenomena. A stochastic Lagrangian model for simulating the dispersion behavior of metallic powder, or powder flow induced by non-spherical particle-wall interactions, is described (Pan & Liou, 2005). The numerical model also takes into consideration particle shape effects. In wall-bounded, gas-solid flows, the wall collision process plays an important role and is strongly affected by particle shape. Non-spherical effects are considered as the deviation from pure spheres shows induced particle dispersion, which has a great impact on the focusibility of the powder stream at the laser cladding head nozzle exit. The parameters involved in non-spherical collision are analyzed for their influencing factors as well as their interrelations.

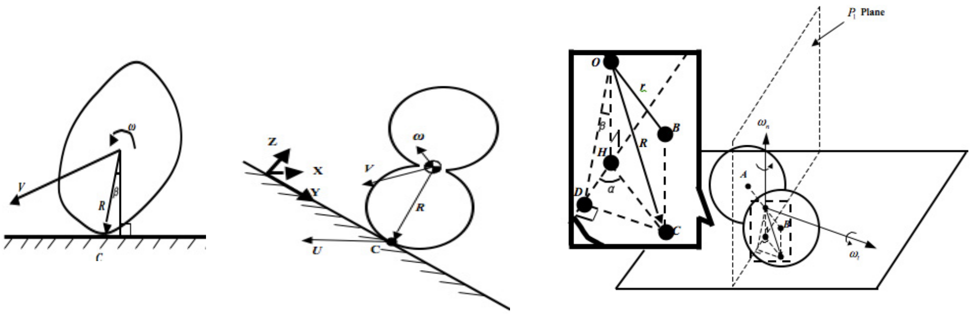


Fig. 6. Particle Collision Diagrams, A) 2D non-spherical particle-wall collision model, B) Local coordinate for collision model, C) of 3-D non-spherical particle-wall collision model (Adapted from Pan & Liou, 2005)

The parameters involved in the 2-D non-spherical model include β and R, as shown in Fig. 6, Part A, where β indicates how much the contact point C deviates from the foot of a vertical from the gravity center of the particle and R shows the actual distance between the contact point and the gravity center. The collision coordinate system used to describe the 3D collision dynamics is defined in Fig. 6, Part B. The contact velocity is computed from:

$$U = V + \omega \times R \tag{4}$$

where V is the particle translational velocity vector, ω is the angular velocity vector, and R is the vector connecting particle mass center to contact point C. The change in the contact point velocity can be obtained by the following equation:

$$\Delta U = [\frac{1}{m} I - R^{\times} J^{-1} R^{\times}]\Delta P \tag{5}$$

where m is particle mass, I is the 3x3 identity matrix, and R^x is the canonical 3x3 skew-symmetric matrix corresponding to R, ΔP denotes the impulse delivered to the particle in the collision, and J^{-1} is the inverted inertia tensor in the local coordinate. As shown in Fig. 6, Part C, a cluster that consists of two identical spheres with equal radius *r* represents the non-spherical powder particle of the 3D model. This representation leads to generalized modeling of satelliting metallic powder particles.

Wall roughness also effects powder dispersion behavior, therefore in this model the roughness effect was included by using the model and parameters proposed by (Sommerfeld & Huber, 1999). The instantaneous impact angle is assumed to be composed of the particle trajectory angle with respect to the plane wall and a random component sampled from a Gaussian distribution function. It was also assumed that each collision has 30% possibility to be non-spherical, which implies the stochastic model was applied in 30% of the total collisions during the feeding process simulation. Simulations using the spherical model (0% non-sphericity) were also conducted.

The non-spherical model successfully predicts the actual powder concentration profile along the radial and axial directions, whereas the spherical particle model underestimates the dispersion and results in a narrow spread of the stream along the radial direction. When compared to the experimental results, the 3D simulated powder stream is in strong agreement, which demonstrates validation of the model. The model also predicts the peak powder concentration or focal point of the power stream for specific cladding head nozzle geometry. It is essential to establish a well-focused powder stream at the exit of the nozzle and to know the ideal stand-off distance in order to increase powder catchment in the melt pool, achieve high material integrity, and reduce material waste.

5.3.3 Tool path modeling and simulation

Process planning, simulation, and tool path generation allows the designer to visualize and simulate part fabrication prior to manufacturing to ensure a successful process. Adaptive multi-axis slicing, collision detection, and adaptive tool path pattern generation for LMD as well as tool path generation for surface machining are the key advantages to the integrated process planning software developed for the LAMP hybrid system (Ren et al., 2010).

Basic planning steps involve determining the base face and extracting the skeleton of an input CAD model (Fig. 7, top left). The skeleton is found using the centroidal axis extraction algorithm (Fig. 7, top right). Based on the centroidal axis, the part is decomposed into sub-components and for each sub-component a different slicing direction is defined according to build direction. In order to build some of the components, not only translation but also rotation will be needed to finish building the whole part because different sub-components have different building directions (Fig. 7, bottom middle), and the laser nozzle direction is always along the z-axis. After the decomposition (Fig. 7, bottom left) results are obtained, the relationship among all the components is determined, and a building relationship graph is created.

From the slicing results and build directions, collision detection is determined. Collision detection is implemented by Boolean operation, which is an intersection operation, on a simulation (Ren et al., 2010). If the intersection result of the updated CAD model and the cladding head nozzle is not empty, then collision will happen in the real deposition process. The deformation of the CAD model following the building relationship graph includes two categories: positional deformation and dimensional deformation. Positional change means translation or rotation of the CAD model. The dimensions will change after every slicing layer is finished. For every updated model, collision needs to be checked before the next slicing layer is added. Following the collision detection algorithm, if a potential collision is detected the sequence of the slicing layers is reorganized (Ren et al., 2010). The output of the collision detection algorithm will be the final list of slicing layers, which comprise the actual building sequence when manufacturing the part.

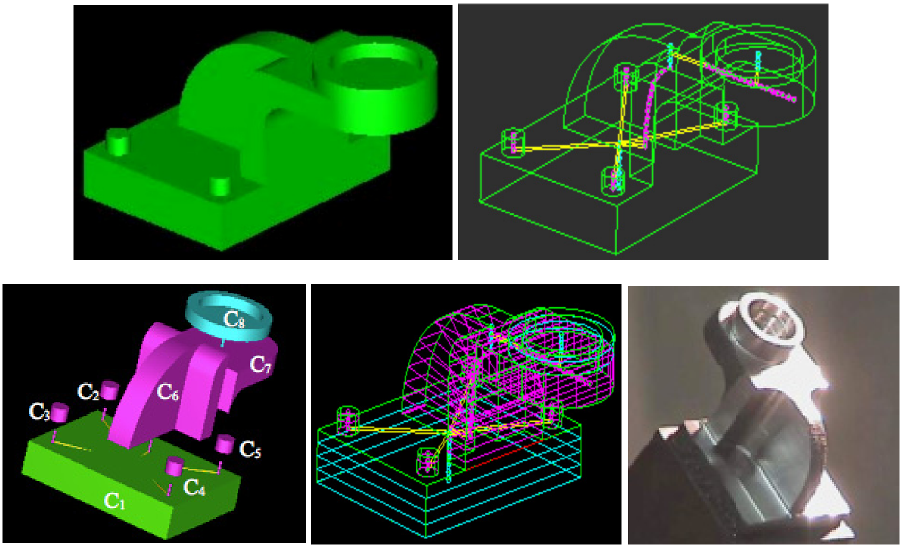


Fig. 7. Process Planning and Fabrication of 3D Part (Adapted from Liou et al., 2007; Ren et al. 2010)

The final piece of process planning is tool path generation. Common tool path patterns are the raster, contour-parallel offsetting, zig-zag, and interlaced. Each pattern has advantages and disadvantages. The adaptive deposition tool path algorithm considers each pattern when predicting the possibility of deposition voids. The goals of the algorithm are to adjust the tool path to remove deposition voids and increase time efficiency. Multiple tool path patterns may be used during fabrication and the algorithm may also prescribe alternating the appropriate tool path pattern when necessary.

Surface finish machining is a sequential step used after deposition to improve manufacturing quality after deposition is finished. The process planning software allows the designer to specify the machining parameters including the feed rate, spindle speed, and depth of cut before determining the number of machining cycles necessary. As with LMD, alignment will be also integrated for 3D geometries to achieve the accuracy without reloading the deposited part to be machined. Again, the tool path will be generated such that a collision-free machining tool path will be generated for the deposited part. A visibility map algorithm (Ruan & Liou, 2003) is applied to detect the collision between the tool and the deposited part.

The final process planning step is to generate the part program. This step is the bridge between the algorithmic results of process planning, quantitative modeling of process parameters, and the realistic operational procedures as well as parameters of the 5-axis manufacturing environment. It will build the map of the process planning results and the real operational parameters and then interpret the final planning tool path as the corresponding movements of the hybrid manufacturing system. The software will combine and refine those movements and translate them into machine executable code. Resulting in a text file composed of three columns of data to needed for the control system to command the laser, powder feeder, and motion system (Ren et al., 2010). The final set of operations is based on the building relationship graph, build directions that avoid collisions, tool path, and time required.

6. Hybrid manufacturing system integration

During the course of this research several integrated manufacturing system designs were analyzed to identify what characteristics comprise a successful hybrid system. Based on this background research, and the experiences of working with and refining the LAMP system, the key elements of a hybrid manufacturing system were identified. The five key elements represent an effective way to design a hybrid manufacturing system, as compared to a reconfigurable or mechatronic design, because the identified elements contain necessary subsystems, are easily modularized, and advocate the use of off-the-shelf hardware and software. Within an integrated system, each element acts as a separate subsystem affording a stable modular design (Gerelle & Stark, 1988).

A strategy for controlling the integrated LMD and machining processes, the 5-axis motion system, and the data corresponding streams provides the basis for fully automating the system. Considering scalability, our integration strategy emphasizes modularity of the integrated components but also modularity of the controlling software. Our control strategy allows data streams to be easily added or removed. Furthermore, our design allows an operator to optimize the control strategy for a particular geometry.

6.1 Physical Integration

Obstacles arise during the development of any manufacturing system; however, by identifying obstacles and solutions the industry as a whole can benefit. Outside of cost and yield, the obstacles of developing a hybrid manufacturing system discussed here cover a range of topics. Table 2 summarizes the obstacles associated with the physical integration of the LAMP hybrid system and provides documented solutions. The documented information in Table 2 does not address every possible integration obstacle, but is meant to be comprehensive from what is found in the literature and personal experience. Issues outside of integration, such as material properties can be found in (Nagel & Liou, 2010).

After central control, integration, and modularity were enforced in the LAMP hybrid system, manufacturing defects and time were significantly reduced, and safety was significantly increased. Material integrity was improved as the laser could be precisely commanded on/off or pulsed as needed during deposition. Furthermore, by integrating the laser power and powder flow commands into the process planning software and automating the distribution of commands, functionally graded parts were manufactured effortlessly.

Issue	Solution	Result	Reference
Adding the laser cladding head to a VMC	A platen with precisely tapped holes for the cladding head mounted to the Z-axis of the VMC	Laser cladding head is securely mounted and future equipment or fixtures can be added	
Protection of Equipment	Retract laser head or position it far enough away from the machining head	Protect laser nozzle	Kerschbaumer & Ernst, 2004
	Mount a displacement sensor on the Z-axis	When cladding head gets too close to X-Y axes the process halts	
Unknown communication protocol	Use reverse engineering to figure out communication protocol	Subsystems can be controlled from a central control system	Stroble et al., 2006
Quality control	Implement control charts, pareto charts, etc.	Manual quality control	Starr, 2004
	Sensor feedback utilized by closed-loop controllers	Automated quality control	Boddu et al., 2003; Doumanidis & Kwak, 2001; Hu & Kovacevic, 2003; Tang, 2007
Transition between additive and subtractive processes	Apply a translation matrix that repositions the X-Y axis for the desired process	Accurate positioning for machining or LMD	
Placement of sensors to monitor melt pool due to high heat of the LMD process	Mount the sensitive vision system in-line with the laser using a dichromatic mirror attachment for the cladding head, and custom hardware mounted to the platen holds the temperature probe at an acceptable viewing angle	Sensors are safe, and the LMD process is accessible	Boddu et al., 2003; Tang & Landers, 2010

Table 2. Physical Integration Issues and Solutions

6.2 Software Integration

Utilization of a central control system directly resulted in automation of the LAMP hybrid system and allowed unconventional possibilities to be explored. To achieve the central controller, a framework consisting of a multi-phase plan and implementation methodology was developed. The automation framework involves controlling the laser, powder feeder, and motion system, and utilizing sensor feedback, all through the NI PXI control subsystem. Open and closed-loop controllers were designed, along with compatibility and proper module communication checking. Moreover, compensation for undesired system dynamics, delays and noise were considered to ensure a reliable and accurate automated manufacturing process. The result of the automation framework is an automated deposition program (developed in LabVIEW) with a customized graphical user interface and data recording capabilities.

Figure 8 is a visual description of the LAMP hybrid system communications layout, including process planning that occurs outside the control system. Once process planning completes the part program, with laser power and powder mass flow rate commands in the form of voltages, the control system parses through the information to automatically fabricate the desired part. While commands are being sent to the physical devices, sensors are monitoring the process and sending feedback to the control system simultaneously, allowing parameters to change in real-time.

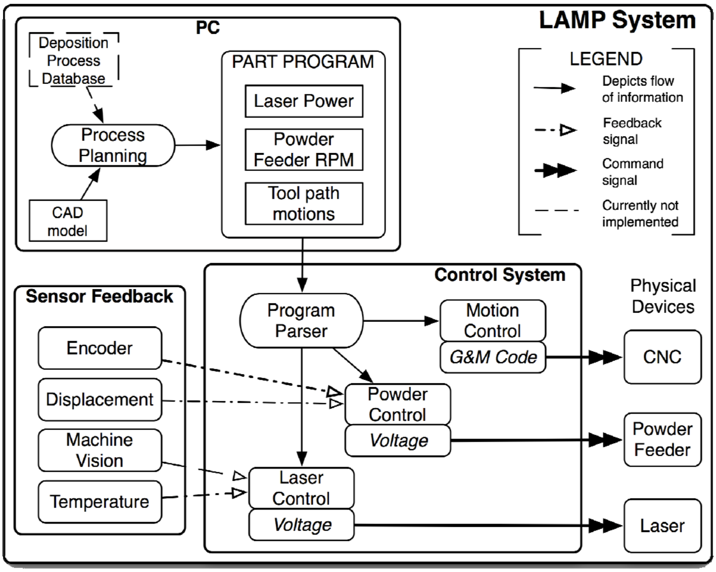


Fig. 8. LAMP Hybrid System Communication Schematic

Unique to the LAMP hybrid system is that the hardware and software are both modular. The automated deposition program that is executed by the control system has three different modes: dry-run, open-loop control, and closed-loop control. Fundamental code within the automated deposition program is shared amongst each of the modes, much like the control system is central to the LAMP hybrid system. Additional portions of code that control the laser, control the powder feeder, utilize feedback, or simply read in, display and record data from sensors are turned on or off by each mode. Code modularity prevents large amounts of the control system software from being rewritten when equipment is upgraded or subsystems are replaced.

During dry-run mode only machine code is distributed by the control system, allowing the user to monitor the VMC motions without wasting materials and energy. This mode is primarily utilized to check uncertain tool paths for instances when the laser should be shut off or when a tool path transition seems too risky. For instance, transitions from one geometry to another may rotate longer than desired at one point causing a mound to form and solidify, which destroys the overall part geometry and could collide with the laser cladding nozzle. Open-loop and closed-loop control modes are provided for fabricating parts and include system monitoring and data acquisition features. The modular software allows multiple closed-loop controllers optimized for a particular geometry to be added as research is completed, such as a feed forward controller (Tang et al., 2007) that regulates powder flow to the melt pool for circular, thin walled structures or thin walled structures with many arcs.

7. Conclusion

In an effort to shorten the time-to-market, decrease the manufacturing process chain and cut production costs, research has focused on the integration of multiple manufacturing processes into one machine; meaning less production space, time, and manpower needed. An integrated or hybrid system has all the same features and advantages of rapid prototyping systems, plus provides a new set of features and benefits. Moreover, hybrid manufacturing systems are increasingly being recognized as a means to produce parts in material combinations not otherwise possible and have the ability to fabricate complex internal geometries, which is beyond anything that can be accomplished with subtractive technologies alone. Internal geometries such as complex conformal cooling channels provide better product thermal performance, which additive fabrication processes create them with ease, giving the manufacturer a better product with little extra cost. As manufacturers and customers dream up more complex products, requiring more advanced equipment and software, hybrid systems will emerge. In short, integrating additive and subtractive technologies to create new manufacturing systems and processes is going to advance the manufacturing industry in today’s competitive market.

Modeling and simulation, both qualitative and quantitative, were shown to be an integral part of hybrid system design and development as well as motivate areas of research that a pure empirical approach does not reveal. Although this research is focused on integrating additive and subtractive processes, the general principles can also be applicable to integrating other unit manufacturing processes (NRC, 1995). Integrated processes can combine multiple processes that fall within the same family, such as different material removal processes, or they can combine processes that are in different unit process families, such as a mass-change process and a microstructure-change process. The results can lead to significant processing breakthroughs for low-cost, high-quality production.

Future work includes applying integrated process and product analysis to various hybrid processes that integrate different manufacturing processes and applying the hybrid system concept to other types of configurations, such as those that include robots. Model-based simulation reveals various new opportunities for simultaneous improvement of part quality, energy and material efficiencies, and environmental cleanness. Thereby, accelerating the hybrid integration process. Other work includes applying an open architecture for the hybrid controller, as such an architecture avoids the difficulties of using proprietary technology and offers an efficient environment for operation and programming, ease of integrating various system configurations, and provides the ability to communicate more effectively with CAD/CAM systems and factory-wide information management systems.

8. Acknowledgment

This research was supported by the National Science Foundation Grant # IIP-0822739, the U.S. Air Force Research Laboratory, and the Missouri S&T Intelligent Systems Center. We would also like to thank all the Missouri S&T LAMP lab researchers that have contributed over the years to make this body of work possible, especially Zhiqiang Fan, Lijun Han, Heng Pan, Lan Ren, Jianzhong Ruan, and Todd Sparks. Their support is greatly appreciated.

9. References

Akula, S. & Karunakaran, K.P. (2006). Hybrid Adaptive Layer Manufacturing: An Intelligent Art of Direct Metal Rapid Tooling Process. *Robotics and Computer-Integrated Manufacturing*, Vol. 22, No. 2, pp. 113-123, 0736-5845

Boddu, M.R., Landers, R.G., Musti, S., Agarwal, S., Ruan, J. & Liou, F. W. (2003) System Integration and Real-Time Control Architecture of a Laser Aided Manufacturing Process. *Proceedings of SFF Symposium*, 1053-2153, Austin, TX, August, 2003

Choi, D.-S., Lee, S.H., Shin, B.S., Whang, K.H., Song, Y.A., Park, S.H. & Lee, H.S. (2001) Development of a Direct Metal Freeform Fabrication Technique Using Co2 Laser Welding and Milling Technology. *Journal of Materials Processing Technology*, Vol. 113, No. 1-3, (June 2001), pp. (273-279), 0924-0136.

Cooper, A.G. (1999). Fabrication of Ceramic Components Using Mold Shape Deposition Manufacturing. Doctor of Philosophy Thesis, Stanford, USA.

Dorf, R.C. & Kusiak, A. (1994). *Handbook of Design, Manufacturing and Automation*, J. Wiley and Sons, 0471552186, New York, N.Y.

Doumanidis, C. & Kwak, Y.-M. (2001) Geometry Modeling and Control by Infrared and Laser Sensing in Thermal Manufacturing with Material Deposition. *Journal of Manufacturing Science and Engineering*, Vol. 123, No. 1, pp. 45-52, 1087-1357

Duley, W.W. (2003). *Laser Processing and Analysis of Materials*, Plenum Press, 0306410672, New York

Fessler, J.R., Merz, R., Nickel, A.H. & Prinz, F.B. (1999) Laser Deposition of Metals for Shape Deposition Manufacturing. *Proceedings of SFF Symposium*, 1053-2153, Austin, TX, August, 1999

Gebhardt, A. (2003). *Rapid Prototyping*, Hanser Publications, 9781569902813, Munich

Gerelle, E.G.R. & Stark, J. (1988). *Integrated Manufacturing: Strategy, Planning, and Implementation*, McGraw-Hill, 0070232350, New York

Han, L., Liou, F. & Musti, S. (2005). Thermal Behavior and Geometry Model of Melt Pool in Laser Material Process. *Journal of Heat Transfer*, Vol. 127, No. 9, pp. 1005, 0022-1481

Hopkinson, N., Hague, R.J.M. & Dickens, P.M. (2006) *Rapid Manufacturing: An Industrial Revolution for the Digital Age*, John Wiley, 0470016132, Chichester, England

Hu, D. & Kovacevic, R. (2003). Sensing, Modeling and Control for Laser-Based Additive Manufacturing. *International Journal of Machine Tools & Manufacture*, Vol. 43, No. 1, (January 2003), pp. 51-60, 0890-6955

Hu, D., Mei, H. & Kovacevic, R. (2002). Improving Solid Freeform Fabrication by Laser-Based Additive Manufacturing. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, Vol. 216, No. 9, pp. 1253-1264, 1253-1264

Hur, J., Lee, K., Hu, Z. & Kim, J. (2002). Hybrid Rapid Prototyping System Using Machining and Deposition. *Computer-Aided Design*, Vol. 34, No.10, (September 2002), pp. 741-754, 0010-4485

Jeng, J.-Y. & Lin, M.-C. (2001). Mold fabrication and modification using hybrid processes of selective laser cladding and milling. *Journal of Materials Processing Technology,* Vol. 110, No. 1, (March 2001), pp 98-103, 0924-0136

Kai, C.C., & Fai, L.K. (1997). *Rapid Prototyping: Principles & Applications in Manufacturing*, John Wiley, 9810245165, New York.

Kerschbaumer, M., & Ernst, G. (2004). Hybrid Manufacturing Process for Rapid High Performance Tooling Combining High Speed Milling and Laser Cladding. *Proc. 23rd International Congress on Applications of Lasers & Electro-Optics (ICALEO)*, 0912035773 , San Francisco, CA, October 2004

Klocke, F. (2002). *Rapid Manufacture of Metal Components*. Fraunhofer Institute for Production Technology, IPT

Kalpakjian, S., & Schmid, S.R. (2003). *Manufacturing Processes for Engineering Materials*. Pearson Education, Inc., 9780130453730, Upper Saddle River, N.J

Lin, J. & Steen, W.M. (1998). Design characteristics and development of a nozzle for coaxial laser cladding. *Journal of Laser Applications*, Vol. 10, No. 2, pp. 55-63, 1042-346X

Liou, F.W., & Kinsella, M. (2009). A Rapid Manufacturing Process for High Performance Precision Metal Parts. *Proceedings of SME Rapid 2009 Conference and Exhibition*, Paper No. TP09PUB18, Schaumburg, IL, May, 2009

Liou, F., Slattery, K., Kinsella, M., Newkirk, J., Chou, H.-N., Landers, R. (2007). Applications of a Hybrid Manufacturing Process for Fabrication and Repair of Metallic Structures, *Rapid Prototyping Journal*, Vol. 13, No. 4, pp. 236–244, 1355-2546

Liou, F.W., Choi, J., Landers, R.G., Janardhan, V., Balakrishnan, S.N., & Agarwal, S. (2001). Research and Development of a Hybrid Rapid Manufacturing process. *Proceedings of SFF Symposium*, 1053-2153, Austin, TX, August, 2001

Mazumder, J., Choi, J., Nagarathnam, K., Koch, J. & Hetzner, D. (1997). The Direct Metal Deposition of H13 Tool Steel for 3-D Components. *JOM*, Vol. 49, No. 5, pp.55-60 1047-4838

Mazumder, J., Schifferer, A., & Choi, J. (1999). Direct Materials Deposition: Designed Macro and Microstructure", Materials Research Innovations, Vol. 3, No. 3, (October 1999), pp.118-131, 1432-8917

NRC (National Research Council) (1995). *Unit Manufacturing Processes: Issues and Opportunities in Research*, The National Academies Press

Nagel, J.K.S. & Liou, F. (2010). Designing a Modular Rapid Manufacturing Process. *Journal of Manufacturing Science and Engineering*, Vol. 132, No. 6, (December 2010), pp. 061006, 1087-1357

Nagel, R., Hutcheson, R., Stone, R., & Mcadams, D., (2009). Process and Event Modeling for Conceptual Design. *Journal of Engineering Design*, Vol. 22, No. 3, (March 2011), pp.145-164, 0954-4828

Nowotny, S., Scharek, S., & Naumann, T. (2003). Integrated Machine Tool for Laser Beam Cladding and Freeforming. *Proc. Thermal Spray 2003: Advancing the Science & Applying the Technology*, 9780871707857, Orlando, FL, May, 2003

Pan, H. & Liou, F. (2005). Numerical simulation of metallic powder flow in a coaxial nozzle for the laser aided deposition process. *Journal of Materials Processing Technology*, Vol. 168, No. 2, pp. 230-244, 0924-0136

Ren, L., Sparks, T., Ruan, J. & Liou, F. (2010). Integrated Process Planning for a Multiaxis Hybrid Manufacturing System. *Journal of Manufacturing Science and Engineering*, Vol. 132, No. 2, pp. 021006, 1087-1357

Ren, L., Padathu, A.P., Ruan, J., Sparks, T., & Liou, F.W. (2008). Three Dimensional Die Repair Using a Hybrid Manufacturing System. *Proceedings of SFF Symposium*, 1053-2153, Austin, TX, August, 2008

Ruan, J., Eismas-Ard, K., & Liou, F. W., 2005, "Automatic Process Planning and Toolpath Generation of a Multiaxis Hybrid Manufacturing System," Journal of Manufacturing Processes, 7(1), pp. 57-68.

Ruan, J., & Liou, F.W. (2003) Automatic Toolpath Generation for Multi-axis Surface Machining in a Hybrid Manufacturing System. *Proc. of ASME IDETC/CIE*, 0791837009, Chicago, IL., September, 2003

Shunk, D.L. (1992). *Integrated Process Design and Development*, Business One Irwin, 9781556235566, Homewood, IL.

Sommerfeld, M., & Huber, N. (1999). Experimental analysis and modeling of particle-wall collisions. *International of multiphase flow*, Vol. 25, pp.1457-1489, 0301-9322

Song, Y.-A., & Park, S. (2006). Experimental Investigations into Rapid Prototyping of Composites by Novel Hybrid Deposition Process. *Journal of Materials Processing Technology*, Vol. 171, No. 1, (January 2006), pp. 35-40, 0924-0136

Starr, M. K., 2004, *Production and Operations Management*, Atomic Dog, 1592600921, Cincinnati, OH.

Stroble, J.K., Landers, R.G., & Liou, F.W. (2006). Automation of a Hybrid Manufacturing System through Tight Integration of Software and Sensor Feedback. *Proceedings of SFF Symposium*, 1053-2153, Austin, TX, August, 2006

Tang, L. & Landers, R.G. (2010). Melt Pool Temperature Control for Laser Metal Deposition Processes, Part I: Online Temperature Control," *Journal of Manufacturing Science and Engineering*, Vol. 132, No. 1, (February 2010), pp. 011010, 1087-1357

Tang, L., Ruan, J., Landers, R., & Liou, F. (2007). Variable Powder Flow Rate Control in Laser Metal Deposition Processes, *Proceedings of SFF Symposium*, 1053-2153, Austin, TX, August, 2007

Venuvinod, P.K., & Ma, W. (2004). *Rapid Prototyping Laser-Based and Other Technologies*, Kluwer Academic Publishers, ISBN: 978-1-402-07577-3, Boston

Wang, B. (1997). *Integrated Product, Process and Enterprise Design*, Chapman & Hall, 0412620200, New York.

Weerasinghe V.W. & Steen, W.M. (1983). Laser Cladding with Pneumatic Powder Delivery. *Proceedings of Laser Materials Processing*, Los Angeles, CA, January 1983

Westkämper, E. (2007). Digital Manufacturing In The Global Era, In: *Digital Enterprise Technology,* Pedro Cunha and Paul Maropoulos, pp. 3-14, Springer, 978-0-387-49864-5, New York

Xue, L. (2006). Laser Consolidation – a One-Step Manufacturing Process for Making Net-Shaped Functional Aerospace Components. *Proc. SAE International Aerospace Manufacturing and Automated Fastening Conference & Exhibition*, Toulouse, France, September, 2006

Xue, L. (2008). *Laser Consolidation--a Rapid Manufacturing Process for Making Net-Shape Functional Components*, Industrial Materials Institute, National Research Council of Canada, London, Ont.

Yang, D.Y., Kim, H.C., Lee, S.H., Ahn, D.G. & Park, S.K. (2005). Rapid Fabrication of Large-Sized Solid Shape Using Variable Lamination Manufacturing and Multi-Functional Hotwire Cutting System. *Proceedings of SFF Symposium*, 1053-2153, Austin, TX, August, 2005

3axis, Direct Metal Laser Sintering (Dmls-Eos), Retrieved on Jan. 14, 2010, http://www.3axis.us/direct_metal_laser_slintering_dmls.asp

12

Environmental Burden Analyzer for Machine Tool Operations and Its Application

Hirohisa Narita
School of Health Sciences, Fujita Health University
Japan

1. Introduction

The manufacturing technologies have been evolving according to development of evaluation methods for quality, productivity and cost. In spite of the need for eco-manufacturing, the relationship between manufacturing technologies and environmental burdens has not been yet revealed. Machining is a major manufacturing activity, hence an analyzer developed evables to estimate the environmental burden due to machine tool operations. The machine tool operations has a big potential influence regarding envirnmental burden, and then the envirnmental burden analyzer is developed based on LCA (life cycle assessment) (SETAC, 1999).

Some environmental burden analyses for machine tool operations have the problem which isn't suited to evaluating cutting conditions in detail (Shimoda, 2000, Touma et al., 2003). For example, the conventional methods can evaluate only the difference among dry, wet and MQL (minimal quantities of lubricant) machining operations, but not the difference among depth of cuts, feed rate, spindle speed and tool path pattern. Furthermore, if removal volume and material type are same, the environmental burden becomes same. That is to say accurate environmental burden can't be provided for deciding the cutting conditions.

One research proposed manufacturing planning with consideration of multi-endpoint environmental effects, but any concrete evaluation ways of machining operation haven't been presented (Hara et al., 2005, Sheng et al., 1998). The other researches discussed environmental burden based on energy consumption(Diaz et al., 2010), but did not cmprehensicely evaluate the related environmental burden. I should thus be able to develop the envirnmental burden analyzer for machine tool operations to realize sustainable manufacturing (OECD, 2009). I also proposed a decision method of cutting conditions to achieve minimum environmental burden with using the analyzer developed. The analyzer will enable to accelerate the development of environmental technologies and eco-industries. A calculation algorithm of environmental burden, a system overview and some application example are described in this paper.

2. Life Cycle Assessment

Generally, LCA is a very useful methodology for estimating the environmental burden of a product or service associated with all stages: row-material production, manufacture, transportation, use, repair and maintenance, and disposal or recycling. LCA has four processes: goal and scope definition, inventory analysis, impact assessment, and interpretation, and then realizes holistic assessment of environmental aspect and helps a more informed decision regarding product design modifications and business strategies.

This paper proposes an application technique of LCA to machining processes. For this purpose, machining process and machine tool models in computer environment is constructed and an environmental burden analyzer is developed. Cutting conditions achieving low environmental burden are also discussed by using the analyzer developed.

3. System overview

Figure 1 explains an overview of the environmental burden analyzer for machine tool operations. A workpiece, cutting tool models and an NC program are entered to the analyzer, all activities including a machine tool operation and a machining process are made an estimation. At that time, electric consumption of a machine tool peripheral devices and motors, cutting tool's wear, coolant quantity, lubricant quantity, metal chip quantity and other factors are calculated. Here, the other factors correspond to electric consumptions of air conditioning, light, AGV's transportation, products washing and etc. Using these calculated factors, emission intensity data and resource data, the environmental burden is obtained, when a part is machined. The emission intensity is the rate of an emission matter for an impact category. For example, quantity of carbon dioxide emitted per joule of energy produced for global warming. The emission intensity data is prepared according to an impact category such as global warming, acidification, toxicity to ecosystem, toxicity to human, eutrophication and nuclear radiation. The resource data also is a machine tool specification data, cutting tool parameters and physical parameters of the cutting force for the estimation of machining process. The cutting tool parameter corresponds to tool's diameter, helical angle, rake angle and number of tooth.

This analyzer can calculate the environmental burden in various cutting conditions, because a machining process is evaluated properly. This is a novel aspect of this research as compared with the conventional approach.

Generally, an impact category must be de determined and relevant emission factors of the impact category are selected for LCA. Global warming is determined as an impact category and carbon dioxide (CO_2), methane (CH_4) and dinitrogen monoxide (N_2O) are selected as the it's relevant emission factors. Influences of halocarbon, and sulfur hexafluoride (SF6) on global warming are well known. However, these relevant emission factors are ignored, because their emission intensities have not been found according to my survey. All emissions are converted to equivalent CO_2 emission by multiplying them by characterization factors, and then total equivalent CO_2 emission is calculated as the environmental burden. The global warming potential (GWP) of 100-year impact (IPCC, 2007) is used as the characterization factors, as shown in Table 1.

	CO_2	CH_4	N_2O
Global warming potential (GWP)	1	25	298

Table 1. Characterization factors of global warming (IPCC, 2007)

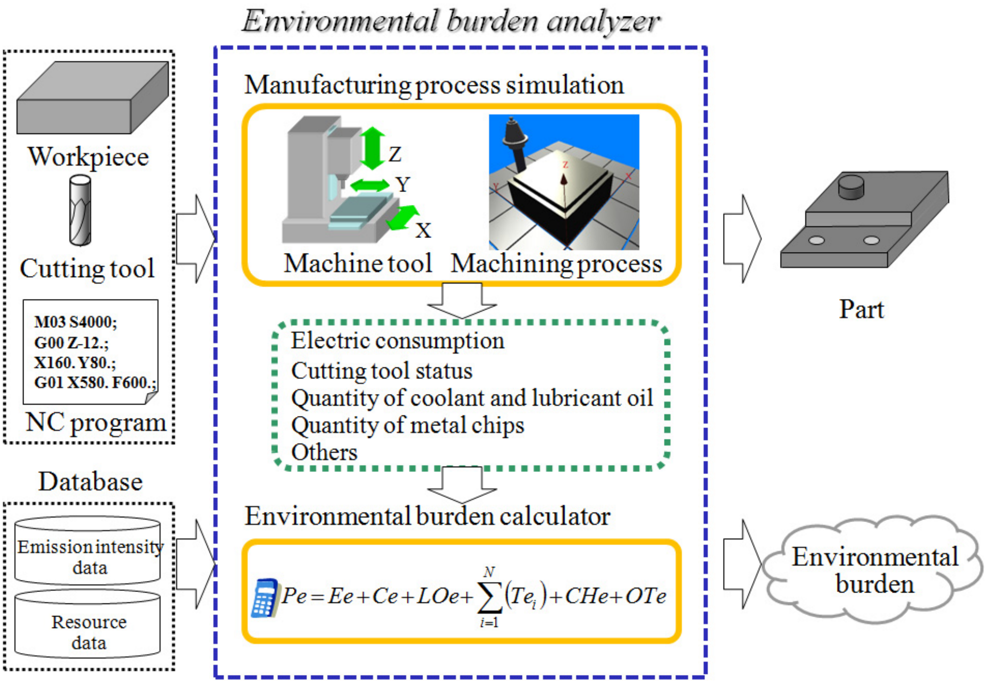


Fig. 1. An overview of the environmental burden analyzer for machine tool operations

4. A calculation algorithm of environmental burden

4.1 Total environmental burden due to a machine tool

Eq.(1) describes total environmental burden due to a machine tool operation, which is obtained from the electric consumption of the machine tool, the coolant, the lubricant oil, the cutting tool's wear, the metal chip and the other factors.

$$Pe = Ee + Ce + LOe + \sum_{i=1}^{N}(Te_i) + CHe + OTe \tag{1}$$

Pe: Environmental burden of machining operation [kg-CO_2]
Ee: Environmental burden of machine tool component [kg-CO_2]
Ce: Environmental burden of coolant [kg-CO_2]
LOe: Environmental burden of lubricant oil [kg-CO_2]
Te: Environmental burden of cutting tool [kg-CO_2]
CHe: Environmental burden of metal chip [kg-CO_2]
OTe: Environmental burden of other factors [kg-CO_2]
N: Number of cutting tool used in an NC program

OTe isn't introduced in this paper. Calculation algorithms of these factors (*Ee*, *Ce*, *LOe*, *Te* and *CHe*) are introduced in detail as follows.

4.2 Electric consumption of machine tool (*Ee*)

Ee means the environmental burden due to the electric consumption in an NC program. Figure 2 shows the electric consumption model of a machine tool, and the environmental burden due to the electric consumption of machine tool is described as follows.

$$Ee = k\ (SME + SPE + SCE + CME + CPE + TCE1 + TCE2 + ATCE + MGE + OAE + COE + CUE + SBE) \quad (2)$$

k: CO_2 emission intensity of electricity [kg-CO_2/kWh]
SME: Electricity consumption of servo motors [kWh]
SPE: Electricity consumption of a spindle motor [kWh]
NCE: Electricity consumption of an NC controller [kWh]
SCE: Electricity consumption of a cooling system of spindle [kWh]
CME: Electricity consumption of a compressor [kWh]
CPE: Electricity consumption of a coolant pump [kWh]
TCE1: Electricity consumption of a lift up chip conveyor [kWh]
TCE2: Electricity consumption of a chip conveyor in machine tool [kWh]
ATCE: Electricity consumption of an auto tool changer (ATC) [kWh]
MGE: Electricity consumption of a tool magazine motor [kWh]
OAE: Electricity consumption of an oil air compressor [kWh]
COE: Electricity consumption of an oil mist compressor [kWh]
CUE: Electricity consumption of a chip air blow compressor [kWh]
SBE: Stand-by energy of a machine tool [kWh]

In Eq. (2), the electric consumption of peripheral devices such as an NC controller, a cooling system of spindle, a compressor, a coolant pump, a lift up chip conveyor, a chip conveyor in machine tool, an ATC, a tool magazine motor and stand-by energy are calculated by their running times. But *CPE*, *TCE1*, *TCE2*, *COE* and *CHE* may not be used according to a machine tool operation with or without coolant usage. We must survey through the

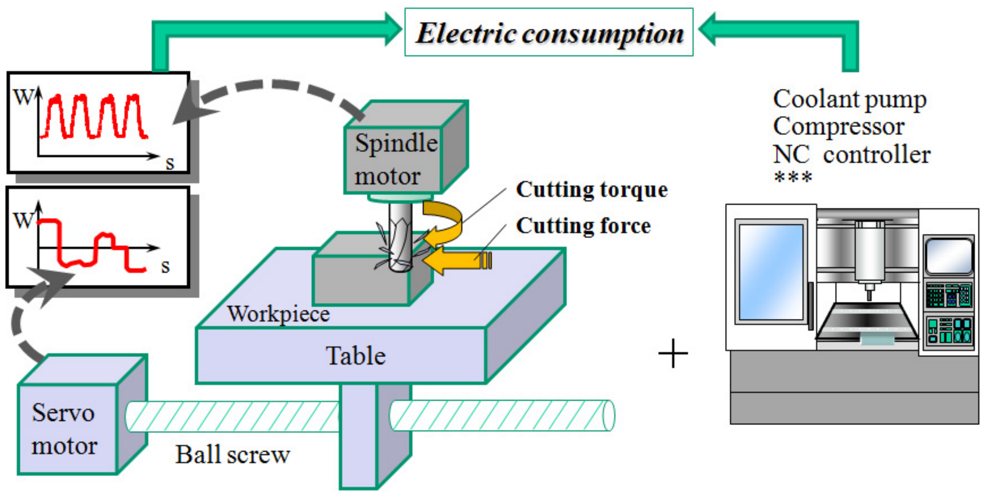


Fig. 2. Electric consumption model for machine tool

machine tool specification previously whether or not to consider these electric consumptions. The electric consumption of the servo motors and spindle motor is also varied dynamically according to the machining process, hence new analysis model must be constructed.

The load torques of servo motors are calculated as follows.

$$T_{Lservo} = T_U + T_M \tag{3}$$

$T_{L\,servo}$: load torque of servo motor [N·m]
T_U : axis friction torque [N·m]
T_M : application torque of ball screw [N·m]

Here, T_U is a torque due to rubber sealing and can't be obtained theoretically, thus its value is decided by an experiment. T_M is calculated as follows.

$$T_M = \frac{(\mu \cdot M \mp f) \cdot l \cdot \cos\theta \pm (M - f) \cdot l \cdot \sin\theta}{2\pi \cdot \eta} \tag{4}$$

μ: Friction coefficient of slide way
η: Transmissibility of ball screw system
l: Ball screw lead [m]
M: Moving part weight (table and workpiece) [N]
f: Cutting force in an axis [N]
θ: Gradient angle from horizontal plane [rad]

This equation is reconstructed by a monitoring method for cutting force (Fujimura and Yasui, 1994) with normal and reverse rotations of the servo motor. θ is 0 in the X- and Y-axes, and $\pi/2$ in the Z-axis. The cutting force in an axis, f, and the load torque of the spindle motor, $T_{L\,spindle}$, are calculated from the cutting force model. Virtual machining simulator I devloped (Narita et al., 2006) is applied for the purpose of estimating the aforementioned cutting force and cutting torque.

The calculated motor torque is converted to electric consumption as follows.

$$P = \frac{2\pi}{60} \times n \times T_L \times t \tag{5}$$

P: Electric consumption [Wh]
TL: Load torque ($T_{L\,servo}$ or $T_{L\,spindle}$) [N·m]
n: Motor rotation speed [rpm]
t: Time [hr]

The electric consumption of the compressors corresponding to *CPE*, *TCE1*, *TCE2*, *COE* and *CHE*, is also varied according to a discharge pressure and the air dryer, hence the following equation is adapted for the estimation.

$$Ec = CEn + CF + ADFr + ADFa \tag{6}$$

Ec: Electric consumption related to compressor [kW]

CEn: Electric motor power of main motor for compressor [kW]
CF: Rated electric motor power of fan motor for compressor [kW]
ADFr: Rated power of cooling machine of air-dryer [kW]
ADFa: Rated power of cooling fan of air-dryer [kW]

4.3 Coolant (*Ce*)

Ce means the environmental burden due to the coolant in an NC program. There are two types cutting fluid, hence two equations are proposed for *Ce*. Regarding water-miscible cutting fluid, coolant is generally used to enhance machining performance and circulated in a machine tool by coolant pump until coolant is made replacement. During this period, some coolants are eliminated by adhesion to metal chips, hence coolant is supplied for this compensation. The reduction of the dilution fluid (water) due to vapor has to be also considered to estimate total coolant. Here, the following equation is adopted to calculate the environmental burden due to the coolant.

$$Ce = \frac{CUT}{CL} \times \left\{ (CPe + CDe) \times (CC + AC) + WAe \times (WAQ + AWAQ) \right\} \qquad (7)$$

CUT: coolant usage time in an NC program [s]
CL: mean interval of coolant update [s]
CPe: environmental burden of cutting fluid production [kg-CO_2/L]
CDe: environmental burden of cutting fluid disposal [kg-CO_2/L]
CC: initial coolant quantity [L]
AC: additional supplement quantity of coolant [L]
WAe: environmental burden of water distribution [kg-CO_2/L]
WAQ: initial quantity of water [L]
AWAQ: additional supplement quantity of water [L]

4.4 Lubricant oil (*LOe*)

LOe means the environmental burden due to the lubricant oil in an NC program. The lubricant oil is used for two main types. One is for a spindle, another is for a slide way. Here, Minute amounts of oil are supplied by a pump to the spindle and the slide way within a specific interval. Grease as a lubricant is not introduced here, but the same equations can be applied to calculate the environmental burden due to grease. The environmental burden due to the lubricant oil is calculated as follows.

$$LOe = \frac{SRT}{SI} \times SV \times (SPe + SDe) + \frac{LUT}{LI} \times LV \times (LPe + LDe) \qquad (8)$$

SRT: Spindle runtime in an NC program [s]
SV: Discharge rate of spindle lubricant oil [L]
SI: Mean interval between discharges [s]
SPe: Environmental burden of spindle lubricant oil production [kg-CO_2/L]
SDe: Environmental burden of spindle lubricant oil disposal [kg-CO_2/L]
LUT: Slide way runtime in an NC program [s]
LI: Mean interval between supplies [s]

LV: Lubricant oil quantity supplied to slide way [L]
LPe: Environmental burden of slide way lubricant oil production [kg-CO_2/L]
LDe: Environmental burden of slide way lubricant oil disposal [kg-CO_2/L]

4.5 Cutting tool (*Te*)

Te means the environmental burden due to the cutting tools in an NC program. All cutting tools are managed by tool life, hence a tool life is compared with machining time to calculate the environmental burden in a machining. Cutting tools, especially for solid end mills, are often renewed by regrinding. The environmental burden due to the cutting tool is calculated as follows by considering the aforementioned processes.

$$Te = \frac{MT}{\sum_{j=1}^{RN+1} TL_j} \times \left(\left(TPe + TDe \right) \times TW + RN \times RGe \right) \tag{9}$$

MT: Machining time [s]
TL: Tool life [s]
TPe: Environmental burden of cutting tool production [kg-CO_2/kg]
TDe: Environmental burden of cutting tool disposal [kg-CO_2/kg]
TW: Tool weight [kg]
RGN: Total number of regrinding processes
RGe: Environmental burden of regrinding [kg-CO_2]

4.6 Metal Chips (*CHe*)

CHe means the environmental burden due to the metal chips in an NC program.

Metal chip recycling, in which a chip compactor, a chip crusher, a centrifugal separator and an arc furnace are used, generates an environmental burden. In this research, the environmental burden is calculated with considering chip weight as follows.

$$CHe = \left(WPV - PV \right) \times MD \times WDe \tag{10}$$

WPV: Workpiece volume [cm^3]
PV: Product volume [cm^3]
MD: Material density of workpiece [kg/cm^3]
WDe: Environmental burden of metal chip processing [kg-CO_2/kg]

4.7 Output example of the developed analyzer

Figure 3 shows an output example of the developed analyzer. The left part shows the machining process instructed by an NC program and estimate cutting force, cutting torque, electric consumption, quantity of cutting oil, quantity of lubricant oil, usage time of cutting tool and metal chip volume. The right part also shows the environmental burden calculated by the aforementioned algorithm. This analyzer can evaluate various environmental burdens by inputting the emission intensities related to the impact categories.

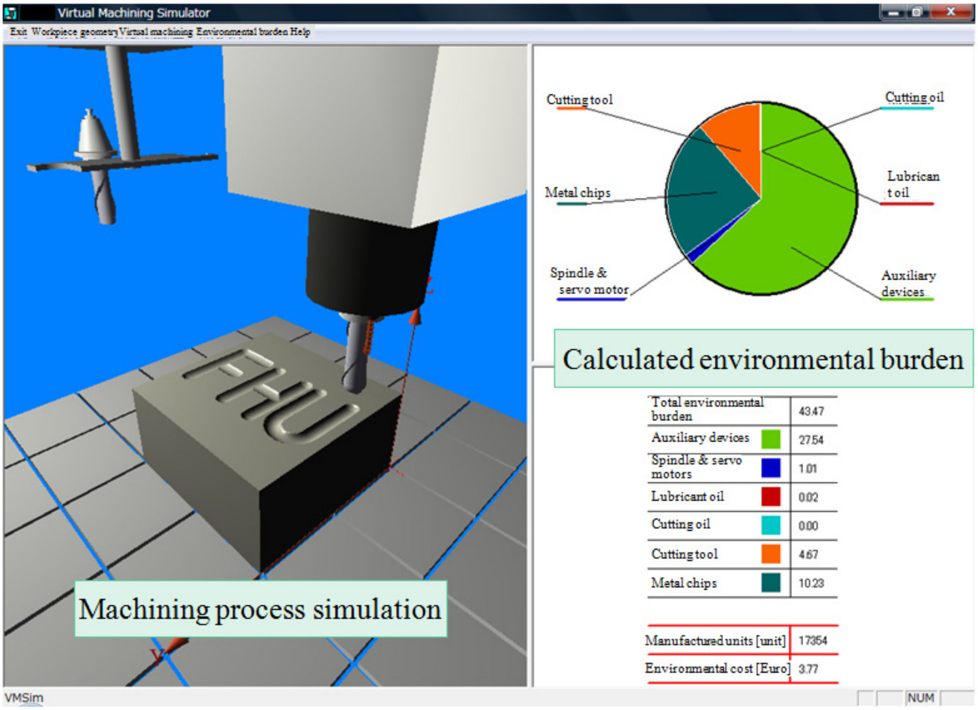


Fig. 3. An output example of the developed analyzer

5. Case study

5.1 Comparison of cutting conditions

For this case study, a machine tool is a vertical machining center (MB-46VA, OKUMA Corp.), a cutting tool is carbide-square end mill with 6mm-diamater and 2-flulte, and a workpiece is a medium carbon steel (S50C) and a compressor is a screw compressors (SCD-110JC, Anest Iwata Corp.).

The parameters to calculate the electric consumption of the servo motor of the machine tool and the electric powers of the peripheral devices of the machine tool are summarized in tables 3 and 4, respectively. These values have been measured and obtained from an instruction manual of the machine tool. The friction torque of servo motors also has been determined by experiment in advance. Table 4 shows the other parameters regarding a machine tool operation.

Table 5 shows the CO_2 emission intensities required to calculate the environmental burden of machining operations. These values were cited from some reports, such as environmental reports, technical reports, homepages and industrial tables (Tokyo Electric Power Company, (2005, Bureau of Waterworks Tokyo Metropolitan Government, 2003, Nansai et al., 2002, Osaka prefecture, 2003, 2005, Mizukami, 2002).

Table part weight [kg]	X: 903, Y:230, Z: 512
Friction coefficient of slide way	All axes: 0.01 (linear guide)
Ball screw lead [mm]	X, Y: 20, Z: 16
Transmissibility of ball screw system	0.95

Table 2. Parameters to calculate the electric consumption of the servo motor

NC controller [kW]	0.16
Cooling system of spindle [kW]	0.45
Compressor [kW]	1
Coolant pump [kW]	0.25
Lift up chip conveyor[kW]	0.1
Chip conveyor in machine tool [kW]	0.6
ATC [Wh]	0.08
Tool magazine [Wh] (1 round)	0.087
Vampire (stand-by) energy [kW]	0.64

Table 3. Electric consumptions and powers of machine tool

Initial cutting fluid quantity [L]	8.75
Additional supplement of cutting fluid [L]	4.3
Initial dilution fluid quantity [L]	175
Additional supplement of dilution fluid [L]	82.25
Mean interval between replacements of coolant in pump [Month]	5
Discharge rate of spindle lubricant oil [mL]	0.03
Mean interval between discharges for spindle lubrication [s]	480
Lubricant oil supplied to slide way[mL]	228
Mean interval between supplies [hour]	2000
Tool life [s]	5400
Total number of regrinding processes	2
Material density of cutting tool [g/cm^3]	11.9
Material density of workpiece [g/cm^3]	7.1
Coolant tank capacity of machine tool [L]	175

Table 4. Other parameters for machine tool operation

Electricity [kg-CO_2/kWh]	0.381
Cutting fluid production [kg-CO_2/L]	0.978
Cutting fluid disposal [kg-CO_2/L]	0.0029
Dilution liquid (water) [kg-CO_2/L]	0.189
Spindle and slide way lubricant oil production [kg-CO_2/L]	0.469
Spindle and slide way lubricant oil disposal [kg-CO_2/L]	0.0029
Cutting tool production [kg-CO_2/kg]	33.7
Cutting tool disposal [kg-CO_2/kg]	0.0135
Regrinding [kg-CO_2/number]	0.0184
Metal chip processing [kg-CO_2/kg]	0.322

Table 5. Equivalent CO_2 emission intensities

Figure 4 shows a part shape and a tool path pattern used for an example and table 6 shows the cutting conditions of NC program. Three cutting condition: Program 1, Program 1 with coolant (water miscible type) and Program 2 are evaluated. Here, a feed rate of immersion to workpiece is 100 mm/min and the tool life is assumed to be increased to 1.5 times of the original one due to the coolant effect.

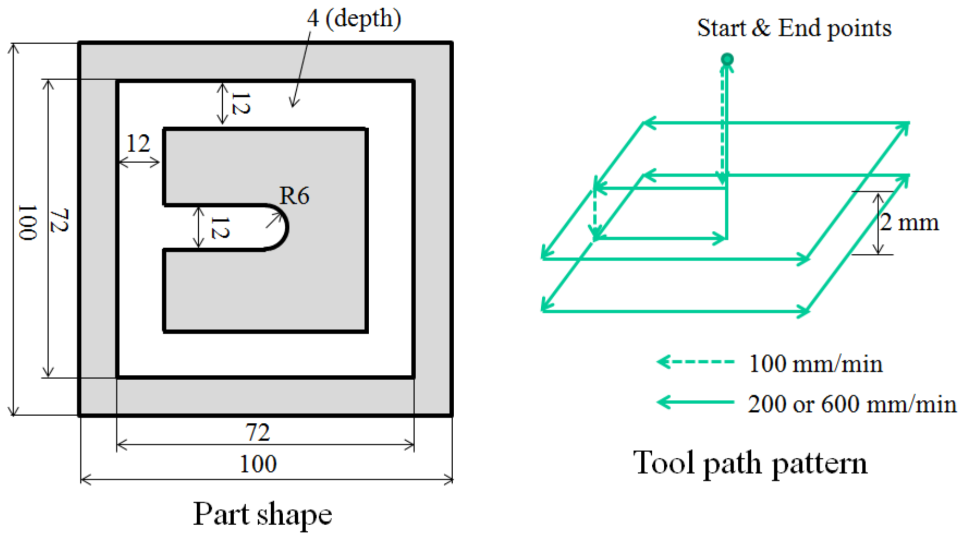


Fig. 4. Part shape and tool path pattern

	Program 1	Program 2
Spindle speed [rpm]	2500	7500
Feed rate [mm/min]	200	600

Table 6. Cutting conditions of NC programs

Figure 5 shows calculated results of three cutting conditions. Program 2 is best in three cutting conditions, because the machining time is very short. The environmental burden due to the cutting tool is reduced by the coolant effect, but the one due to electric consumption of peripheral devices are increased by the usage of coolant pump. As a matter of course, the one of coolant is increased but small. It is found that main reason of the increase of environmental burden due to the coolant usage is the one due to the peripheral devices as shown in this figure. As shown in this case study, the developed analyzer can evaluate various cutting conditions in details.

5.2 Determination method to realize low environmental burden

The environmental burden due to the peripheral devices must be reduced in order to reduce total environmental burden as shown in Fig.5. The one due to the peripheral devices is proportional to time, hence high speed milling might be effect.

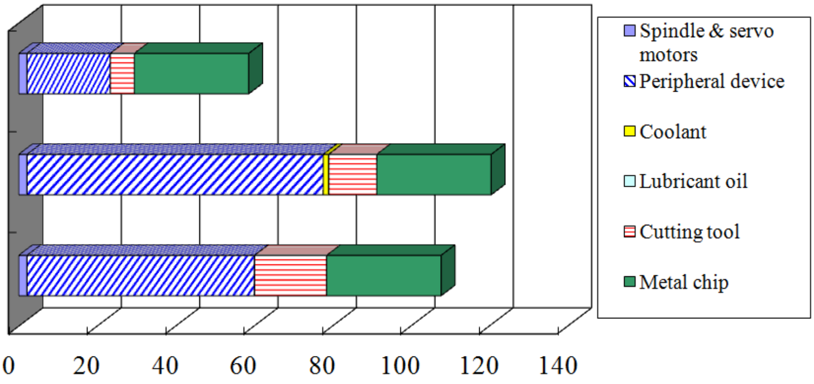


Fig. 5. Comparison of various cutting conditions

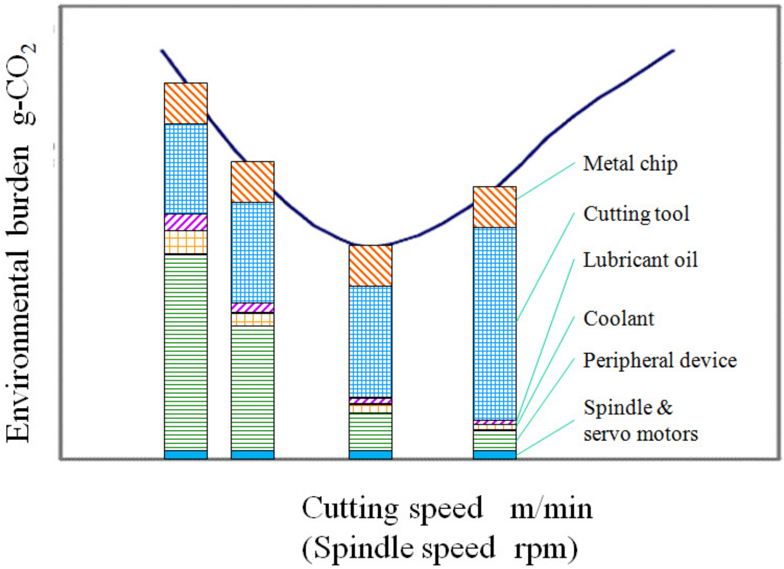


Fig. 6. Tendency due to high speed milling

Here, the relationship between the environmental burden and the cutting speed (spindle speed) is discussed when a feed per tooth, a radial depth and an axial depth of cuts are constant. Figure 6 shows a tendency due to the high speed milling. A tool wear becomes extremely large [Jiang, 2011] and a tool life will shorten in high speed millings, hence the environmental burden due to the cutting tool will increase. But the one due to the electric consumption, coolant and lubricant oil is proportional to time. That is to say there is a trade-off relation between the one due to the cutting tool and the one due to the electric consumption, coolant, lubricant oil. However the one due to the spindle and servo motors is very small (it is constant when the feed per tooth is constant), and the one due to the metal chip is constant, and then these environmental burdens are ignored for discussing. As shown in fig.6, there is a cutting condition to realize the minimum environmental burden. Hence, an optimum cutting speed (spindle speed) can be obtained automatically by calculating an approximate equation with using least-square method and exploring the cutting conditions achieving the minimum environmental burden with using iterative calculation. An optimum cutting speed (spindle speed) is attempted to be calculated as an example by embedding the aforementioned functions to the analyzer. A parabolic equation is applied for the approximate equation in this research.

A real tool wear data (Anzai, 2003) is used in order to confirm the tendency depicted in Fig.6. For this case study, cutting tool is a ball end mill with R10 and 2-flute, and workpiece is PX5. Cutting speed is varied from 50 to 550m/min, the axial depth is 0.5 mm, the radial depth is 0.8 mm, the feed per tooth is 0.15 mm/tooth and the cutting length is 56.25m. The coolant is also used for this cutting. Figure 7 describes the relation of tool wears according to cutting speed. Here, a flank wear is used to distinguish its tool life and then the threshold of maximum tool wear is assumed to be 0.8 mm, and the tool life in time domain is obtained.

Figure 8 shows a relation of equivalent CO_2 emission according to the cutting speeds. The approximate equation is obtained regarding the plotted data as follows.

$$y = 3.38 \times 10^{-2} x^2 - 27.0x + 6.02 \times 10^3 \tag{10}$$

Where, y means equivalent CO_2 emission and x means cutting speed. The minimum cutting speed is obtained by the iterative calculation and becomes 398.9 m/min (about 12702 rpm).

A research introduces an importance about the decision of an optimum cutting condition achieving low environmental burden with using virtual reality technology before a real machining operation (Shao, 2010) based on my previous research (Narita, 2009), but any concrete ways to decide the optimum cutting condition from the view point of the environmental burden haven't been proposed so far. This is the first proposition how to decide the optimum cutting conditon achieving low environmental burden as show in this example. I believe the feasibility of the environmental burden analyzer can be described in this paper.

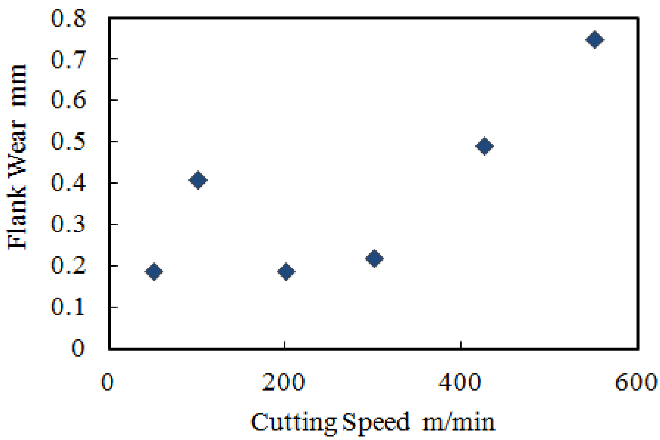


Fig. 7. Tool wears according to cutting speed (Anzai, 2003)

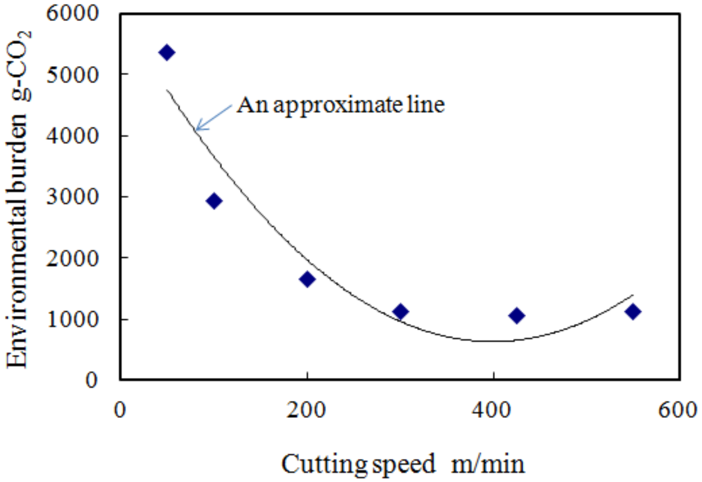


Fig. 8. Environmental burden vs. cutting speed

6. Conclusion

Conclusions are summarized as follows:

1. A algorithm to calculate environmental burden due to machine tool operations was proposed and the environmental burden analyzer for machine tool operations was developed.
2. A decision method of cutting conditions to achieve minimum environmental burden with using the developed analyzer was also proposed.
3. The feasibility of the environmental burden analyzer and the decision method of cutting conditions to achieve minimum environmental burden were demonstrated through examples.

7. Acknowledgment

I would like to express my sincere appreciation to the stuff of OKUMA Corp. for thoughtful support.

8. References

Anzai, M., The Cutting Characteristic of Mold Steel Under The High-Speed Milling Condition, *Nachi Technical Report*, Vol. 11 A1, 2006 (in Japanese).

Bureau of Waterworks Tokyo Metropolitan Government, (2003), Environmental report of Tokyo waterworks 2002(in Japanese).

Diaz, N., Choi, S., Helu, M., Chen, Y., Jayanathan, S., Yasui, Y., Kong, D., Pavanaskar, S. and Dornfeld, D., , (2010), Machine Tool Design and Operation Strategies for Green Manufacturing, *Proceedings of 4th CIRP International Conference on High Performance Cutting* , Gifu, Japan, pp.271-276.

Fujimura, Y., Yasui, T. , (1994), Machine tool and manufacturing system, KYORITSU SHUPPAN Co. Ltd., (in Japanese).

Hara, Y., Tanaka, T. and Saito, Y. , (2005), Evaluation Method of the Power Consumption for Machine Tool, Proceeding of the 2005 JSPE Autumn Annual Conference (in Japanese) , Kyoto, Japan, pp.744-745.

IPCC Fourth Assessment Report (AR4) , (2007), Changes in Atmospheric Constituents and in Radiative Forcing.

Jiang, B., Qi, C., Xia, D., Cao, S., and Chen B. , (2011), Dynamics Characteristic of Cutter Wear in High Speed Milling Hardened Steel, Advanced Science Letters Vol. 4, 3103–3107

Mizukami, H., Yamaguchi, R., Nakayama, T., Maki, T., , (2002), Off-gas Treatment Technology of ECOARC, *NKK Technical Report (in Japanese)*, No.176, pp.1-5.

Nansai, K., Moriguchi, Y. and Tohno, S. , (2002), Embodied Energy and Emission Intensity Data for Japan Using Input-Output Tables (3EID)–Inventory Data for LCA-, Center for Global Environment Research, National Institute of Environmental Studies, Japan.

Narita, H., Chen, L.Y., Fujimoto, H., Shirase, K. and Arai, E. , (2006), Trial-Less Machining Using Virtual Machining Simulator for Ball End Mill Operation, *International Journal of the Japan Society of Mechanical Engineers*, Series C, Vol.49, No.1, pp. 50-55.

Narita, H. and Fujimoto, H., Analysis of Environmental Impact Due to Machine Tool Operation, International Journal of Automation Technology, Vol.3, No.1, pp.49-55, 2009.

Organisation for Economic Co-operation and Development (OECD), (2009), Sustainable Manufacturing and Eco-innovation: Towards a Green Economy, OECD June 2009 Policy Brief.

Osaka prefecture, (2003), guideline index of waste material incinerator in Osaka prefecture, Osaka prefectural announcement No.618 (in Japanese).

Osaka prefecture, (2005), Osaka prefectural statistical yearbook 2004 (in Japanese).

SETAC, (1993), Guidelines for Life-Cycle Assessment: A Code of Practice.

Shao, G., Kibira, D. and Lyons, K., (2010), A Virtual Machining Model For Sustainability Analysis, *Proceedings of ASME 2010 International Design Engineering Technical Conference & Computers and Information in Engineering Conference*, DETC2010-28743, August 15-18, Montreal, Quebec, Canada.

Sheng, P., Bennet, D., Thurwachter, S., / B.F. von Turkovich, (1998), Environmental-Based Systems Planning for Machining, *Annals of the CIRP*, Vol.47/1, p.409-414.

Shimoda, M. , (2000), LCA case of machine tool, *Symposium of 2002 Japan Society for Precision Engineering Spring Annual Meeting "Leading-Edge Trend of Environmental Impact Evaluation for Inverse Type Design and Manufacturing"*, Nagoya, Japan, p. 37-41, (in Japanese).

Tokyo Electric Power Company, (2005), The Earth, People & Energy TEPCO Sustainability Report 2004

Touma, S., Ohmori, S., Kokubo, K., Tateno, M. , (2003), Evaluation of Environmental Burden in Eco-friendly Machining Method using Life Cycle Assessment Method –Estimation of Carbon Dioxide Emission in Eco-friendly Turing Method– *Journal of the Japan Society for Precision Engineering*, (in Japanese), Vol.69, No.6, p.825-830.

13

Digital Manufacturing Supporting Autonomy and Collaboration of Manufacturing Systems

Hasse Nylund and Paul H Andersson
Tampere University of Technology
Finland

1. Introduction

This chapter discusses on the challenges and opportunities of digital manufacturing supporting the decision making in autonomous and collaborative actions of manufacturing companies. The motivation is the change towards more networked collaboration caused by, for example, globally distributed markets and specialization of manufacturing companies to their core competences, their autonomous activities. This situation has led to increasingly complex manufacturing activities in the manufacturing network and the importance of collaboration has become a critical factor. In most cases companies seek to respond to the challenges through cooperation rather than expanding their own operations. The autonomy means that the parties involved in the manufacturing activities do their own tasks by themselves independently from other parties while the collaboration involves the activities that one party cannot do by itself and therefore, co-operation of several parties are required. This kind of situation can be clearly seen in networked manufacturing activities involving several companies, but similarly, inside a company and its one facility, same kind of autonomous and collaborative activities can be recognized. In the discussion, the dimensions of autonomy and collaboration are considered in designing and developing manufacturing systems, as well as in improving the daily operations.

The rest of this Chapter is structured as follows. Section 2 discusses on the main issues behind the research, including Competitive and Sustainable Manufacturing, changeability in manufacturing as well as support from digital manufacturing. In Section 3, a structure for manufacturing systems and entities is proposed, which is the base for the design and development activities of manufacturing systems discussed in Section 4. An academic research environment is introduced in Section 5 describing several of the theoretical aspects discussed before. Section 6 gives a brief conclusion on the topics discussed.

2. Background

The focus of the discussion is on mechanical engineering industry of discrete part manufacturing for business-to-business (B2B) industry, including their part manufacturing and product assemblies. These kinds of products are typically highly customized and tailored to customer needs and requirements with low or medium demand (Lapinleimu, 2001). This type of production usually involves several companies and is formed as a supply network. For example, the production includes a main company, its suppliers and suppliers of suppliers as well as customers and customers of customers. The current manufacturing paradigm, in the above context, has evolved from the early craft manufacturing via mass manufacturing towards mass customization. Typical characteristics that have been recognized include (Andersson, 2007):

- Globally local systems spread over industrial ecosystems and manufacturing networks of their own pros and cons.
- Managing the networked manufacturing, where the importance of procurement and management of knowledge flow increase.
- Specialization to one's core competences and collaborating with others in the manufacturing network.

Early discussions considered whether these characteristics could be fulfilled with developing existing flexible manufacturing systems (FMSs), or to shift to reconfigurable manufacturing systems (RMSs) paradigm. At some point, more ambitious goals were set with the aim to describe a manufacturing system with autonomous entities having the needed level intelligence to be changeable to organize themselves to altered situations, and to identify what new entities will be required. At the same time, a manufacturing system is required to be competitive in order to survive in the markets as well as sustainable to reduce or eliminate unwanted activities and outputs.

2.1 Competitive and sustainable manufacturing

The well-known definition of sustainability is: "The Sustainable Development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (World Commission on Environment and Development [WCED], 1987), thereafter (WCED, 1987). This political statement is the root cause for today's key global challenges and related problems that call for a drastic change of paradigm from economic to sustainable development. Competitive Sustainable Manufacturing (CSM) is seen as a fundamental enabler of such change (Jovane, 2009).

Sustainable development has been recently increasingly emphasized around the world; in Europe (Factories of Future Strategic Roadmap and the Manufuture initiative), the USA (Lean and Mean), and Japan (Monozukuri and New JIT). The CSM paradigm widens the classical view of sustainability to interact with the Social, Technological, Economical, Environmental, and Political (STEEP) context (AdHoc, 2009). Sustainable manufacturing is a multi-level approach where product development, manufacturing systems and processes as well as enterprise and supply chain levels need to be considered, with metrics identified for each level (Jawahir et al., 2009).

The CSM is one of the strategic research areas within the Department of Production Engineering (TTE) at Tampere University of Technology (TUT). Figure 1 presents the main areas of the CSM approach, consisting of three main pillars, Sustainable, Lean and Agile Manufacturing. Lean manufacturing aims to combine the advantages of craft and mass production, while avoiding the drawbacks such as the high costs of craft production and rigidity of mass production systems (Womack et al., 1990). For example, the Lean Enterprise Institute (2008) defines Lean manufacturing as "a business system for organizing and managing product development, operations, suppliers, and customer relations that requires less human effort, less space, less capital, and less time to make products with fewer defects to precise customer desires, compared with the previous system of mass production."

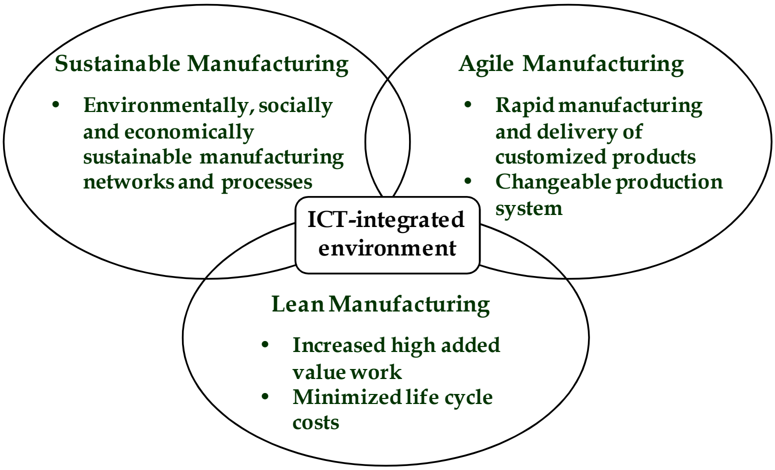


Fig. 1. The cornerstones of the CSM at the Department of Production Engineering (Nylund et al., 2010)

Agile manufacturing can be defined as an enterprise level manufacturing strategy of introducing new products into rapidly changing markets (Nagel & Dove, 1991) and an organizational ability to thrive in a competitive environment characterized by continuous and sometimes unforeseen change (Kidd, 1994). Agile manufacturing highlights the need to adapt to changes in the business environment, and generally agility is defined as ability to react to and take advantage of changes and opportunities, see for example (Sharifi & Zhang, 1999; Gould, 1997).

Sustainable development is the development that meets the needs of the present without compromising the ability of future generations to meet their own needs (WCED, 1987). It consists of three structural pillars namely society, environment, and economy, whilst at the same time it also involves operational aspects such as the consumption of resources, natural environment, economic performance, workers, products, social justice and community development (Jayachandran et al., 2006). When these three pillars of Lean, Agile, and Sustainable are considered as one system, Lean emphasized the stability of a system that can be referred as the autonomy while agility adds the needed capability to change to new situations, therefore focusing more on the collaboration. These two have their main focus on economic issues while sustainability adds the viewpoints of energy and environmentally friendly manufacturing.

2.1.1 Changeability in manufacturing systems

Wiendahl et al. (2007) suggest changeability as an important factor in the competitiveness of manufacturing companies in addition to the classical factors of cost, quality, and time. Changeability is defined on the five structuring levels of an enterprise: changeover ability, reconfigurability, flexibility, transformability and agility. Agility, which was discussed in the context of CSM, is seen from a manufacturing enterprise level and refers to the ability of an enterprise to effect changes in its systems, structure and organization (Gunasekaran & Yusuf, 2002).

Transformability is changeability at a factory level. It includes, for example, facilities, organization and employees. The whole factory is oriented towards the market to offer the right products and services (Wiendahl et al., 2007). Into a detailed level of manufacturing activities the term changeover ability is used. It is related to single workstations that perform manufacturing processes in order to manufacture product features.

Reconfigurability and flexibility are the most widely examined structuring levels of changeability in the context of manufacturing systems. An FMS is configured to deal with part variations within its scope. The functionality and capacity of FMSs are pre-designed, while flexibility is inherent and built-in a priori (ElMaraghy, 2005). Because of the fixed flexibility of FMS, it is not flexible enough for rapid and cost-effective reconfiguration in response to changing markets (Mehrabi et al., 2000). An RMS is composed of general-purpose hardware and software modules that are reused in reconfiguration tasks. Modules are replaced or added only if necessary. An RMS has the ability to change capacity and functionality to bring about the needed flexibility, i.e. to bring about exactly the functionality and capacity needed exactly when needed (Koren, 1999).

2.2 Support from digital manufacturing

The tools and principles of digital manufacturing, factories, and enterprises can offer significant value to all aspects of manufacturing systems during their life cycles. However, there are no commonly used or agreed definitions for those, but they usually share the idea of managing the typically isolated and separate manufacturing activities as a whole by the means of Information and Communications Technology (ICT) (Nylund and Andersson, 2011). Typical examples often found from the definitions, based on literature, are (see, for example: Bracht & Masurat, 2005; Maropoulos, 2003; Souza et al., 2006):

- An integrated approach to develop and improve product and production engineering technologies.
- Computer-aided tools for planning and analysing real manufacturing systems and processes.
- A collection of new technologies, systems, and methods.

Typical tools and principles of digital manufacturing on different structuring levels are, for example (Kühn, 2006):

- Computer-aided technologies, such as computer-aided design (CAD) and computer-aided manufacturing (CAM), e.g. offline programming for virtual tool path generation to detect collisions, analyse material removal and optimise cycle times.
- Visual interaction applications, e.g. virtual environments and 3D-motion simulations that offer realistic 3D graphics and animations to demonstrate different activities.
- Simulation for the reachability and sequences of operations as well as internal work cell layout and material handling design. These include, for example, realistic robotics simulation (RRS) and ergonomics simulation.

- Discrete event simulation (DES) solutions including the need for and the quantity of equipment and personnel as well as evaluation of operational procedures and performance. DES can also be focused on e.g. factories and supply chain or network sales and delivery processes as well as to complex networked manufacturing activities, including logistical accuracy and delivery reliability of increasing product variety.

The above are examples of typical application areas of digital manufacturing. In each case, the activities rely on up-to-date and accurate information and knowledge. The total information and knowledge of a manufacturing system can be explained with explicit and tacit components (Nonaka and Takeuchi, 1995). The explicit part of the knowledge can be described precisely and presented formally in ICT-systems. The skills of humans are explained as the tacit dimension of knowledge, which, presented digitally, may lead to unclear situations and can be wrongly understood. The importance of the transformation from tacit to explicit knowledge has been recognized as one of the key priorities of knowledge presentation (Chryssolouris et al., 2008).

Challenges exist both in the autonomous and collaborative parts of the digitally presented manufacturing entities. The internal part should include only the needed information and knowledge to fully describe the autonomous activities while the collaboration mostly relies on effective sharing of information and knowledge and therefore both the communication language and content should be described formally. Effective knowledge management consists of four essential processes: creation, storage and retrieval, transfer, as well as application, which are dynamic and continuous phenomenon (Alavi and Leidner, 2001). Examples of the application areas of the digital part are:

- Email messages, Internet Relay Chat (IRC), Instant Messaging, message boards and discussion forums.
- More permanent information and knowledge derived from the informal discussions, stored in applications such as Wikipedia.
- Internet search engines and digital, such as dictionaries, databases, as well as electronic books and articles
- Office documents, such as reports, presentations, as well as spreadsheets and database solutions.
- Formally presented information systems, such as Enterprise Resource Planning (ERP), Product Data Management (PDM), and Product Lifecycle Management (PLM).

The importance of the possibilities offered by ICT tools and principles is ever more acknowledged, not only in academia, but also in industry. The Strategic Multi-annual Roadmap, prepared by the Ad-Hoc Industrial Advisory Group for the Factories of the Future Public-Private Partnership (AIAG FoF PPP), lists ICT as one of the key enablers for improving manufacturing systems (AdHoc, 2010). The report describes the role of ICT at three levels; smart, virtual, and digital factories.

- Smart factories involve process automation control, planning, simulation and optimisation technologies, robotics, and tools for competitive and sustainable manufacturing.
- Virtual factories focus on the value creation from global networked operations involving global supply chain management.

- Digital factories aim at a better understanding and the design of manufacturing systems for better product life cycle management involving simulation, modelling and management of knowledge.

Both digitally presented information and knowledge as well as computer tools and principles for modelling, simulation, and analysis offer efficient ways to achieve solutions for design and development activities. General benefits include, for example:

- Experiments in a digital manufacturing system, on a computer model, do not disturb the real manufacturing system, as new policies, operating procedures, methods etc. can be experimented with and evaluated in advance in a virtual environment.
- Solution alternatives and operational rules can be compared within the system constraints. Possible problems can be identified and diagnosed before actions are taken in the real system.
- Modelling and simulation tools offer real-looking 3D models, animations, and visualisations that can be used to demonstrate ideas and plans as well as to train company personnel.
- Being involved in the process of constructing the digital manufacturing system tasks increases individuals' knowledge and understanding of the system. The experts in a manufacturing enterprise acquire a wider outlook compared to their special domain of knowledge as they need to gather information also outside their daily operations and responsibilities.

3. Structure of manufacturing entities and systems

The proposed structure of manufacturing systems consists of manufacturing entities as well as their related domains and activities. An entity, being autonomous, is something that has a distinct existence and can be differentiated from other entities. The term 'entity' has similarities to other terms, such as: object, module, agent, actor, and unit. A domain is an expert area in which two or more entities are collaborating. Domains have certain roles in the system and their own responsibilities and specific objectives. An activity is a set of actions that accomplish a task that is related to the entities and domains, as well as to their context.

3.1 Structure of manufacturing entities

Figure 2 illustrates the general viewpoints of the proposed structure of manufacturing entities. The structure is explained with internal structure of individual manufacturing entities. It is derived from the principles behind the term 'holon' and the concept of Holonic Manufacturing Systems (HMS). The term holon comes from the Greek word 'holos', which is a whole and the suffix '–on', meaning a part. Therefore the term holon means something that is at the same time a whole and a part of some greater whole (Koestler, 1989).

In HMS, holons are autonomous and co-operative building blocks of a manufacturing system, consisting of information processing part and often a physical processing part (Van Brussel et al., 1998). In this approach, the information part is divided into digital and virtual parts differentiating the digitally presented information and knowledge from the computer models representing the existing or future possible real manufacturing entities. The digital part barely exists as clearly consisting separate part. It can be distributed in several information systems both globally and locally and in information rich computer models, the virtual parts of the manufacturing entities.

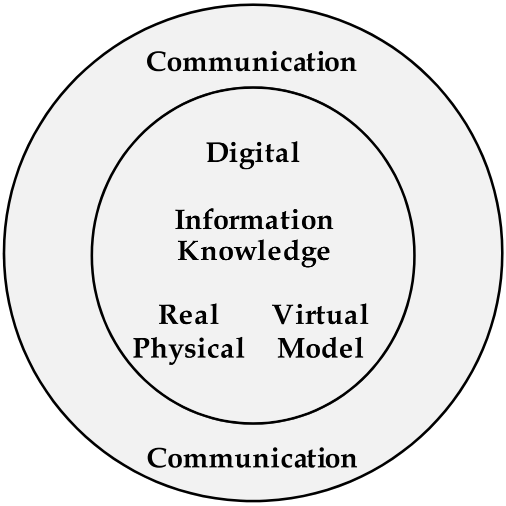


Fig. 2. Internal structure of manufacturing entities

The digital, virtual, and real parts combined present the autonomy of a manufacturing entity. The communication part is responsible of both the language and content of the messages between manufacturing entities. Therefore, it enables the manufacturing entities to collaborate with each other (Nylund & Andersson, 2011). As the autonomous entities exist distributed, independently from each other, they can be developed separately. At the same time, the communication part enables the investigation of the entities in an integrated fashion, and to develop the whole system they form.

The division into digital, virtual, and real is intentionally missing the tacit dimension, as it is intended to be used in decision making processes by humans, based on their skills and knowledge. At the end, the humans are the ones that are making the decisions, or are the ones that are creating the decision making mechanisms.

3.2 Structure of manufacturing systems

A manufacturing system consists of manufacturing entities with different roles as well as their related domains and activities. Figure 3 shows a general presentation of manufacturing entities of products, orders, and resources as well as their connecting domains of process, production, and business. The focus is on the manufacturing activities that are related to the transformation of raw material to finished products and their associated services as well as the flow of information and knowledge that is related to the physical manufacturing of customer orders.

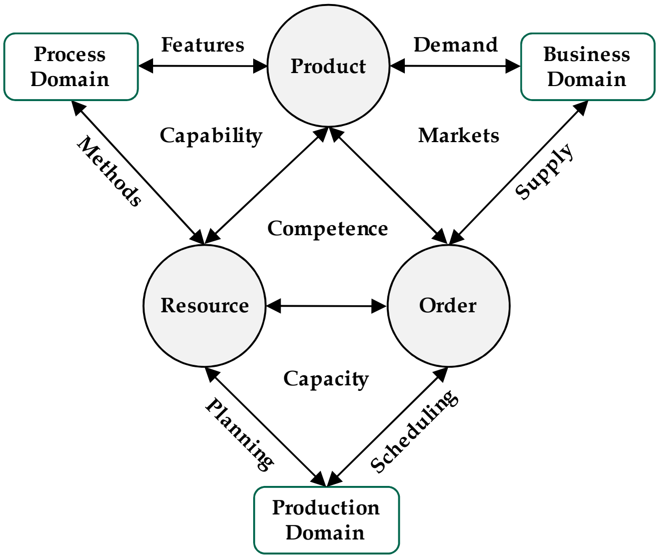


Fig. 3. Structure of manufacturing systems

The proposed structure is loosely based on the HMS reference architecture Product-Resource-Order-Staff Architecture (PROSA) (Van Brussel et al., 1998). The PROSA explains the relations between the entities with the information and knowledge they exchange while in this approach the relations are explained with activities occurring between the entities. Brief descriptions of the entities and domains are:

- *Products* represent what the manufacturing system offers to its customers. The characteristics of the products specify the requirements for the manufacturing system, i.e. what the system should be able to do.
- *Resources* embody what is available to manufacture the products. The characteristics of the resources determine what kinds of products can be manufactured.
- *Orders* represent instances of products that are ordered by customers. They define the volume and variation requirements of the products ordered, as well as the capacity and scalability requirements for the manufacturing system.
- *The process domain* represents the capabilities that are needed to manufacture the products. It connects the development activities of products and resources.
- *The production domain* defines the capacity and scalability to manufacture changing volumes and variations in customer orders. It handles the material and information flow of the manufacturing system.
- *The business domain* is responsible for markets, i.e. for the right products being available for the customers to gain enough orders.

3.3 Structuring levels in manufacturing

A fractal is an independently acting manufacturing entity that can be precisely described (Warnecke, 1993). Fractals are structured bottom-up, building fractals of a higher order.

Entities at the higher levels always assume only those responsibilities in the processes which cannot be fulfilled in lower order (Strauss & Hummel, 1995). This is similar to holons and holarchies, as at every fractal level of holons the level above is the holarchy of the holons at a lower level. Similarly, the autonomy of the holons is not considered in the holarchy, but instead dealing with and organizing the co-operation of the holons is the responsibility of the holarchy.

In Figure 4, four different structuring levels, manufacturing units, stages, plants, and networks, are distinguished.Manufacturing units correspond to individual machine tools that have certain manufacturing methods. The units are designated to manufacture the features of work pieces that have similarities in, for example, size and shape as well as tolerances and material properties. Typical areas are computer-aided design (CAD) and computer-aided manufacturing (CAM), e.g. offline programming for virtual tool path generation to detect collisions, analyse material removal and optimise cycle times (Kühn, 2006).

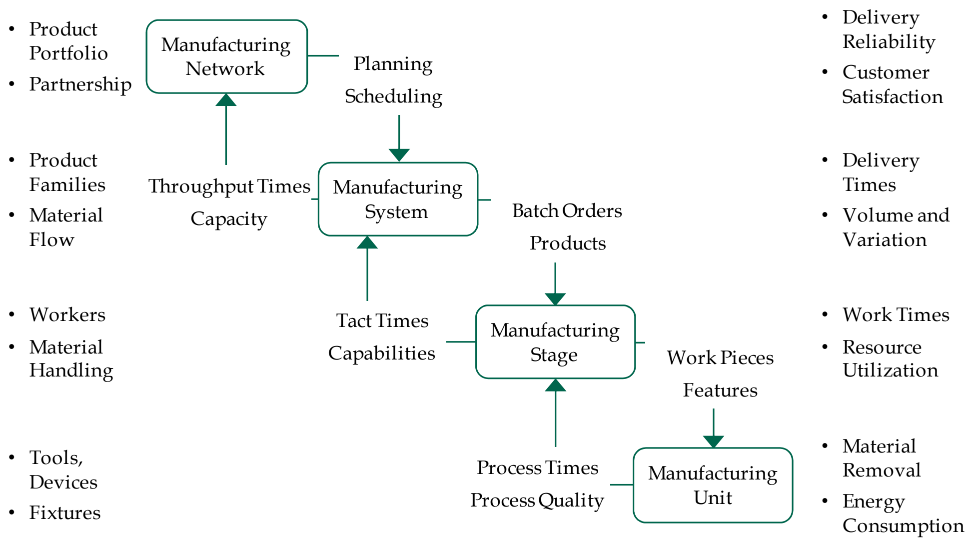


Fig. 4. Examples of structuring levels of manufacturing and their connections

Manufacturing stages are physical or logical manufacturing areas, e.g. manufacturing cells, consisting of one or more manufacturing units and their co-operation. Additionally, the manufacturing stages include internal material handling in moving the work pieces between the manufacturing units as well as buffers and stocks to hold batches of the work pieces. In manufacturing stages the focus can be on simulation for the reachability and sequences of operations as well as internal work cell layout and material handling (Kühn, 2006).

Manufacturing plants are composed of manufacturing stages, warehouses for storing the products as well as internal logistics to transfer material between the stages and material storing areas. They typically correspond to factories and have customers who can be other companies or internal customers, such as an assembly plant. Typical simulation issues concern the layout design and material flow analysis as well as planning and controlling the manufacturing activities. Simulation studies on a manufacturing plant level are usually conducted using discrete event simulation (DES) including the need for and the quantity of equipment and personnel as well as evaluation of operational procedures and performance. Manufacturing networks consist of factory units, which can exist globally. One of the key differences between plants and networks is that entities in the network often belong to different companies that may have contradictory goals in their strategies. Simulation can be focused on traditional supply chain sales and delivery processes as well as to complex networked manufacturing activities, including logistical accuracy and delivery reliability of increasing product variety.

4. Digital manufacturing support for manufacturing activities

A digitally presented manufacturing system contains the information and knowledge of manufacturing entities and activities that it is reasonable to represent in a digital form. This, at its best, makes possible efficient collaboration between all the manufacturing activities and related parties. The discussion on digital manufacturing support is based on a previously developed framework for extended digital manufacturing systems (EDMS). An EDMS can briefly be defined as follows (Nylund and Andersson, 2011):

- an integrated and collaborative environment for humans, machines, and information systems to act and interact;
- to enhance the research, development and management activities of products, production systems, and business processes,
- supporting knowledge-intensive decision-making in the entirety of their lifecycles.

4.1 From ideas to innovative solutions

Figure 5 represents a process from ideas and the need for change to innovative solutions. It consists of a chain of activities where the results evolve towards more precise solutions. Each phase has its enablers as inputs and the activity creates results as outputs. The results affect the enablers in the following phases of the process. The process is also iterative as it is possible to go back to previous phases in order to change or refine them. The need for change can arise, for example, from social, technological, economic, environmental, and political aspects.

The changes can also derive from voluntary ideas that are seen to improve the competence of the system. If the process has not been developed previously, the current system has to be analysed to create the digital information and knowledge of what currently exists. The synthesis of the existing system and possible changes form the new requirements for the future system. The combination of feasible new possibilities and existing capabilities forms the solution principles. The results are digital entities and abstract and conceptual descriptions, including the objectives and preliminary properties of the future system.

When the descriptions evolve towards a more detailed level, possible technologies can be investigated, resulting in alternative solutions. The solution alternatives can be modelled as virtual entities that include, in addition to their digital description, for example, 3D models with their own operating rules, motion, and behaviour. Combining the existing and new virtual entities forms a rough simulation model. The solution that is implemented has to be verified to make sure that the behaviour and co-operation of the entities in the system are modelled correctly. The verified simulation model can be used to run test experiments. By analysing the results from the simulation model and comparing them with known or predicted outcomes, the behaviour of the simulation model can be validated. When the simulation model is verified and validated, it can be used for manufacturing experiments. The experiments are used to analyse the behaviour of the system, and can lead towards innovative solutions.

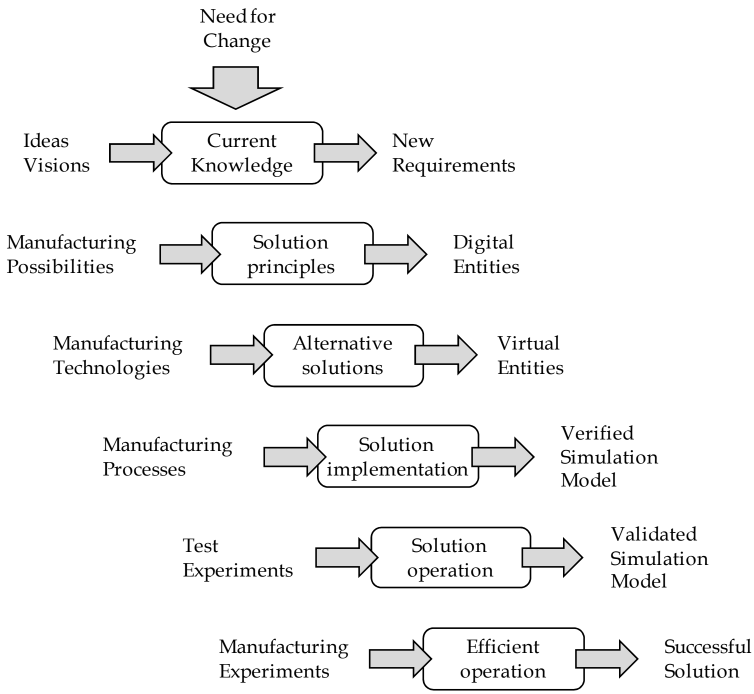


Fig. 5. The process from ideas to innovative solutions

4.2 Manufacturing process and flow development

Figure 6 shows a theoretical example of process and flow development. The manufacturing process part corresponds to the process domain, presented in Figure 3, where the capabilities for a manufacturing network are developed. The part of the manufacturing flow presents the production domain in Figure 3, aiming for the right capacity and scalability of the manufacturing network to meet the customer demands. The existing capabilities are combined with new possibilities, requirements, and constraints in the production network creating the synthesis of existing and what new capabilities will be required. These derive from, for example, new possible markets, customers, and competition i.e. what is important in the future that the current capabilities cannot fulfil. The new possible capabilities are tested virtually using computer-aided technologies in connection with the digitally presented information and knowledge.

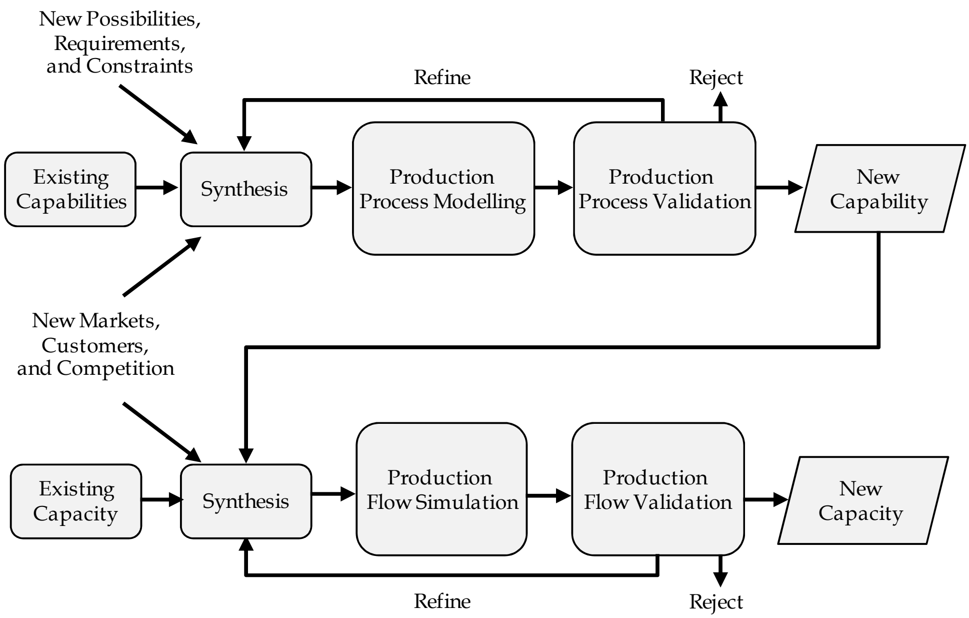


Fig. 6. Production process modelling and production flow simulation.

The resulted new capability is validated both to ensure that it does what it is supposed to do and that it meets the performance requirements, such as cost, quality, and time efficiency as well as the social and environmental issues. The production flow simulation in Figure 6 follows the same idea as the production process modelling. The new capability can add to the total capabilities of the network if something new is implemented, or change the existing capabilities if something already existing is reconfigured. It is not enough that all the needed capabilities exist.

The production flow simulation aims to define how much capabilities are required to produce the changing volume and variation of customer orders at the right time. Typical areas are the controlling, planning and scheduling of the activities. To investigate the production process modelling in more detail, five categories between product requirements and resource capabilities can be recognized, see Figure 7:

- Existing capability: The capabilities exist for all of the product requirements without any need for changes to the system. The products can be manufactured as the service requests have service providers.
- Possible existing capability: At least some of the product requirements need further investigation as to whether the capabilities exist or not. The requirements are close to the existing capabilities and, using modelling and simulation, the capabilities can be verified.
- Capability after reconfiguration: There is no existing capability but it may be possible to reconfigure the system so that it has the capabilities. By modelling the reconfigured system the possibility can be verified.

- Capability after implementation: The system does not have the needed capability. It may be possible if new capabilities are added to the system. Again this can be verified using modelling and simulation.
- No capability: The result may also be that there are no capabilities and they cannot be implemented either. This leads to the need for an alternative solution, which leads to a result that fits into one of the first four categories.

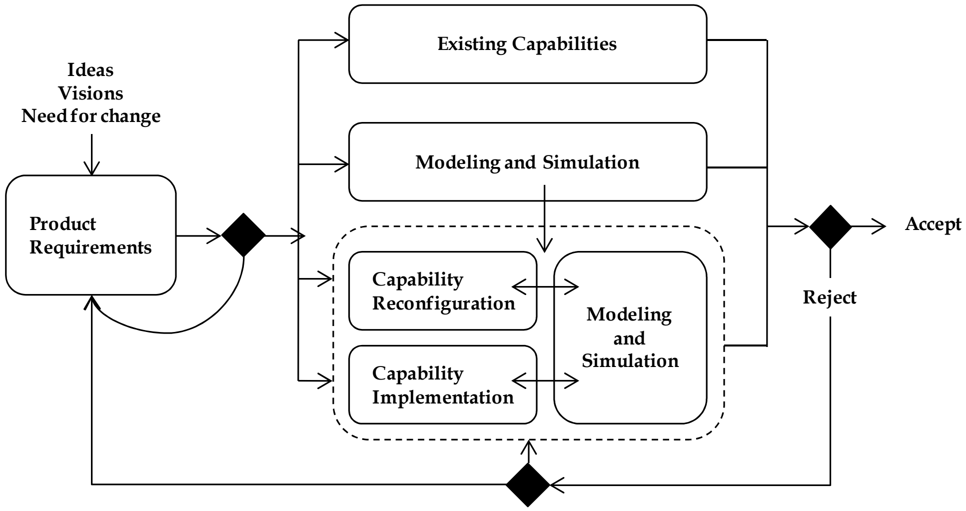


Fig. 7. Alternative outcomes of capability modelling and simulation

When it is known that the capabilities exist for all the product requirements, the efficiency of the capabilities still needs to be evaluated against factors such as cost, quality, and time. It has to be decided if the solution alternative is good enough. It can be further investigated in the capacity loop or it can be rejected and sent back to the capability loop. If all the needed capabilities exist, the capacity of the system has to be checked. The same five categories can be used in capacity evaluation. If it is known that there is enough capacity, nothing else has to be done. Modelling and simulation can be used to verify that there is enough capacity. It can also be used in capacity reconfiguration and implementation issues. Modelling and simulation of capacity has the same constraints as in the case of capabilities. The capacity for existing volume and variation still has to exist when new products are considered as an addition to existing products. In the capacity loop, the solution can be accepted or rejected, as in the capability loop. If the solution is rejected, it can be sent back to the capability loop or further back into the design requirements loop.

4.3 Manufacturing system operation

Operation of a manufacturing system can be viewed from the time dimensions of past, present, and future. The past represents what has happened i.e. it can be said to be the digital memory of the system. The time dimension of the present, what is happening now, is used to operate the current system by monitoring the state of the system and comparing it to the desired state. The future dimension makes it possible to plan future manufacturing activities ahead and to compare different changes in strategies. Figure 8 shows the connection of the time dimensions into the operation of manufacturing systems.

The past presents the data collected from the system activities when they happened. It can be used to analyse previous manufacturing activities in order to find out what happened and the reasons why it happened. In finding the root causes for phenomena, the system can learn from its past and prevent unwanted situations in the future. Rules for the autonomy of the manufacturing entities, as well as for their collaboration, can be enhanced and new rules can be created. The present here means the near future, where no major changes are planned.

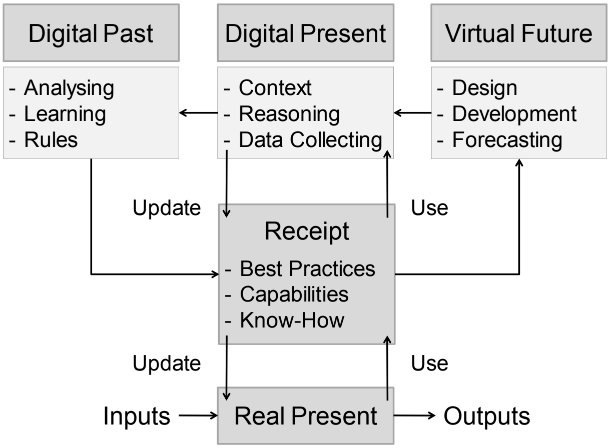


Fig. 8. Digitally co-existing past, present, and future time dimensions

It is, for example, the use of existing resources and the planning and scheduling of customer orders that have already been placed. In the present the digital and real existences co-exist. As the system operates the activities are logged, creating new history data to be analysed and to aid decision-making. The state of the real manufacturing system can be seen in the digital manufacturing system and actions can be taken with the state of the system as a starting point.

The dimension of the future relies on the information and knowledge gathered from the system previously. Future design and development decisions are syntheses of existing capabilities and requirements combined with future goals and possibilities. The viewpoint of the future can be divided into tactical decisions and visions. Tactical decisions consider the near future into which the manufacturing system is heading. Future visions are similar to tactical decisions, the difference being the time horizon.

The outcome of future visions is more obscure but there are more possibilities to be investigated. The information and knowledge from analysing the past, collecting data from the present, and forecasting the future is stored in the form of receipts. A receipt holds the capabilities of a system, constantly updating and refining the best practices in conjunction with human skills and know-how. The receipts are the basis of the operations in the real present, the only time dimension in the real world.

4.4 Continuous analysis and improvement

A manufacturing system can be seen as multiple autonomous manufacturing entities interacting and co-operating in a complex network of manufacturing activities. The activities are explained as services, which hold the information and knowledge needed to explain the manufacturing activities. It is required that the activities are known exactly, in that they are understood by all related parties.

Describing the activities as services in a digital format creates a formal way to present the services. This makes possible efficient collaboration in a digital manufacturing system between entities that can be humans, machines, or information systems. The information and knowledge is kept as the autonomous property of the manufacturing entities and the communication between the entities includes only the information that is needed to fully describe the collaboration activity.

The communication between the manufacturing entities is loosely based on service-oriented architecture (SOA), which consists of self-describing components that support the composition of distributed applications (Papazoglou & Georgakopoulos, 2003) enabling the autonomous manufacturing entities to negotiate and share their information and knowledge. The basic conceptual model of the SOA architecture consists of service providers, service requesters, and service brokers (Gottschalk, 2000). The roles of manufacturing entities in a digital manufacturing system based on SOA are briefly explained as follows:

- Service requesters are typically product entities when they are realized as order entities. The order entities call on the services they require to be manufactured.
- Service providers include the manufacturing resource entities which have the capabilities needed to provide the services that are requested.
- Service broker plays a role of an actor that contains the rules and logics of using the services. Its function is to find service providers for the requesters on the basis of criteria such as cost, quality, and time.

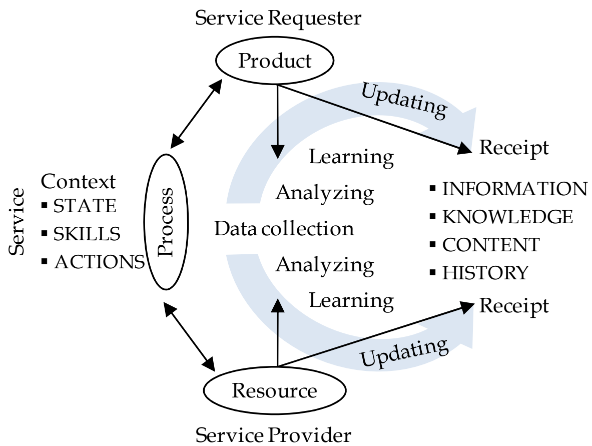


Fig. 9. An example of a service between a service provider and a service requester

Figure 9 shows an example of a service occurring in the process domain between products and resources. A service consists of two different entities, i.e. the product and resource entities having the roles of service requesters and providers. The actual service, being the manufacturing activity, is twofold, consisting of a context and receipt. The context is the environment, real or virtual, where the service takes place, whilst the receipt is the digital description of the service. The product entity requests a service, which is provided by the resource entity. The service, whether it is happening in a virtual or real environment, has a certain context that is in a certain state. The state is a basis for the actions happening during the service, and the result is based on the skills of the service provider. During the service data are collected from the process. The collected data are analyzed, forming information that is the basis for learning from the service. When something is learned, it is used to update the receipt, which will be the basis for future services.

When a certain product entity uses a service provided by a certain resource entity, the data collection, analysis, learning, and updating phases include adding the same data and information to the knowledge of both entities. The knowledge of a resource entity is updated with several product entities using the services it provides. In a similar fashion, the knowledge of a product entity consists of all the services it requests. A service can be seen as a hierarchy in which a service on the upmost level divides iteratively into multiple sub-services until the level on which the individual part features are requested. This means that an entity requesting a service gets information about the possible service provider entities, but it does not know how the service request is fulfilled. For example, a service request for the manufacturing of a product is a request on the macro level. The macro level service request is divided into multiple sub-services on the meso level and the meso-level service makes similar requests on the micro level. The upper level only needs the information about whether the service request can be fulfilled or not. The hierarchy of the services may be limited by the service requester as it may state special requirements for the service that limit the selection of possible providers. For example, a customer may require certain parts of ordered products to be manufactured in a specific manufacturing plant.

5. Academic research environment

Several of the theoretical issues discussed in this chapter have been implemented into an academic research environment of which real machinery exists in the TTE heavy laboratory, see Figure 10. The digital part of the environment has been constructed as a modular ICT architecture and the virtual part exists as simulation and calculation models. The aim of the environment is to offer a research platform that can be utilised in:

- Designing, developing and testing current and future research topics.
- Prototyping possible solutions for industrial partners in ongoing research projects.
- Utilizing it as an educational environment for university students and company personnel to introduce the latest results in the area of intelligent manufacturing.

The initial version of the environment was introduced during the Tampere Manufacturing Summit seminar, which was held in Tampere, Finland, in June 2009. Since then the environment has been discussed in scientific research papers as well as in seminar and conferences.

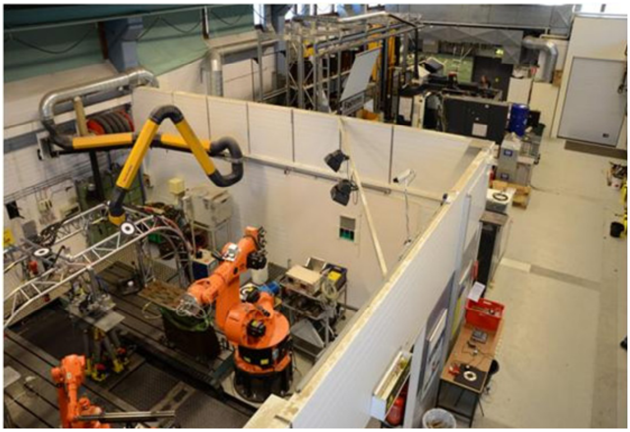


Fig. 10. The research environment in TTE heavy laboratory

5.1 General description of the research environment

The research environment consists of typical manufacturing resources and work pieces as physical entities. The resources of the research environment, offering different manufacturing capabilities are, see Figure 11:

- Machine tools (a lathe and a machining centre) for machining operations.
- Robots for material handling and robotized machining operations.
- Laser devices for e.g. machining and surface treatment.
- A punch press, existing only virtually, for the punching of sheet metal parts. The real punch press is located at a factory of an industrial project partner company.

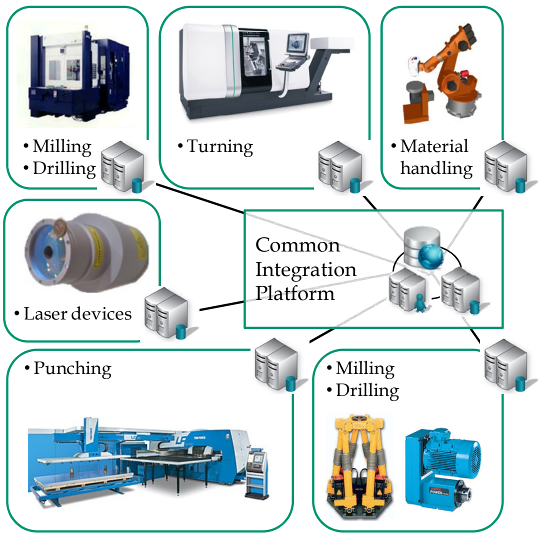


Fig. 11. The machine tools and devices of the research environment

The work pieces, which can be manufactured in the environment, are fairly simple cubical, cylindrical, and flat parts in shape. They have several parameterized features that can be varied within certain limits, e.g. dimensions (width, length, and depth), number of holes, internal corner radiuses, and sheet thickness. The main reasons for the parameterization are, firstly, that the number of different parts can be increased with the variation without having a large number of different types of parts. Secondly, the parameters can be set in a way where changing the parameters also requires capabilities of different kind i.e. different manufacturing resources are required. This gives more opportunities to compare alternative ways to manufacture the work pieces based on selected criteria, such as the cheapest or fastest way to manufacture a work piece.

5.2 Viewpoints of the environment

The research environment can be seen from the digital, virtual, and real viewpoints. Figure 12 shows the digital, virtual, and real views of the whole research environment. The environment can be viewed from three different structuring levels; the whole environment, machining and robot cells, as well as the individual machine tools and robots. The real part of the environment exists in a heavy laboratory and is divided into two main areas, one including the robots and laser devices, and the second consisting of the machine tools. The real manufacturing entities on each structuring level have their corresponding computer models and simulation environments as their virtual parts.

The information and knowledge of the environment is stored in local databases of the manufacturing entities as well as in a common Knowledge Base (KB) for the whole environment, those presenting the digital part of the environment. The actual connection is enabled by and executed via the KB, see (Lanz et al., 2008), as all communication activities use or update it. The KB is the base for the ICT-related research and development activities of the research environment. It is a system where the data of the environment can be stored and retrieved for and by different applications existing in the environment.

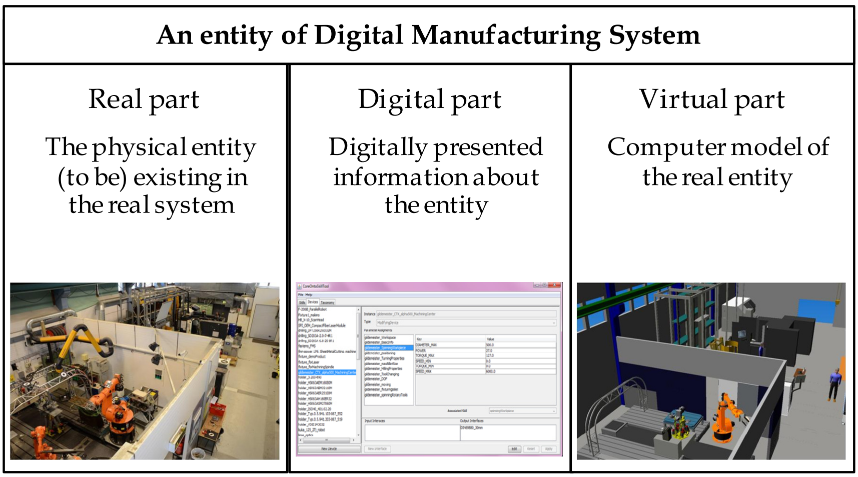


Fig. 12. Digital, virtual, and real viewpoints of the research environment

5.3 Scenarios for manufacturing tests

Figure 13 presents an overall view of the process of digital, virtual and real manufacturing tests that can be performed using the research environment. The product and manufacturing information and knowledge holds what is known about the manufacturing resources of the environment and products that have been manufactured in the environment. The manufacturing methods of the resources are described as capabilities of the research environment i.e. what is known that can be manufactured within the environment. The product requirements are described similarly including all manufacturing features of products that have been previously manufactured. When the ability to manufacture a new product will be examined, firstly a CAD model of the product is required. The CAD model will be analyzed using a feature recognition property of the research environment. For each product feature, a service request is created. The request is sent to the process planning part of the environment to compare the requirements of the new request to existing capabilities of the environment. If a suitable service exists i.e. there exists a process plan for the product feature, the result will be an existing service and no further examination is required. Otherwise, the new service request will be tested for its manufacturability.

The manufacturing tests can roughly be divided into three categories being digital, virtual and real test manufacturing. The digital test manufacturing is basically comparing a set of parameterized values of the service request to the formally described capabilities of the manufacturing resources. The process is quite rapid as it is happening in a computer and no visualization or animation is required. It is the most favourable choice if time is limited.

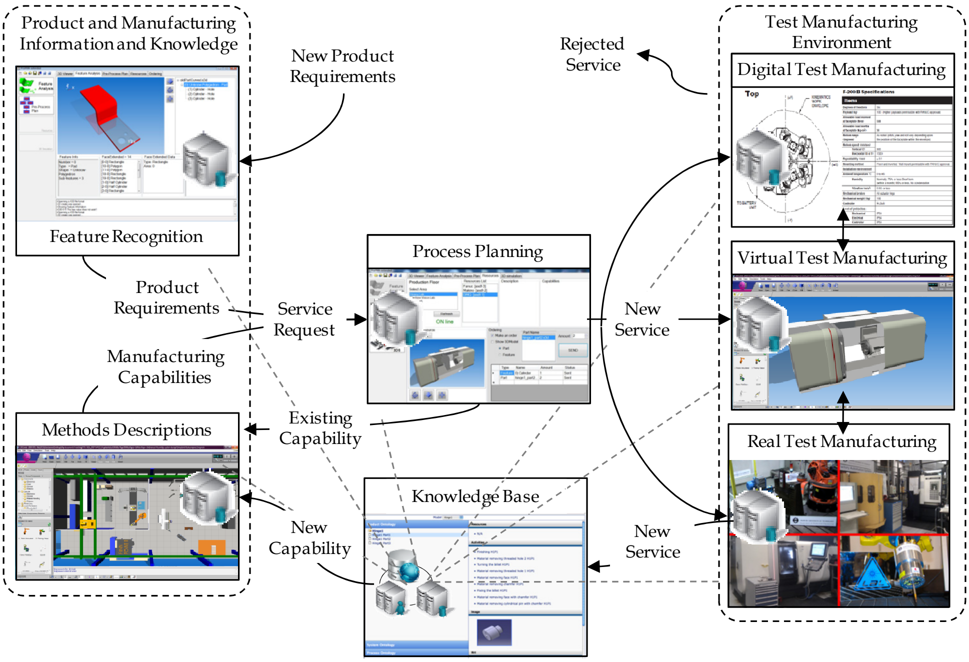


Fig. 13. An overall view of the manufacturing tests of the research environment.

The second choice would be a virtual test manufacturing i.e. typically modelling and simulation. It requires more time as human interaction is required during the process. The time required is dependent if existing simulation models can be used or new simulation models need to be constructed. In the case, where the existing simulation models cannot be used, new ones are required to be built. The creation of a new simulation model may be to reconfigure the existing virtual system to meet the new requirements, or implementing something new into the system if the system does not have all the required capabilities or manufacturing resources. In these alternatives, the test manufacturing is still carried out with computers i.e. it does not interrupt the use of the real manufacturing resources.

The real test manufacturing will be the choice if the digital or virtual manufacturing tests are not accurate enough to fully trust or understand the results gathered from the test. The real test manufacturing requires the physical resources and the time used will reduce the time for daily operations to manufacture customer orders. In some cases it is also reasonable to conduct additional tests with real manufacturing resources to reduce the risk of implementing fault processes. The responsibility of selecting, whether digital or virtual test manufacturing would be enough, is to be determined by humans, based on their skills and knowledge of the matter in hand, and has to be evaluated separately for each time a decision needs to be made. After the manufacturing tests have been conducted, the alternative is either a rejected or accepted new service. The result of rejected service could happen if the product feature cannot be manufactured within the system, or even if it could be manufactured, it is e.g. too expensive, uses too much time or does not output desired quality. In these cases, the results can be fed back to the product development to consider it the feature can be redesigned. In the case where the new service is accepted, it is added as a new capability of the environment and new process plan will be created. This will increase the known capabilities of the environment as each test manufacturing test adds new information and knowledge to the digital part of the environment, which will be available for the future test manufacturing cases.

5.4 Performance metrics

The measurements of the manufacturing environment can be divided into direct and indirect measures. The direct measurements are achieved using the sensors and measurement devices in the environment, and the metrics can be calculated immediately. Examples of the direct measurements are:

- Process quality assurance, a real time measurement using force, acceleration, and acoustic emission (AE) sensors.
- Process stability monitoring following the electricity variation of the robot servomotor caused by the cutting forces.
- Energy consumption monitoring using a Carlo Gavazzi EM21 72D energy meter.

In the case of the indirect measurements the logged data are stored in the history section of the KR. The data can be analyzed and to create the desired performance metrics. Table 1 summarizes the performance metrics from the viewpoints of manufacturing operation, production supervising, and business management.

Performance metric	Manufacturing Operator	Production Supervisor	Business Management
Cost	Continuous improvement to reduce the cost per part	Using the most cost-efficient production choices	The gain more profit and offer cheaper products to customers
Quality	To assure the manufacturing process efficiency and stability	Delivery reliability and Just-In-Time manufacturing	Improved customer satisfaction and decreased reclaims
Material consumption	To use near-net shape blank material	To reduce waste, material and energy use to meet the sustainability requirements	Meeting the requirements of legislation and expectations of the society by reducing the unwanted effects
Waste			
Energy consumption	To have real energy consumption results		
Production load and time metrics	To reduce the time per part and to update any changes in the manufacturing process times	To efficiently plan and schedule production to utilize the capacity of the system	To know how much customer orders can be placed and to give more precise delivery dates
Resource utilization			

Table 1. Different views to utilize the performance metrics

6. Conclusion

This Chapter discussed on the possibilities of digital manufacturing to support efficient activities of designing, developing and operating manufacturing systems. A structure of individual manufacturing entities and whole systems was proposed. Describing entities of a manufacturing system as independent, yet closely related existences of digital, virtual and real enables more efficient and effective manufacturing activities from early conceptual ideas to successful solutions. Even when describing the manufacturing entities independently, they are required to be closely integrated with each other and that can be done via domains of manufacturing related activities of products, resources, and business. Again, when the entities and domains are combined, the integrated fashion should also be invested separately in different structuring levels of manufacturing, yet again closely integrated between the structuring levels

By keeping the entities the same during their whole lifecycle reduces the loss of information and knowledge and enables more efficient manufacturing activities. These we discussed from several aspects i.e. a path from early ideas and needs to efficient solutions, development of manufacturing processes and flow, as well as how a system can learn from its daily operations by collecting and analysing data from the activities that can help in learning thus improving the way to do things in future.

An academic research environment was discussed on how these theoretical aspects can be implemented into a manufacturing environment. As the environment is constantly developed, some of the issues have been fully implemented while some other areas remain as a future of the environment. This is due to the fact that the current and future research topics lead the development of the environment.

7. Acknowledgment

The research presented in this paper is co-financed by Tekes (the Finnish Funding Agency for Technology and Innovation), TUT Foundation and several major companies in Finland. The authors would like to thank all the colleagues involving on the development of the academic research environment.

8. References

Ad-hoc Industrial Advisory Group. (2010). Factories of the Future PPP - Strategic Multi-annual Roadmap.

Alavi, M. & Leidner, D.E. (2001). Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research issues. *MIS Quarterly*, Vol. 25, pp. 107-136, ISSN: 0276-7783

Andersson, P.H. (2007). FMS in 2010 and beyond. *Tampere Manufacturing Summit 2007*, Tampere, Finland, June 2007

Bracht, U. & Masurat, T. (2005). The Digital Factory between vision and reality. *Computers in Industry*, Vol. 56, pp. 325-333, ISSN: 0166-3615

Chryssolouris, G., Mavrikios, D., Papakostas, N., Mourtzis, D., Michalos, G. & Georgoulias K. (2008). Digital manufacturing: history, perspectives, and outlook. *Proceedings of the Institution of Mechanical Engineers Part B: Journal of Engineering Manufacture*, Vol. 222, No. 5, pp. 451-462, ISSN (printed): 0954-4054. ISSN (electronic): 2041-2975

ElMaraghy, H.A. (2005): Flexible and reconfigurable manufacturing systems paradigms. *International Journal of Flexible Manufacturing Systems, Special Issue on Reconfigurable Manufacturing Systems*, Vol. 17, pp. 261-276, ISSN (printed): 0920-6299. ISSN (electronic): 1572-9370

Gottschalk, K. (2000). Web Services architecture overview. IBM developerWorks, Whitepaper

Gould, P. (1997). What is agility?. *Manufacturing Engineer*, Vol. 76, No. 1, pp. 28-31, ISSN

Gunasekaran, A. & Yusuf, Y.Y. (2002). Agile manufacturing: a taxonomy of strategic and technological imperatives. *International Journal of Production Research*, Vol. 40, No. 6, pp. 1357-1385, ISSN (printed): 0020-7543. ISSN (electronic): 1366-588X

Jawahir, I.S., Badurdeen, F., Goldsby, T., Iyengar, D., Gupta, A., Metta, H., Stovall, C. & Ladd. C. (2009). Assessment of Product and Process Sustainability: Towards Developing Metrics for Sustainable Manufacturing. *NIST Workshop on Sustainable Manufacturing*, Gaithersburg, MD, USA, October 2009

Jayachandran, R., Singh, R., Goodyer, J. & Popplewell, K. (2006). The design of a sustainable manufacturing system: A case study of its importance to product variety manufacturing. *Intelligent production machines and systems, 2nd I*PROMS virtual international conference*, ISBN 13: 978-0-08-045157-2, Pham, D.T., Eldukhri, E.E. & Soroka A.J. (eds.), pp.650-656

Jovane, F. (2009). The Manufuture Road towards Competitive and Sustainable HAV Manufacturing, *Tampere Manufacturing Summit*, Tampere, Finland, June 2009

Kidd, T. (1994). *Agile Manufacturing: Forging new frontiers*. Addison-Wesley Reading, 1st edition, ISBN-13: 978-0201631630, Boston, MA.

Koestler, A. (1989). *The Ghost in the Machine*. Arkana Books, ISBN-13: 978-0140191929, London, UK

Koren, Y., Heisel, U., Jovane, F., Moriwaki, T. Pritschow, G., Ulsoy, G. and Van Brussel H. (1999). Reconfigurable manufacturing systems. *CIRP Annals*, Vol. 48, No. 2, pp. 527–540, ISSN: 0007-8506

Kühn, W. (2006). Digital Factory - Integration of simulation enhancing the product and production process towards operative control and optimization. *International Journal of Simulation Modelling*, Vol. 7, No. 7, pp 27-39, ISSN: 1726-4529

Lanz, M., Kallela, T., Velez, G. & Tuokko, R. (2008). Product, Process and System Ontologies and Knowledge Base for Managing Knowledge between Different Clients. *Proceedings of the 2008 IEEE International Conference on Distributed Human-Machine Systems*, pp. 608-513, ISSN: 1094-6977

Lapinleimu, I. (2001). Ideal Factory, Theory of Factory Planning, Produceability and Ideality. Doctoral dissertation, Tampere University of technology, Finland, Publications 328

Lean Enterprise Institute. (2008). *Capsule Summaries of Key Lean Concepts*. accessed September 15, 2011, Available from: http://www.lean.org/WhoWeAre/NewsArticleDocuments/key_lean_definitions.html

Maropoulos, P.G. (2003). Digital enterprise technology-defining perspectives and research priorities. *International Journal of Computer Integrated Manufacturing*, Vol. 16, Nos. 7-8, pp. 467-478, ISSN (printed): 0951-192X. ISSN (electronic): 1362-3052

Mehrabi, M.G., Ulsoy A.G. & Koren, Y. (2000). Reconfigurable manufacturing systems: Key to future manufacturing. *Journal of Intelligent Manufacturing*, Vol. 11, pp. 403–419, ISSN (printed): 0956-5515. ISSN (electronic): 1572-8145

Nagel, P. and Dove, R. (1991). 21st century manufacturing enterprise strategy. Iacocca institute, ISBN-13: 978-0962486630, Bethlehem, PA.

Nonaka, I. & Takeuchi, H. (1995). The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. Oxford University Press, ISBN-13: 978-0195092691, Oxford, USA

Nylund, H., Koho, M. & Torvinen, S. (2010). Framework and toolset for developing and realizing competitive and sustainable production systems. *Proceedings of the 20th International Conference on Flexible Automation and Intelligent Manufacturing*, July 2010, Oakland, CA, USA, pp. 294-301.

Nylund, H. & Andersson, P.H. (2011). Framework for extended digital manufacturing systems. *International Journal of Computer Integrated Manufacturing*, Vol. 24, No. 5, pp. 446 – 456, ISSN (printed): 0951-192X. ISSN (electronic): 1362-3052

Papazoglou M.P. & Georgakopoulos, D. (2003). Service Oriented Computing. *Communications of the ACM*, Vol. 46, No. 10, pp .25-28, ISSN: 0001-0782

Sharifi, H. & Zhang, Z. (1999). A methodology for achieving agility in manufacturing organizations: An introduction. *International Journal of Production Economics*, Vol. 62, No. 1-2, pp.7-22, ISSN: 0925-5273

Souza, M.C.F., Sacco, M. & Porto A.J.V. (2006). Virtual manufacturing as a way for the factory of the future. *Journal of Intelligent Manufacturing*, Vol. 17, pp. 725-735, ISSN (printed): 0956-5515. ISSN (electronic): 1572-8145

Strauss, R.E. & Hummel, T. (1995). The new industrial engineering revisited - information technology, business process reengineering, and lean management in the self-organizing "fractal company". *Proceedings of 1995 IEEE Annual International Engineering Management Conference*, Singapore, June 1995, pp. 287-292.

Van Brussel, H., Wyns, J., Valckenaers, P., Bongaerts, L. & Peeters, P. (1998). Reference architecture for holonic manufacturing systems: PROSA. *Computers in Industry*, Vol. 37, pp. 255-274, ISSN: 0166-3615

Warnecke, H.J. (1993). The Fractal Company: A Revolution in Corporate Culture, Springer-Verlag, ISBN-13: 978-3540565376, Berlin, Germany.

WCED (1987) Our Common Future. World Commission on Environment and Development (Brundtland Commission), Oxford University Press, ISBN-13: 978-0192820808, Oxford.

Wiendahl, H.-P., ElMaraghy, H.A., Nyhuis, P., Zäh, M.F. & Wiendahl H.-H. (2007). Changeable Manufacturing - Classification, Design and Operation. *CIRP Annals*, Vol. 56, No. 2, pp.783-809, ISSN: 0007-8506

Womack, J.P., Jones, D.T. & Roos, D. (1990). The machine that changed the world, Rawson, ISBN-13: 978-0060974176

14

Modelling and Implementation of Supervisory Control Systems Using State Machines with Outputs

Moacyr Carlos Possan Junior[1,2] and André Bittencourt Leal[2]
[1]*Whirlpool Latin America*
[2]*Santa Catarina State University – UDESC*
Brazil

1. Introduction

The growth of the complexity of automated systems in industry has occurred extensively in recent years due to the production demand increase, quality improvements and flexibility to restructure the manufacturing systems in order to satisfy new procedures. Nevertheless, the evolution of control devices and their functionalities, such as processor speed, memory and network communication has advanced in parallel with the factories' requirements. Although the evolution of automation in industrial processes is a fact, there is still a scarcity of formal methods for analysis, project and implementation of control systems for Discrete Event Systems (DES) in order to reduce the development time, reduce human resources investment, and satisfy the operational requirements for certain systems in an effective way. Furthermore, the occurrence of programming bugs resulting in errors due to interruption in the process and losses due to poorly designed software is obviously unacceptable nowadays in this market that has a just-in-time mindset and is strongly focused on profits.

Usually, the projects for supervisory control systems are based on the knowledge of the system's practitioner, according to his experience in programming. The usage of formal methods is sparse, such that the reuse of documentation and source code as well as the dissemination of the knowledge generated are both impaired. Moreover, the automation of manufacturing systems has brought an increase in the complexity of control systems, so that to elaborate and implement robust and reliable control logic is not a trivial task. In order to minimize the risks due to programming errors and to permit a formal method for modelling DES, Ramadge & Wonham (1989) introduced the Supervisory Control Theory (SCT), which guarantees optimal control logic (nonblocking and minimally restrictive) for these systems.

A Discrete Event System (DES) consists of a system with discrete states that are driven by events. In other words, its state evolution depends on the occurrence of discrete asynchronous events over time (Cassandras & Lafortune, 2008). Discrete Event Systems are quite common in the industry nowadays and the events may be classified as uncontrollable and controllable. Examples of controllable events are the start and end of an operation and examples of uncontrollable events are the activation and deactivation of a presence sensor.

The Supervisory Control Theory (SCT) is already widespread in the academic environment, using the automata theory as a base to model the control systems. However, such a theory is not common in the industrial environment. Therefore the resolution of supervisory control problems has been done without the usage of a formal procedure.

The SCT allows the solution of control problems in a systematic manner. This technique guarantees that the resulting supervisor will satisfy the specifications imposed by the designer, avoiding general control issues, such as blocking. Besides, due to its heuristic nature, the SCT facilitates the code writing changes before implementation in a controller, in case there is some inclusion/exclusion of devices or changes in the system layout.

Although the SCT provides an automatic method to synthesize control systems for DES, when analysing the monolithic supervisor obtained from the SCT, it is difficult to visualize the process dynamics in an easy way as the system complexity grows. That occurs due to the large number of states and no distinction about what kind of event, uncontrollable or controllable, has priority to occur when the supervisor is in a certain state. Furthermore, the implementation of this supervisor in a controller will require considerable non-volatile memory, neither being an elegant solution nor justifying the adoption of a formal method. PLC (Programmable Logic Controller) implementation of DES supervisory control was discussed in many works, as in (Ariñez et al., 1993; Lauzon, 1995; Leduc & Wonham, 1995; Leduc, 1996; Lauzon et al. 1997; Fabian & Hellgren, 1998; Dietrich et al., 2002; Hellgren et al., 2002; Liu & Darabi, 2002; Music & Matko, 2002; Queiroz & Cury, 2002; Chandra et al., 2003; Hasdemir et al., 2004; Manesis & Akantziotis, 2005; Vieira et al., 2006, Noorbakhsh & Afzalian 2007; Hasdemir et al., 2008; Silva et al., 2008; Leal et al., 2009; Uzam et al., 2009; Moura & Guedes, 2010).

In brief, other works presented methodologies using extended automata with variables in an attempt to minimize the exponential growth of states resulting from the automata composition, such as (Chen & Lin, 2000; Yang & Gohari, 2005; Gaudin & Deussen, 2007; Skoldstam et al., 2008), amongst others.

This chapter intends to propose a formal methodology to model control systems for industrial plants through extended automata called Mealy state machines (Mealy, 1955) and subsequent implementation in Programmable Logic Controllers (PLCs) using the Ladder language. An algorithm proposed by Possan (2009), which explores the benefits of SCT to design a supervisor, is used to convert the automaton which represents the supervisor to a finite state machine with outputs. Another algorithm is then used to reduce the state machine to have a simplified structure implemented in a PLC. The simplified state machine representation is a different way of viewing a DES, resulting in a systematic way to implement the source code in a controller with reduced memory usage. The code implementation takes into consideration some common issues found in synchronous controllers, such as PLCs.

A Mealy state machine is a finite state machine with outputs, composed of an oriented graph where the nodes are called states and arcs are called transitions. It is a powerful model to represent the behavior of processes in general.

The chapter is structured as follows: Section 2 introduces the proposed methodology; Section 3 covers the monolithic approach defined by the SCT; Section 4 shows a system to be used as example to illustrate the modelling and implementation; Section 5 presents in detail the transformation algorithm to obtain a Mealy state machine; Section 6 relates to the state machine simplification procedure; Section 7 presents common issues found during supervisor implementation in synchronous controllers while Section 8 describes how to implement the simplified state machine in PLC using the Ladder language. Finally, Section 9 covers the conclusion.

2. Methodology for the control system design

Figure 1 shows an overview of the proposed methodology. It starts with the supervisor synthesis. The synthesis is done based on the SCT and a monolithic supervisor is obtained. The supervisor is then used as input for the transformation algorithm to obtain the state machine. The machine is then simplified to have a reduced number of state transitions. The simplified machine, on the other hand, represents a model to generate the code for a controller (PLC, microcontroller or any other data processing unit). This chapter is focused on the state machine implementation using the Ladder language for PLCs.

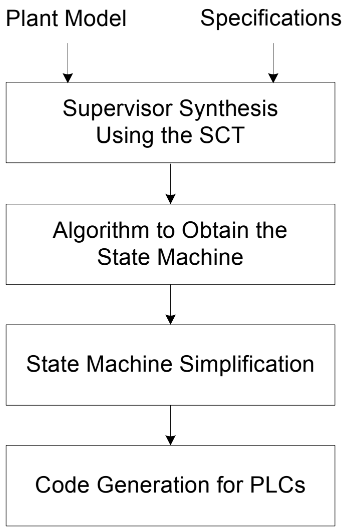


Fig. 1. Proposed methodology

The synthesis of the supervisor and the obtaining and simplification of the state machine will be described in the next three sections.

3. Supervisor synthesis using the SCT

The monolithic approach for the synthesis of an optimal supervisor (nonblocking and minimally restrictive) is based on three main steps:

a. obtain a model for the physical system (plant) to be controlled;
b. obtain a model which represents the specifications to be respected; and
c. synthesize a nonblocking and optimal control logic.

The plant model is built through the synchronous composition (Cassandras & Lafortune, 2008) of all the existing subsystem models in the system. The same procedure is done to build the specification model. The plant and the target language, obtained through the synchronous composition of the plant with the specification, will be used as input to obtain the monolithic supervisor. This procedure can be done using the computational tool named *Grail for Supervisory Control* (Reiser et al., 2006).

In the SCT, the plant is assumed to spontaneously generate events. The supervisor observes the string of events generated by the plant and might prevent the plant from generating a subset of the controllable events, thus disabling them. However, the supervisor has no means of forcing the plant to generate an event, as shown in Figure 2-a.

In practice, the modelled behavior of the plant does not correspond exactly to the real behavior due to the assumption that controllable events are not generated by the plant, as presumed by the SCT. This is because in most real systems, the events modelled as controllable correspond to commands that actually must be generated by the control system. These commands must be sent by the controller to the actuators because they would not occur spontaneously. Thus, the implementation is performed according to the structure proposed by Queiroz & Cury (2000) to keep such coherence. Figure 2-b shows the representation for real systems, where the electric signals coming from the sensors (responses from the plant) correspond to the observed (uncontrollable) events while the electric signals sent to the actuators (actions in the plant) correspond to the disabled (controllable) events.

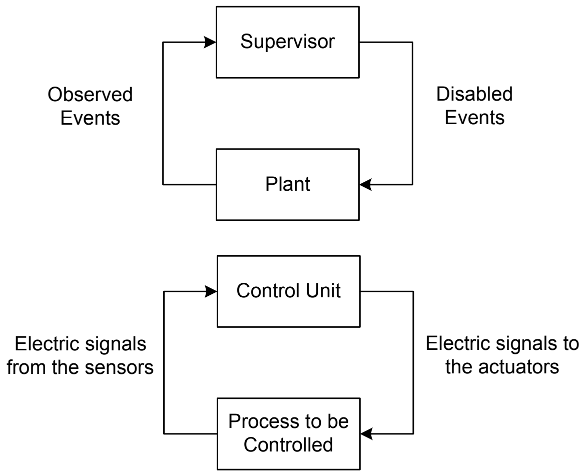


Fig. 2. Monolithic approach: (a) Ramadge-Wonham Framework, (b) Similar representation for a real system

4. A motivation example

In this section, a manufacturing system is used to demonstrate the proposed methodology. This system is composed of three apparatus and two intermediary buffers with a capacity of one which are available between the apparatus, as illustrated in Figure 3.

The apparatus are represented by *Ai*, where *i* = *1, 2, 3* and the buffers are represented by *Bj*, where *j* = *1, 2*.

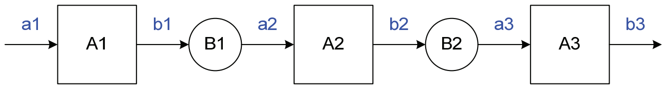


Fig. 3. Manufacturing system

The controllable events that correspond to the start of the apparatus' operation are represented by a_x, while the uncontrollable events that correspond to the end of operation are represented by b_x, where $x = 1, 2, 3$.

The plant and specifications are modelled using automata and the controllable events are represented with a dash. The behavior of each apparatus (or subsystem) can be modelled by the automaton shown in Figure 4. Notice that state *0* is double circled. This is a marked state that represents a completed task.

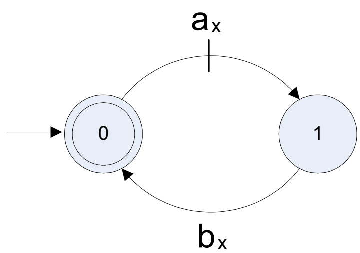


Fig. 4. Automaton for the apparatus Ax

When modelling this plant, only the apparatus were taken into consideration. The buffers were considered only in the in the control specification model.

The specifications for this system are restrictions of coordination to avoid overflow (the apparatus finishes its task but the output buffer is already full) or underflow (the apparatus starts working without any item to fetch from the input buffer). Those restrictions point out the idea that it is necessary to alternate b_1-a_2 and b_2-a_3, respectively. This means that the start of operation for an apparatus (event a_{x+1}) will only be allowed when its input buffer is loaded (event b_x), as shown in Figure 5.

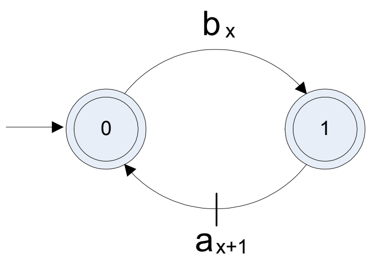


Fig. 5. Automaton for the control specifications

The synchronous composition of the plants with the specifications will result in the target language (Ramadge & Wonham, 1989). The calculation of the minimally restrictive and nonblocking supervisor is based on an iterative process which identifies and eliminates bad states in the automata that models the target language.

The monolithic supervisor found with the usage of *Grail* (Reiser et al., 2006) has 18 states and 32 state transitions.

The supervisor can be reduced for a later comparison with the state machine. This procedure is done to simplify the supervisor size and also the number of transitions (Su & Wonham, 2004). An algorithm for supervisor reduction is used to obtain fewer states and transitions than the original supervisor (Sivolella, 2005).

Figure 6 shows the automata which represents the reduced supervisor. The disabled events for each state are represented by red dashes.

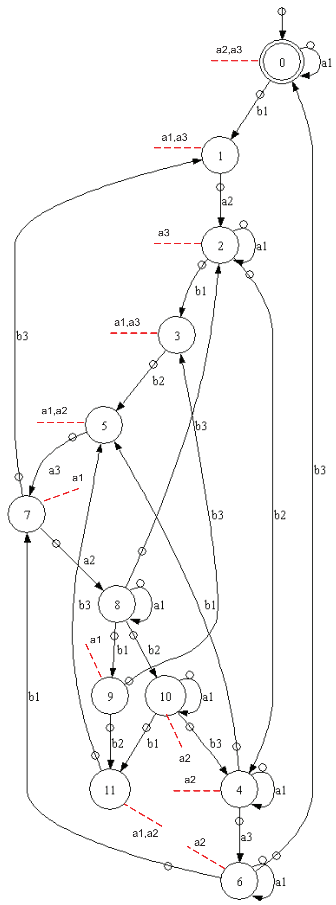


Fig. 6. Reduced supervisor and disabled events for the manufacturing system

Notice that the reduced supervisor has 12 states and 20 state transitions. The supervisor reduction algorithm creates self-loop transitions. However, these transitions are not relevant because they do not cause a state change.

Transitions among the states can occur either by controllable events (a_x) or by uncontrollable events (b_x). In case the designer intends to implement the code in the controller based on this model, he will have to decide on the kind of event to give priority. Furthermore, it is not trivial to visualize what are the active apparatus and what events modelled as uncontrollable are expected to occur at a certain state of the supervisor.

5. Algorithm to obtain the state machine with outputs

The state machine obtained with the proposed algorithm consists of a Mealy machine (Mealy, 1955). In a Mealy machine, a transition can have one or more output actions (set one or more controllable events) and any output action can be used in more than one transition. The output actions are not associated with the states, which are passive. Thus, the actions can be associated with more than one state.

A simple example of a Mealy machine is shown in Figure 7. Transitions and actions are separated by dashes. It is a machine with two states where, when in State 1, Transition 1 makes the machine to go from State 1 to State 2 and takes Action 1. When in State 2, Transition 1 plus Transition 2 makes the machine go from State 2 to State 1 and Action 2 is taken.

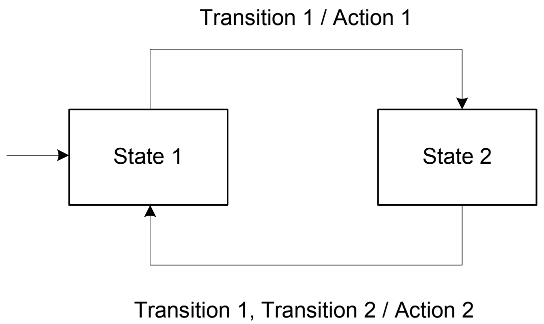


Fig. 7. Example of a Mealy machine

In this proposal, the transition between two states is performed by means of one or more uncontrollable events in the system. For each transition, an output action may be generated.

The algorithm proposed to create the state machine works in an iterative process, looping through the input data to obtain the states, transitions and actions that compose a finite state machine. As soon as all the input data is processed, the state machine is completed. The algorithm is shown in Figure 8.

Input data for the algorithm is the information about the plant, the supervisor and the list of disabled events. Output data are the states, transitions and actions which compose the state machine.

In the transformation process from the supervisor to the state machine, the output actions correspond to the controllable events in SCT while the transitions among the states correspond to the uncontrollable events.

The initialization considers the initial states of the input data. These data compose the starting point of the state machine representing the condition where the physical process has not started.

The next step is to create a states queue. The queue is required to store the states that are being obtained iteratively to be treated after the treatment of a current state has finished. The queue consists of a First In, First Out (FIFO) structure. A while loop is suggested to treat all the states available until the states queue becomes empty.

```
Read the supervisor (S) automata and the disabling map (DM)
Split the supervisor automata according to its events controllability – Su (uncontrollable) and Sc (controllable)
Create the initial state for the machine = initial state of supervisor automata
Create a list with the next states which are obtained from the current states that need to be processed
while (States List > 0) do
    Read state from the list and consider as a current state
    Create a Transitions List for this state
    for (Transitions List) do
        Read Su, MD, Sc
        if (Sc evolved) then
            Update MD e Sc
        else
            Create a new next state
        end if
        if (State was already created) then
            do nothing
        else
            Add the next state to the States
            List according to the next state
            from the supervisor automata
        end if
    end for
    Decrement States List
end while
Remove unreachable states
Save the Monolithic Mealy State Machine
```

Fig. 8. Algorithm to obtain the state machine

For each state, a transition queue is created as well. A "for loop" is suggested to treat all the valid transitions for that specific state until all the transitions available in the queue are processed.

In order to create the list of valid transitions for a certain state in the state machine, it is first necessary to divide the plant and the supervisor in two parts according to the controllability of the events and their transitions. Thus G_u and S_u will have the list of automata state transitions due to uncontrollable events. The same happens for the controllable part.

The data reading sequence begins at the uncontrollable part (G_u and S_u). After that, the algorithm evaluates the disabled events for the actual state of the supervisor. This will define if the controllable part of the system (G_c and S_c) can evolve or not.

In other words, the resulting state machine gives priority to the occurrence of uncontrollable events, waiting to receive some response from the physical system to make a decision about what controllable events to disable.

The valid transitions for a certain state are the uncontrollable events that create state evolutions from the current active states in the uncontrollable part of the plant and supervisor. The combination of more than one uncontrollable event is also considered as a transition.

After the uncontrollable part is processed, the controllable part is processed. The controllable disabled events are evaluated according to the current state the supervisor is in. While the disabled events are forbidden to occur, the remaining ones may originate the actions.

The valid actions for a certain state are the controllable events that are not disabled at the supervisor state and cause states evolution in the controllable part of the plant and the supervisor.

When an action occurs, the algorithm checks if the controllable part of the supervisor has evolved. If so, then the disabled events for the destination state are evaluated in order to verify if another action may occur. This step assures that all the actions possible to occur for the same transition are processed. It means that more than one action can occur for the same transition.

If S_c has not evolved, then a new state is created. This state is compared with the other states and if it already exists, it is ignored. Otherwise, it is added to the states queue to be treated later.

For each transition due to the uncontrollable events, G_u and S_u obtained from the SCT evolve. The same happens for the controllable part in the occurrence of an action.

This methodology is consistent with the Mealy machine approach, where the outputs (actions) depend on the current state and valid inputs (transitions).

After the procedure for creating a transition and corresponding actions is finished, the algorithm will treat the remaining transitions available in the transitions queue. When all the transitions in the queue are treated, the algorithm will evaluate the next state available in the states queue. These iterative processes are performed until the states queue is empty. This means that the finite state machine has been completed.

The state machine for the manufacturing system with the proposed algorithm is shown in Figure 9, composed of 8 states and 22 state transitions.

The states are named according to the apparatus that are operating at a certain point in time and the buffers that are full. The transitions are represented by the uncontrollable events, and the taken actions, if any, are separated from the transitions by a slash (/). The disabled events are represented by red dashes. Notice that in this model the transitions may occur due to more than one uncontrollable event, and the actions, if any, may be due to more than one controllable event.

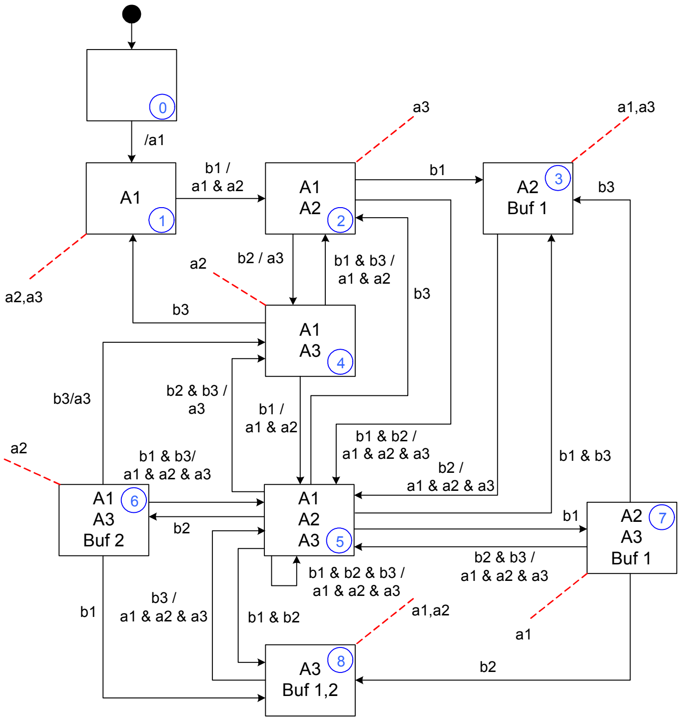


Fig. 9. Mealy state machine

6. State machine simplification

When the state machine model considers the treatment of more than one uncontrollable event, this may result in the disadvantage of an exponential growth of transitions, depending on the number of plants modelled and how many may be enabled at the same time. The number of transitions created for a certain state is on the order of $2^n - 1$, where n represents the number of uncontrollable events present in the model and possible to occur for that state. Therefore, for larger and more complex systems, the code size would be affected significantly to satisfy that condition, being a convincing reason for not using such a methodology. An alternative solution for that is to consider a reduced state machine where the state transitions are represented only by transitions that result in actions. Transitions that do not result in actions are represented by self-loops inside their current state. For example, for state 2 of the state machine represented in Figure 9, transition b_1 could be represented as a self-loop. Although the state machine evolves to state 3, there is no action taken during this transition.

Although they do not result in actions, these transitions are important to represent the plant dynamics and cannot be simply disregarded during the implementation process. The controller program must capture their occurrence and store it in some internal variable to be used in the decision-making process when other transitions occur. The algorithm presented in Figure 10 describes the process to reduce a monolithic Mealy machine.

```
1: Read the Mealy machine
2: for State = from i to n do % where n is the number of states
3:     for Transition = from i to t do % where t is the number of transitions from the current state
4:         if (Transition results in action) then
5:             Keep the transition in the reduced machine
6:         else
7:             Create a self-loop represented by dashed lines in the current state
8:         end if
9:     end for
10: end for
11: Remove unreachable states
12: Save the Reduced Monolithic Mealy State Machine
```

Fig. 10. Algorithm to reduce the state machine

It is important to emphasize that such a procedure to reduce the Mealy machine does not necessarily result in a minimal state machine.

Figure 11 shows the reduced state machine for the manufacturing system. This state machine contains only four states and seven transitions. The transitions illustrated by solid lines represent the occurrence of an action, regardless of whether the transition occurred from one state to another. The transitions illustrated by self-loops in dashed lines inside a current state represent that, although a transition has occurred, an action is not fired.

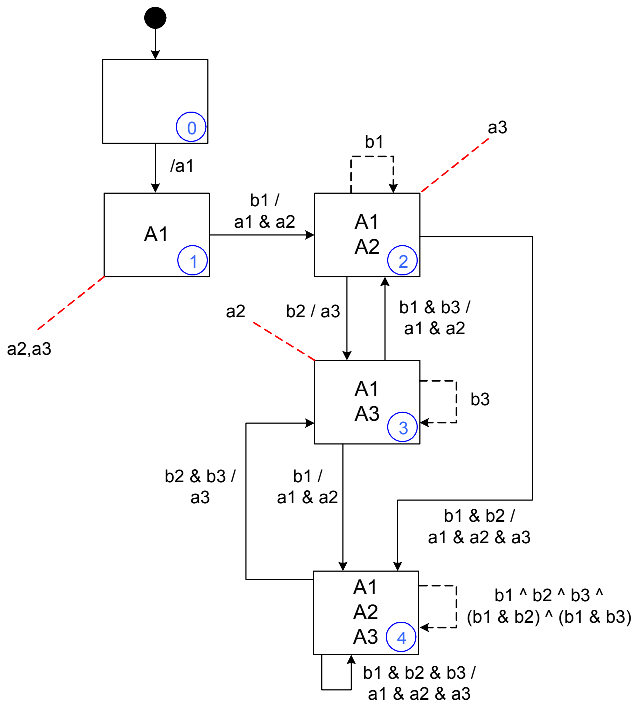


Fig. 11. Reduced Mealy state machine

Consider for instance state *2* of the reduced state machine. If transition b_2 occurs, the state machine evolves to state *3*. In the case that transition b_1 occurs, the model illustrates that situation as a self-loop represented by dashed lines, which means that the practical implementation in the controller must guarantee the proper storage of this information in some internal variable.

When event b_2 occurs with transition b_1 already enabled, transition b_1 & b_2 will be activated, so that the state machine evolves to state 4 and actions a_1, a_2 and a_3 are taken.

The disabled actions shall be represented in the corresponding states to illustrate the control actions forbidden to occur. The states which became unreachable if compared to the original state machine are eliminated by this reduced model.

Transitions due to more than one uncontrollable event and action due to more than one controllable event are still represented in this model. That is relevant information which helps the designer when he intends to implement the control system, so that the program allows several events to be executed inside the same scan cycle of the synchronous controller.

In addition, a new logic operator appears in this model. The operator "^" represents an *exclusive or* condition. In order to understand its function, consider state *4* of the reduced state machine, for example. If any of the transitions listed in the self-loop with dashed lines occur, namely, b_1 ^ b_2 ^ b_3 ^ (b_1 & b_2) ^ (b_1 & b_3), none of the actions will be taken and those transitions will remain enabled. Actions will be taken only when the transitions b_2 & b_3 or b_1 & b_2 & b_3 become valid. The operator "^" appears only for the states with a self-loop represented by dashed lines.

7. Issues with Implementing supervisors in synchronous controllers

According to (Fabian & Hellgren, 1998), "the supervisor implementation is basically a matter of making the Programmable Logic Controller (PLC) behave as a state machine". However, that is not a simple task. Certain issues appear during the implementation process in synchronous controllers, such as PLCs and computers. Those problems exist regardless of the model used to represent the supervisors, by means of automata, Petri nets or colored Petri nets (Basile & Chiacchio, 2007). Such problems are explored in the structure presented here and a solution is presented where possible, in accordance with the Ladder language definition described by the IEC-61131-3 (1993) standard.

7.1 Causality

The SCT considers that all events are generated spontaneously by the plant and that the supervisor tracks the sequence of events generated by the plant and also acts to disable the controllable events in order to avoid any infringement of the control specifications. However, in most practical applications, the controllable events are not generated spontaneously by the physical plant, but only the feedback due to sent commands. So, the question "who generates what?" must be answered (Fabian & Hellgren, 1998), or in other words, who is responsible for the generation of certain kind of events, the supervisor or the plant? (Vieira, 2007).

The supervision scheme proposed by the SCT was already illustrated in Figure 2-a. In this structure, the plant is the responsible for events generation, both controllable and uncontrollable, while the supervisor is just an observer and disables a set of controllable events to satisfy control requirements.

However, in most practical applications, the events modelled as controllable correspond to commands that, indeed, shall be generated by the PLC and sent to the actuators, because they do not occur spontaneously. The plant generates only the uncontrollable events, as a feedback to the stimulus sent by the PLC through the controllable events. Figure 2-b describes the control structure usually employed in practice (Malik, 2002).

In this chapter, the causality problem is solved by representing the model using state machines with outputs instead of automata. Therefore, when this transformation is performed, the SCT changes to a model with inputs and outputs, as suggested by Malik (2002). The signals originated by the sensors correspond to the uncontrollable events while the signals generated by the actuators correspond to the controllable events.

7.2 Event detection

The synchronous nature of PLCs may create issues during the detection of uncontrollable events as described below.

7.2.1 Signals and events

Some implementation problems are due to the lack of an easy way to translate supervisors based on discrete event systems, which are symbolic, asynchronous and occur in discrete instants of time in a synchronous universe and are based on signals such as those from the PLC (Fabian & Hellgren, 1998). In order to avoid a discrepancy between the theory and practice, a signal cannot generate more than one uncontrollable event during an operational cycle of the PLC (Basile & Chiacchio, 2007).

7.2.2 Avalanche effect

The PLC signals may assume boolean values and are sampled periodically. Thus, in order to implement supervisors according to the SCT in PLCs, the events are associated with changes in the PLC input signals, which may cause what is called the *avalanche effect*. This effect occurs when a value change in an input signal is registered as an uncontrollable event and makes the software jump to an arbitrary number of states during the same scan cycle of the PLC. This may occur specifically if a certain uncontrollable event is used to trigger several state transitions in a list, creating behavior similar to an avalanche.

Figure 12 shows an example of the avalanche effect for a conventional supervisor implementation in PLC. The supervisor transitions from state *0* to state *2* with the occurrence of event b_1. It should transition from state *1* to state *2* only with the occurrence of a new event b_1, or transition from state *1* to state *3* with the occurrence of an event b_2. However, the implementation proposed on the right side of Figure 12 does not restrict that event b_1 permits the transition from state *1* to state *2* in the same scan cycle of the PLC. So, the avalanche effect introduces a transition directly from state *0* to state *2*, so that, even if event b_2 happened, it would not have any effect on the controller dynamics. This is clearly unsatisfactory.

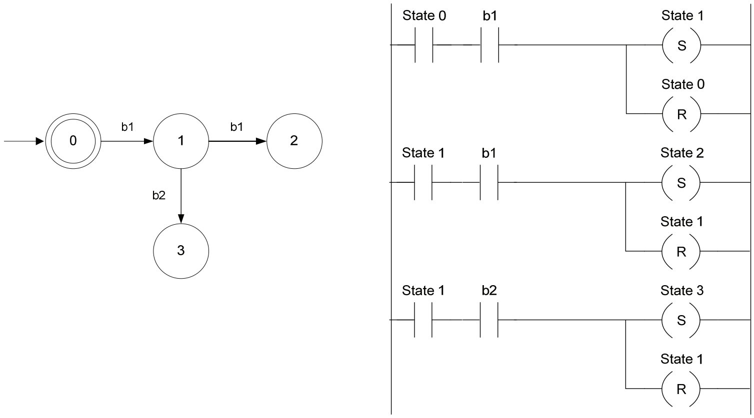


Fig. 12. Example of the Avalanche Effect

In order to solve this problem, Vieira (2007) proposes a procedure where a variable is associated with all the transitions available for the supervisor. That variable is activated every time a new state transition occurs, being deactivated at the end of the section of source code where the supervisor structure is implemented. Every state transition includes the negation of this variable as a condition. Such a solution solves the problem, but has the disadvantage of allowing the occurrence of only one uncontrollable event inside the same PLC scan cycle.

The solution proposed in this chapter consists of the deactivation of the variable corresponding to the event which occurred as soon as a state transition due to that event happens. In this way, the supervisor deactivates that specific event which occurred, allowing transitions due to other events available in the model still to happen. During a scan cycle, if events related to asynchronous specifications occur, they will be treated in this same cycle. Such a proposal solves the avalanche effect problem and permits several uncontrollable events to be treated inside the same PLC scan cycle. Moreover, the constraint that more than one event is treated in the same scan cycle results in another problem called *loss of information* (Vieira, 2007). Such a problem occurs in situations where several events happen and only one is treated per scan cycle. Then, the information related to the occurrence of the remaining events is lost. That issue does not occur in the methodology proposed here due to the possibility of treating all other events which occurred in the same scan cycle. This is possible because there is not just one general variable which is activated when a state transition occurs, but a specific treatment for each event, where the other events remain enabled to fire transitions.

7.2.3 Inability to recognize the order of events

This simultaneity problem was covered first by Fabian & Hellgren (1998) considering the implementation of supervisors in PLCs. When two or more input signals change their values between two input readings, those changes will be stored as a simultaneous change of uncontrollable events, so that it is not possible to identify which one occurred first, because the PLC performs a synchronous reading of its inputs.

It is noticeable that signal changes may be simultaneous or not, but the events will always be stored as simultaneous during the reading. That problem is called *simultaneity* (Fabian & Hellgren, 1998). Thus, in order to avoid implementation problems, the supervisor shall have a control action which does not depend on the different interleaving of the uncontrollable events. Fabian & Hellgren (1998) define such a property as *interleave insensitivity* of uncontrollable events. These authors proposed an algorithm to detect if a supervisor is interleaving insensitive. If not, the order in which the uncontrollable events occur cannot be recognized if the controller utilized is synchronous.

8. Code generation for PLCs

The scan cycle of a program in the PLC follows the sequence: input read, control logic execution and output write. That synchronous behavior of the PLC forces the outputs to be updated only at the end of the scan cycle. Due to that, the activation of the actuators requires a specific treatment. Looking at the state machine's structure, it may happen that as soon as it finishes the operation, an apparatus may be requested to start a new operation cycle inside the same scan cycle in which the previous operation had just been finished. If this happens during a scan cycle, that variable would assume a low logic level (end of operation) and return to a high logic level (start of operation). However, as the PLC writes the values stored in its internal memory to the physical outputs only when its execution cycle is finished, it does not recognize the process of operation end/start, so that its physical output will be kept active all the time during the scan cycle. In order to avoid that, variables are added to represent the evolving of the plants and guarantee the proper synchronism during the system dynamics. Those variables are called *Plant i*, with *i* varying from *1* to *n*, where *n* is the number of plants modelled in the global system. That variable is enabled every time an apparatus finishes its operation. This procedure guarantees that the apparatus is not requested to start operating again inside the same scan cycle when its operation end is detected. It will be turned on again only in the next scan cycle.

As in the state machine model, for the implementation, the variables b_i represent the uncontrollable events (transitions) while variables a_i (actions) represent the controllable events.

The Ladder code can be split into five blocks, called by a main organizational block in the following order: initialization, inputs, transitions/actions, disabling and outputs. They implement the reduced state machine shown in Figure 11.

8.1 Initialization

The initialization starts the state machine and puts it into its initial state, and enables the controllable event a_1 in order to start the process, as shown in Figure 13. Other variables, such as the ones responsible for uncontrollable events *bi* and the evolution of plants *Plant i*, are disabled.

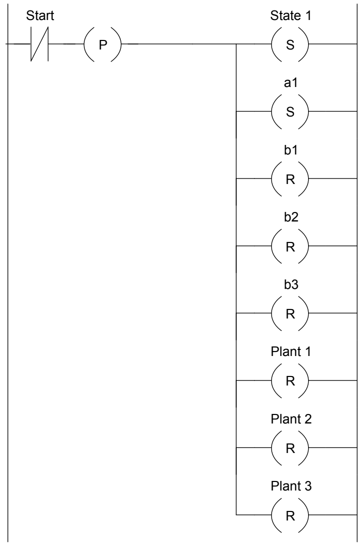


Fig. 13. Initialization

8.2 Inputs

The variables responsible for transitions b_i will be activated in the controller only when an edge rise occurs (from logic level 0 to 1) in the corresponding inputs. Therefore, a pulse detector is required for each input in order to signal the corresponding event, as shown in Figure 14. The variables related to the uncontrollable events are activated in this block. During the program execution, if they result in some action, they will be disabled. If they result only in self-loops where no action is taken, as represented by dashed lines in the state machine, these variables remain active until the moment that some transition which results in an action occurs later.

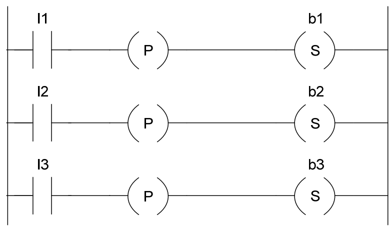


Fig. 14. Inputs

8.3 Transitions/actions

The requirement for a transition to occur is that the state machine must be in a certain state and one or more uncontrollable events that result in some action happens. If these requirements are satisfied, then the next state is activated and the current one is deactivated. The uncontrollable events responsible for the transition are deactivated to avoid occurrence of the avalanche effect (Fabian & Hellgren, 1998). When an uncontrollable event happens for a certain plant, the variable corresponding to that plant, *Plant i*, is activated in order to avoid that the action corresponding to the start of the operation for that apparatus occurs during the same scan cycle, being treated only in the next scan cycle. This is due to the specific treatment required for the actions, as described previously. The actions, if any, will be activated to allow the corresponding plant to evolve in the same scan cycle or only in the next scan cycle if forbidden to evolve in the current cycle. Figure 15 shows this block for the manufacturing system. Verify the representation of the reduced state machine shown in Figure 11 in order to compare the theoretical model with the practical one. Consider for instance that the state machine is in state *3*. Here, two transitions are possible: due to both events b_1 and b_3 or only due to event b_1. The transition b_1 & b_3 (due to events b_1 and b_3) must always appear first in the Ladder diagram due to this problem of simultaneity (Fabian & Hellgren, 1998). This is because in the case that events b_1 and b_3 occur and the transition only due to event b_1 is implemented first, the latter will be executed in the program and the variable b_1 will be deactivated. In this way, the transition due to events b_1 and b_3 will never happen in practice.

Consider now that the state machine is in state 4 and the transition b_1 & b_2 & b_3 occurs. The state machine remains in the same state and, as all the apparatus finish their operation, no action will be taken in the current scan cycle. However, as the actions a_1, a_2 and a_3 are activated, in the next scan cycle all the apparatus will be turned on again. Thus, such a methodology guarantees that it is necessary to wait only one scan cycle for the enabled actions to be taken.

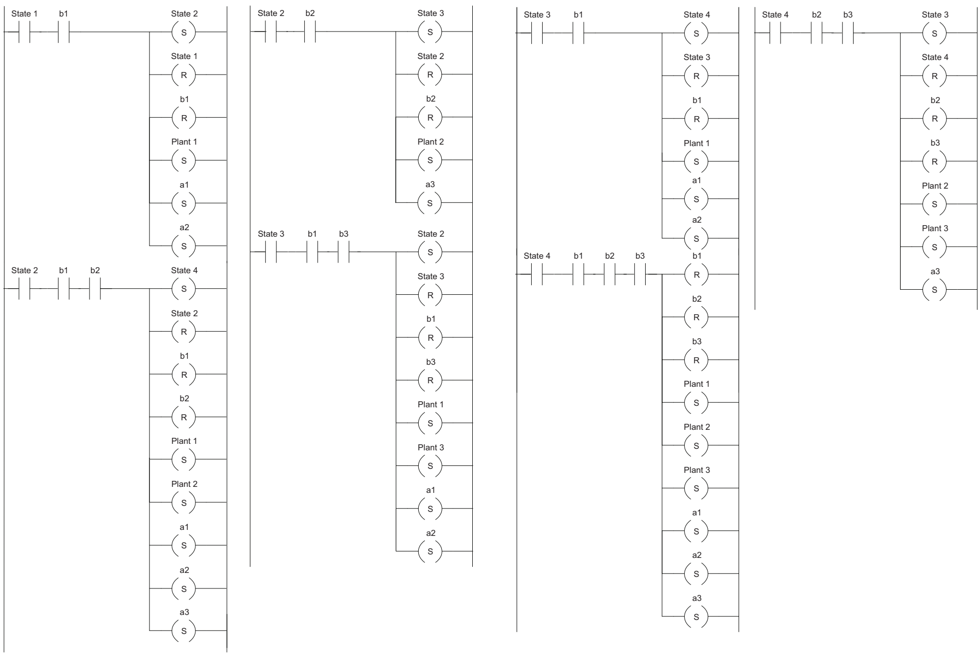


Fig. 15. Transitions/Actions

8.4 Disabling

This block is responsible for disabling the control actions in the machine states. This means that if the state machine reaches a certain state, the actions related to this state are forbidden to occur, being disabled. Figure 16 shows the disabling block for the manufacturing system. The first two rungs on the Ladder code represent the states where the corresponding control actions are disabled. Furthermore, the disabling may occur if some uncontrollable event is signalled but it does not allow an action to be taken (represented by a self-loop with dashed lines in the reduced state machine), so that the start of operation in the corresponding plant is forbidden. In the Ladder diagram, it is enough to disable the action related to a certain *Plant i* when the uncontrollable event *bi* occurs. In other words, when b_1 occurs, it is enough to disable *Plant 1*. When b_2 occurs, it is enough to disable *Plant 2*, and so on. This is because although the event *bi* occurred, it did not effectively generate a transition in the state machine, and therefore, the plant was not disabled, as shown in the remaining rungs of the Ladder diagram.

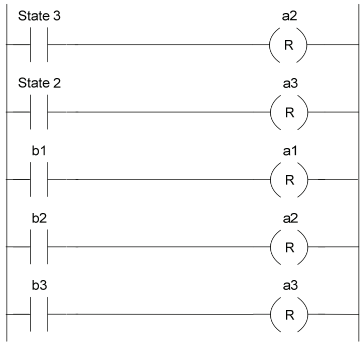


Fig. 16. Disabling

8.5 Outputs

A physical output is activated only if the action related to a controllable event is taken and its corresponding variable, *Plant i*, is not enabled, as shown in Figure 17. If such situation occurs, the coil Q_i which represents the physical output of the PLC will be energized. Yet, at the end of the program, the variables *Plant i* will be disabled in order to return to the initial condition before starting a new scan cycle.

Besides being a simplified implementation model, this solution has the advantage of not restricting more than one uncontrollable event to be treated in the same PLC scan cycle.

In order to have a better understanding of the logic implementation of this supervisory control, consider first that the start of operation is given by the action a_1. After a few PLC scan cycles, the apparatus A1 finishes its operation and the transition due to the uncontrollable event b_1 is generated. According to the reduced state machine, the actions a_2 and a_1 are enabled to occur and the state machine transitions to *state 2*. However, as *Plant 1* finished its operation, only the physical output Q_2 is activated in the same scan cycle. The physical output Q_1 will be activated only in the next scan cycle. Next, the state machine will behave differently depending on which apparatus finishes its operation first. If apparatus

A2 finishes its operation, then transition b_2 will be generated, action a_3 will be taken and the state machine transitions to *state 3*. If apparatus A1 finishes its operation, then transition b_1 will be generated resulting in a self-loop with dashed lines inside *state 2*, because no action is taken. If the PLC identifies the changes in the input signals corresponding to the end of operation in both apparatus A1 and A2 (transitions b_1 and b_2, respectively), then actions a_1, a_2 and a_3 will be taken and the state machine transitions to *state 4*. Similarly, the remaining transitions and actions follow the same dynamics, as illustrated in the reduced state machine model presented in Figure 11.

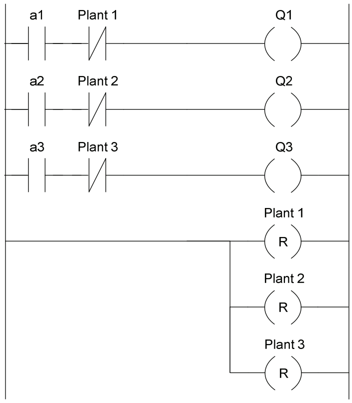


Fig. 17. Outputs

9. Conclusion

The proposed methodology presents a new model to represent supervisors for Discrete Event Systems (DES). The Supervisory Control Theory (SCT) is a convenient methodology to obtain supervisors from simpler models, where automata are useful to represent plants and control specifications so that, starting from simpler models, supervisors may be represented by state machines. In general, the approach based on state machines consists of a model rich in information. It is appropriate to emphasize the importance of the modelling process for the supervisors, because if some control specification changes or a new subsystem is added to the plant, it is necessary to remake the synthesis for the state machine.

The proposed algorithm to transform the automata which represents a supervisor in a state machine allows a reduction in the number of states in the model. That reduced approach can be used as a reference to implement the control system in a data processing unit. It is necessary only to take into account the implementation aspects related to the controller used, based on the constructive aspects of its hardware. In this chapter, solutions were described to avoid the problems usually found when implementing supervisors in synchronous controllers, using a PLC as a target with the source code implemented in the Ladder language.

The example of a manufacturing system demonstrates some aspects related to the optimization in the code size generated resulting in an economy in the non-volatile memory usage and the possibility of treating several events inside the same scan cycle of the controller. For large systems, this approach results in an improvement of the temporal dynamics of the control when several input signal changes and several actions shall be taken in the same scan cycle, ensuring synchronism and minimizing problems due to communication delays.

10. Acknowledgment

The authors thank the financial support provided by CNPq and FAPESC, and also thank Whirlpool Latin America and the UDESC for their support in the realization of this work.

11. References

Ariñez, J.F.; Benhabib, B.; Smith, K.C. & Brandin, B.A. (1993). Design of a PLC-Based Supervisory-Control System for a Manufacturing Workcell, *Proceedings of the Canadian High Technology Show and Conference*, Toronto, 1993.

Basile, F. & Chiacchio, P. (2007). On the Implementation of Supervised Control of Discrete Event Systems. *In: IEEE International Transactions on Control Systems Technology,* Vol. 15, No. 4, pp. 725 – 739, ISSN: 1063-6536

Cassandras, C.G. & Lafortune, S. (2008). *Introduction to Discrete Event Systems.* 2nd Ed. Springer Science, ISBN-13:978-0-387-33332-8, New York, USA.

Chandra, V.; Huang, Z. & Kumar, R. (2003). Automated Control Synthesis for an Assembly Line Using Discrete Event System Control Theory. *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, Vol. 33, No. 2, (May 2003), pp. 284-289, ISSN 1094-6977.

Chen, Y. L. & Lin, F. (2000). Modeling of Discrete Event Systems using Finite State Machines with Parameters. *In: IEEE International Conference of Control Applications,* Anchorage, Alaska, USA.

Fabian, M. & Hellgren, A. (1998). A PLC-based implementation of supervisory control for discrete systems. *In: Proceedings of the 37th IEEE Conference on Decision and Control,* Vol. 3, pp. 3305-3310.

Gaudin, B. & Deussen, P.H. (2007). Supervisory Control on Concurrent Discrete Event Systems with Variables. *In: Proceedings of the IEEE American Control Conference,* pp. 4274-4279, ISSN: 0743-1619, New York, NY, USA.

Harel, D. (1987). Statecharts: A Visual Formalism for Complex Systems. *Science of Computer Programming,* Vol. 8, No. 3 (June 1987), pp. 231-274.

Hasdemir T., Kurtulan, S., & Gören, L. (2004). Implementation of local modular supervisory control for a pneumatic system using PLC. In: *Proceedings of the 7th International Workshop on Discrete Event Systems (WODES'04)*, pp. 27-31, Reims, France.

Hasdemir, T.; Kurtulan, S.; & Gören, L. (2008). An Implementation Methodology for Supervisory Control Theory. *International Journal of Advanced Manufacturing Technology*, Vol. 36, No. 3, (March 2008), pp. 373-385. ISSN 0268-3768.

IEC-61131-3 (1993). International Electrotechnical Commission. Programmable Logic Controllers – Part 3: Programming Languages.

Lauzon, S. C. (1995). *An implementation methodology for the supervisory control of flexible-manufacturing workcells*, M.A. Sc. Thesis, Mechanical Engineering Dept. University of Toronto, Canada.

Lauzon, S. C.; Mills, J. K.; & Benhabib, B. (1997). An Implementation Methodology for the Supervisory Control of Flexible Manufacturing Workcells, *Journal of Manufacturing Systems*, Vol. 16, No. 2, pp. 91-101.

Leal, A. B.; Cruz, D.L.L.; Hounsell, M.S. (2009). Supervisory Control Implementation into Programmable Logic Controllers. In: *Proc. of the 14th IEEE International Conference on Emerging Technologies and Factory Automation*, pp. 899-905, Mallorca, 2009.

Leduc, R. J. (1996). *PLC Implementation of a DES supervisor for a manufacturing testbeb: an implementation perspective*, M.A.Sc. Thesis, Dept. of Electrical and Computer Engineering, Univ. of Toronto, Canada.

Leduc, R. J. & Wonham W. M. (1995). PLC implementation of a DES supervisor for a manufacturing testbed. In: *Proceedings of 33rd Annual Allerton Conference on Communication, Control and Computing*, pp. 519-528, University of Illinois.

Liu, J. & Darabi, H. (2002). Ladder logic implementation of Ramadge-Wonham supervisory controller, *Proceedings of the 6th International Workshop on Discrete Event System (WODES'02)*, pp. 383-389. ISBN: 0-7695-1683-1. Zaragoza, Spain, October 2002.

Malik, P. (2002). Generating Controllers from Discrete-Event Models. *In : Proceedings of the Modelling and Verifying Parallel Processes Movep 02*. pp. 337-342, Nantes, France.

Manesis, S. & Akantziotis, K. (2005). Automated synthesis of ladder automation circuits based on state-diagrams. *Advances in Engineering Software*, 36, pp .225–233.

Mealy, G.H. (1955). A Method to Synthesizing Sequential Circuits. *Bell Systems Technical Journal*. pp. 1045-1079.

Moura, R. & Guedes, L.A. (2010). Control and Plant Modeling for Manufacturing Systems using Basic Statecharts. *In: Programmable Logic Controller*, Guedes, L.A. (Ed), pp. 33-50, INTECH, ISBN 978-953-7619-63-3, Croatia.

Music, G. & Matko, D. (2002). Discrete event control theory applied to PLC Programming, *Automatica*, Vol 43, 1-2, pp. 21-28.

Noorbakhsh, M. & Afzalian, A. (2007). Implementation of supervisory control of DES using PLC. In: *Proc. 15th Iranian Conference on Electrical Engineering (ICEE)*, Tehran, Iran.

Possan, M.C.P. (2009). *Modelling and Implementation of Supervisory Control Systems Based on State Machines with Outputs* (in Portuguese). Thesis (Master Degree), Santa Catarina State University, Joinville, Brazil.

Queiroz, M.H. & Cury, J.E.R. (2000). Modular supervisory control of large scale discrete-event systems. *Discrete Event Systems: Analysis and Control.* Kluwer Academic Publishers (*Proc. WODES 2000),* Ghent, Belgium, pp. 1-6).

Queiroz, M. H. de & Cury, J. E. R. (2002). Synthesis and implementation of local modular supervisory control for a manufacturing cell, *Proc. of the 6th Int. Workshop on Discrete Event Systems WODES*, pp. 1-6. ISBN: 0-7695-1683-1. Zaragoza, Spain, October 2002.

Ramadge, P.J. & Wonham, W.M. (1989). The control of discrete event systems. *In: Proceedings of IEEE, Special Issue on Discrete Event Dynamics Systems*, Vol. 77, No. 1, pp. 81-98.

Reiser, C.; Da Cunha, A.E.C. & Cury, J.E.R. (2006). The Environment Grail for Supervisory Control of Discrete Event Systems. *In: Proceedings of the 8th International Workshop on Discrete Event Systems,* Michigan, USA, pp. 1-6.

Silva, D. B.; Santos, E. A. P.; Vieira, A. D. & Paula, M. A. B. (2008). Application of the supervisory control theory in the project of a robot-centered, variable routed system controller, *Proceedings of the 13th IEEE International Conference on Emerging Technologies and Factory Automation – ETFA'08*, pp. 1-6.

Sivolella, L. F. (2005). *Contributions for Supervisors Reduction for Discrete Event Systems* (in Portuguese). Thesis (Master Degree), IME, Rio de Janeiro, Brazil.

Skoldstam, M. & Akesson, K. & Fabian, M. (2008). Supervisor Control Applied to Automata Extended with Variables - Revised. Technical Report. Chalmers University of Technology, Gothenburg, Sweden.

Su, R. & Wonham, W. M. (2004). Supervisor reduction for discrete-event systems, *Discrete Event Dynamic Systems*, Vol. 14, No. 1, pp. 31-53.

Uzam, M.; Gelen, G.; & Dalci, R. (2009). A new approach for the ladder logic implementation of Ramadge-Wonham supervisors, *Proc. of the 22nd Int. Symp. on Information., Commun. and Automation Technologies*, pp. 1-7. ISBN 978-1-4244-4220-1, Bosnia 2009.

Vieira, A.D. (2007). *Method to Implement the Control of Discrete Event Systems with the Application of Supervisory Control Theory* (in Portuguese). PhD. Thesis, UFSC, Florianópolis, Brazil.

Vieira, A. D.; Cury, J. E. R. & Queiroz, M. (2006). A Model for PLC Implementation of Supervisory Control of Discrete Event Systems. *Proc. of the IEEE Conf. on Emerging Technologies and Factory Automation*, pp. 225–232. Czech Republic, September 2006.

Yang, Y. & Gohari, P. (2005). Embedded Supervisory Control of Discrete-Event Systems. *In: IEEE International Conference on Automation Science and Engineering,* pp. 410-415, Edmonton, Canada.

15

Integrated Process Planning and Scheduling and Multimode Resource Constrained Project Scheduling: Ship Block Assembly Application

Raad Yahya Qassim
Department of Ocean and Naval Engineering, COPPE,
Federal University of Rio de Janeiro
Brazil

1. Introduction

Planning and scheduling are two major tasks in manufacturing system management, which have a direct bearing on the competitive position of enterprises inserted in diverse manufacturing fields, such as the chemical, aerospace, semiconductor, and shipbuilding industries. Inadequate planning and scheduling are considered to be a major cause of the gap between desired and actual enterprise manufacturing performance in terms of inventory level, throughput, manufacturing cost, and facility location. In order to bridge this gap, not only a feasible but also an optimal production network and supply chain schedule is required, which satisfies all types of constraints within the manufacturing – marketing environment, with a view to achieving what has become known as enterprise-wide optimisation (EWO). Until the start of the last decade, the focus of most multiple facility production firms operating at multiple sites has been on operations optimisation at the single facility level. Solutions obtained at this level are suboptimal at the multiple facility level. In order to obtain an optimal solution at the latter level, the scope of optimisation must be enlarged to model the multiple facility supply chain of the enterprise as a whole, including the interaction with suppliers and customers; see for example (Laínez *et al*., 2010), Munõz et al. , (2011), Stray et al., (2006), and Monostori *et al*., (2010).

Within EWO, there exists a wide spectrum of optimisation problems whose nature depends on the type of manufacturing environment that is under consideration. The concern in this paper is focused on two such problems:

1. integrated process planning and scheduling problem (IPPSP);
2. multimode resource-constrained project scheduling problem (MRCPSP).

In view of their important role in manufacturing management practice, this pair of problems has attracted significant interest in the academic literature. In the IPPSP, the two functions of process planning and production scheduling, which have usually been solved sequentially, are integrated and solved simultaneously, with a view to providing greater flexibility at the factory floor level. In the MRCPSP, the schedule of a project is obtained allowing each activity one or more modes of execution, whereby with each mode a time duration, execution cost, and resource consumption level are associated. For recent surveys of the IPPSP and MRCPSP, the reader may refer to (Shen *et al*., 2006) and (Hartmann & Briskorn, 2010), respectively. The IPPSP arises normally at the shop floor level in batch process and discrete part manufacturing environments. The MRCPSP occurs in project – oriented production environments, such as those in shipbuilding, aerospace, and highway construction; see (Martínez *et al*., 1997 and Martínez & Pérez, 1998). The general objective of this paper is to relate these two hitherto separately considered problems in manufacturing system management, in order to open vistas for the formulation and solution of new problem variants, which may benefit from the employment of methods and techniques that have been already developed separately for the IPPSP and MRCPSP. A specific application is provided by the development of a mathematical programming model for ship block assembly, which is one of the major final stages in shipbuilding; see for example (Yu-guang *et al*., 2011).

1.1 Integrated process planning and production scheduling problem

In discrete manufacturing system management, two major tasks are process planning and production scheduling. In process planning, the question of how an item will be manufactured is answered. This is done by the determination of the sequence of operations that are necessary to produce the item under consideration. An item may be a part or an assembly of several parts. In general, an item may be manufactured in number of ways. This leads to the existence of a multiplicity of process plans. On the other hand, the question that has to be answered in production scheduling is the following: when is a job be dispatched to the shop floor, and what amount of each resource is allocated for its manufacture under prevailing conditions at a specific instant of time?

In general, process planning consists of the determination of operations and parameters that are required to convert raw materials or intermediate items into a finished item, such as a part or an assembly. The process planning task comprises the interpretation of design data, selection and sequencing of operations to manufacture the item, selection of machines, tools, and other resources, along with the corresponding quantitative data such as machine speeds and resource amounts. It is common practice that process planning is carried out in two stages: preliminary and detailed, whereby feasible process plans and optimal process plans are generated, respectively. Specific aspects of process planning vary with the application environment, such as machining, welding, and assembly; see, for example (Kong *et al*., 2011).

Production scheduling is normally carried out for several time horizon spans and corresponding levels of detail. This approach is known as hierarchical production planning (HPP); see, for example, (Bang & Kim 2010). At the shortest time horizon span with corresponding most extensive level of detail, stands the shop floor scheduling problem, which is the focus of this paper. Here, one has a set of jobs, each of which possesses its own set of process plans, along with available resource amounts at the shop floor level. One requires the sequencing of jobs in time and the corresponding allocation of resources to each job; see, for example, (Li, *et al*., 2010).

In practice, the tasks of process planning and production scheduling are carried out sequentially, whereby for each item a set of alternative process plans are first determined. The generation of each process plan of an item is based on the assumption of the unlimited availability of the set of resources that are necessary for the execution of the plan on the shop floor. For each operation within a process plan, an execution time and an amount of each resource that is required for its execution, are associated. In the sequential approach, one member of the set of process plans so generated is selected according to one or more criteria, such as execution cost and execution time. Once the process plan selection problem is solved, and the process plan is fixed for each item, the production scheduling task is carried out for a set of jobs, each of which corresponds to an item, employing available resource amounts on the shop floor. This decomposition of the overall problem into a sequence of two subproblems for the generation of process plans and production schedules defines the sequential approach.

The sequential approach possesses the undeniable advantage of the simplification by decomposition of a large problem into a pair of smaller subproblems; however, the decomposition constitutes in essence a heuristic, in that the solution that it provides is an approximation to the solution of the original process planning / production scheduling problem. As a heuristic, it may provide adequate solutions albeit not optimal. In fact, an optimal solution can only be envisaged if a simultaneous approach is adopted, whereby the original problem is attacked without its decomposition into a pair of subproblems. This has become to be known as the integrated process planning and scheduling (IPPSP). Clearly, the simultaneous approach, whilst providing an optimal solution, involves a larger and therefore more difficult problem than that involved in the sequential approach. An additional and important practical advantage of the simultaneous approach is that under certain conditions in the shop floor, it may provide feasible solutions when this is not possible with the sequential approach. An example of this arises when job due dates are excessively tight when process plans are fixed prior to job arrival on the shop floor, as is the case in the sequential approach. In contrast, in the simultaneous approach, the degree of freedom provided by selecting process plans along with job production scheduling decisions may allow for meeting job due dates.

Having provided a motivation and a justification for research efforts on the IPPSP, the next question that arises pertains to its formulation and solution. A start is made by providing a typical problem statement of the IPPSP as follows. Given a set of independent jobs, with each of which a delivery due date is associated, that is to be manufactured employing one of a set of process plans, each of which consisting of a sequence of operations, with each of which a processing time and a set of resource amounts is associated, it is required to determine a minimum makespan production schedule. Clearly, in the IPPSP what is being sought is an optimal schedule. Consequently, mathematical programming (MP) provides a natural framework for the formulation and solution of the IPPS problem. Extensive work has been reported on variants of the IPPSP employing MP models; see (Li, et al., 2010) for further details.

1.2 Multimode resource constrained project scheduling

A major task in project management is project scheduling whose objective is the sequencing of project activities, which may be executed in one or more modes, subject to logical precedence between activities and limited resources. This has given rise to the resource – constrained project scheduling problem (RCPSP). Extensive work has been carried out on the RCPSP; see [5] for a recent literature review. An important extension of the RCPSP is the multimode variant (MRCPSP), whereby an activity may be executed in more than one mode, whereby mode is a proxy for intensity, in that the activity work content may be executed at various speeds and corresponding resource consumption rates.

A typical problem statement of the MRCPSP may be provided as follows. Given a project that consists of a set of activities with a corresponding set of logical precedence relations, a set pair of renewable and nonrenewable resources, it is required to determine a minimum makespan schedule. Obviously, the goal is the generation of an optimal schedule, and for this mathematical programming (MP) constitutes a natural framework for the formulation and solution of the MRCPSP. An extensive literature exists on MP modeling of the MRCPSP; see (Hartmann & Briskorn, 2010), for a recent review.

2. Process plan – Execution mode relation

The similarity between two a priori different problems that arise in manufacturing systems management, IPPSP and MRCPSP, stems from the analogous roles of process plan and activity mode in the IPPSP and MRCPSP tasks, respectively. A job that is to be produced may be viewed as an activity to be completed, whilst a process plan corresponds to an execution mode. Extending the analogy, multiple jobs correspond to multiple projects in the corresponding environment.

If on the one hand, a process plan is similar to an execution mode, there exists an important conceptual difference. The former consists of a set of operations which possesses a corresponding set of precedence relations, whilst the latter is indivisible. Nevertheless, this conceptual difference results in practical implications when MP models are formulated and solved for the IPPSP and MRCPSP. In particular, the number of constraints in the former is higher in the IPPSP than in the MRCPSP of equal size, due to the existence of precedence relations in a process plan as opposed to the indivisibility of an execution mode. On a closer look, this difference may be conveniently removed if one considers the division of activities into subactivities, as suggested in (Nicoletti & Nicoló, 1998). The introduction of the concept of subactivity may not be merely semantic, as it may possess a practical aspect, such as activity preemption. The interruption of activities has been shown to provide a manner of fast tracking. In practice, an activity may only be interrupted at a finite number of points during its execution. These points may serve as a useful basis for the definition of subactivities.

3. Ship block assembly

Normally, the assembly of a ship block is part of a shipbuilduing project, which comprises several blocks. As a result, the assembly of each block should be appropriately scheduled, so as to be compatible with the schedule of the shipbuilding project as a whole. In general, it is desirable to determine the earliest delivery date of each block. The problem of interest then is to determine the minimum makespan assembly schedule for each block. This problem may be viewed as a project scheduling problem with resource and material supply constraints; see, for example, Dodin & Elimam, 2001, Alfieri *et al*., 2010, and Sajadieh, *et al*., 2009.

The following assumptions are made with a view to facilitating modelling of the problem.

1. With each activity, a duration time and a set of direct predecessor activities, and a set of direct successor activities are associated.
2. With each activity, a set of non – renewable resources, these being materials, is associated, this set being necessary to start and finish the activity.
3. With each member of the material set, a procurement lead time and a spatial area are associated, this set being necessary to start and finish the activity.

4. With each activity, a set of renewable resources, such as spatial area and manpower, is associated.

Spatial area, which is a member of the set of renewable resources, is employed for two purposes: execution of an activity and storage of associated materials from delivery time to the start time of the execution of the activity.

The following notation is employed:

i,j – indices for activities; $i \in I$
k – index for non-spatial area renewable resource; $k \in K$
m – index for non-renewable material source; $m \in M$
n – index for spatial area renewable resource; n=1
t – index for time period; $t \in T$
D_i – duration time of activity i
EF_i – earliest finish time of activity i
ES_i – earliest start time of activity i
F_i – set of direct predecessor activities of activity i
G_m – lead time material m
H_{imn} - spatial area renewable resource n required by material m associated with activity i
LF_i – latest finish time of activity i
LS_i – latest start time of activity i
P_{ik} – non-spatial area renewable resource per unit time required by activity i
Q_n – available non - spatial area renewable resource n
R_{mt} – inventory level of material non – renewable resource m in time period t
S_n – available spatial area renewable resource n
X_{it} – binary variable ∈ { 1 if activity i is finished in time period t, 0 otherwise }
Y_{it} – binary variable ∈ { 1 if activity I is in execution in time period t, 0 otherwise }
Z_{imt} – binary variable ∈ { 1 if material m associated with activity I is ordered in time period t, 0 otherwise }

The model may be stated as follows.

$$\text{Minimise } \Sigma_{\tau=EFI,\ldots LFI} \ \tau X_{I\tau} \ , \tag{1}$$

subject to

$$\Sigma_{\tau=EFj,\ldots,LFj} \tau \ X_{j\tau} + D_i \leq \Sigma_{\tau=EFi,\ldots,LFi} \ \tau X_{i\tau} \ , \ \forall j \in F_i \ ; \forall i \in I \ , \tag{2}$$

$$\textstyle\sum_{\tau=1,\ldots,T} \tau \ Z_{im\tau} \leq \sum_{\tau=EFi,\ldots,LFi} \tau \ X_{i\tau} - D_i - G_m \ , \forall i \in I \ ; \ \forall m \in M \ , \tag{3}$$

$$\Sigma_{i \in I} \ P_{ik} \ Y_{it} \leq Q_k \ , \ \forall k \in K \ ; \ \forall t \in T \ , \tag{4}$$

$$\Sigma_{m \in M} \quad R_{mt} \leq S_n \ , \ \forall t \in T \ , \tag{5}$$

$$R_{mt} = R_{m,t-1} + \Sigma_{i \in I} H_{imn} \ (Z_{im,t-Gm} - X_{it}) \ , \forall m \in M \ ; \ \forall t \in T \ , \tag{6}$$

$$R_{m0} = 0 \ , \forall m \in M \ , \tag{7}$$

$$\Sigma_{\tau=EFi,\ldots,LFi} \ X_{i\tau} = 1 \ , \forall i \in I \ , \tag{8}$$

$$\Sigma_{\tau=EFi,\dots,LFi} \; Y_{i\tau} = D_i \;, \forall i \in I \;, \quad (9)$$

$$\Sigma_{\tau=1,\dots,T} \; Z_{im\tau} = 1 \;, \forall i \in I \;;\; \forall m \in M \;, \quad (10)$$

$$X_{it}, Y_{it}, Z_{imt} \in \{0, 1\} \;, \forall i \in I \;;\; \forall m \in M \;;\; \forall t \in T \;. \quad (11)$$

Expression (1) defines the project makespan objective function that is to be minimised. Constraints (2) ensure precedence between an activity and each of its direct predecessor activities. Constraints (3) ensure that all materials necessary for each activity are available before the start of an activity. Constraints (4) and (5) guarantee that each renewable resource is not exceeded in each time period. Constraints (6) ensure the material balance over the planning horizon. Constraints (7) fix the initial inventory level of each material at the start of the planning horizon. Constraints (8) ensure that an activity is finished once in the planning horizon. Constraints (9) define the time duration of each activity. Constraints (10) guarantee that each material is ordered once for each activity over the planning horizon. Constraints (11) define the domain of the decision variables.

3.1 Three – level assembly problem

Consider for illustrative purposes the three – level block assembly example shown in Fig.1. This problem possesses features which belong to both the IPPSP and MRCPSP. As it arises in shipbuilding practice, it is naturally viewed as a variant of the IPPSP. With a view to highlighting the connection with the MRCPSP, it is modelled in this Subsection as belonging to the MRCPSP. The initial and final activities are denoted by i= 0 and i= 8, respectively. It is assumed that there exists a single non-spatial renewable resource, namely manpower, and it is denoted by k=1; furthermore, it is assumed that the spatial area renewable resource is common to all activities and materials in the assembly workshop and it is denoted by n=1. The model of this example may be stated as follows.

$$\text{Minimise } \Sigma_{\tau=EF8,\dots,LFT8} \; \tau \, X_{8\tau} \;, \quad (12)$$

subject to

$$\Sigma_{\tau=EF0,\dots,LF0} \; \tau \, X_{0\tau} + D_1 \le \Sigma_{\tau=EF1,\dots,LF1} \; \tau \; X_{1\tau} \;, \quad (13)$$

$$\Sigma_{\tau=1,\dots,T} \; \tau \, Z_{11\tau} \le \tau X_{1\tau} - D_1 - G_1 \;, \quad (14)$$

$$\Sigma_{i=1,\dots,8} \; P_{i1} \; Y_{1t} \le Q_1 \;, \quad (15)$$

$$\Sigma_{m=1,\dots,8} \; R_{m1} \le S_1 \;, \quad (16)$$

$$R_{12} = R_{11} + \Sigma_{i=1,\dots,8} \; H_{i11} \; (Z_{i1,1-G1} - X_{11}) \;, \quad (17)$$

$$R_{10} = 0 \;, \quad (18)$$

$$\Sigma_{\tau=EF1,\dots,LF1} \; X_{1\tau} = 1 \;, \quad (19)$$

$$\Sigma_{\tau=EF1,\dots,LF1} \; Y_{1\tau} = D_1 \;, \quad (20)$$

$$\Sigma_{\tau=EF1,\dots,LF8} \; Z_{11\tau} \;. \quad (21)$$

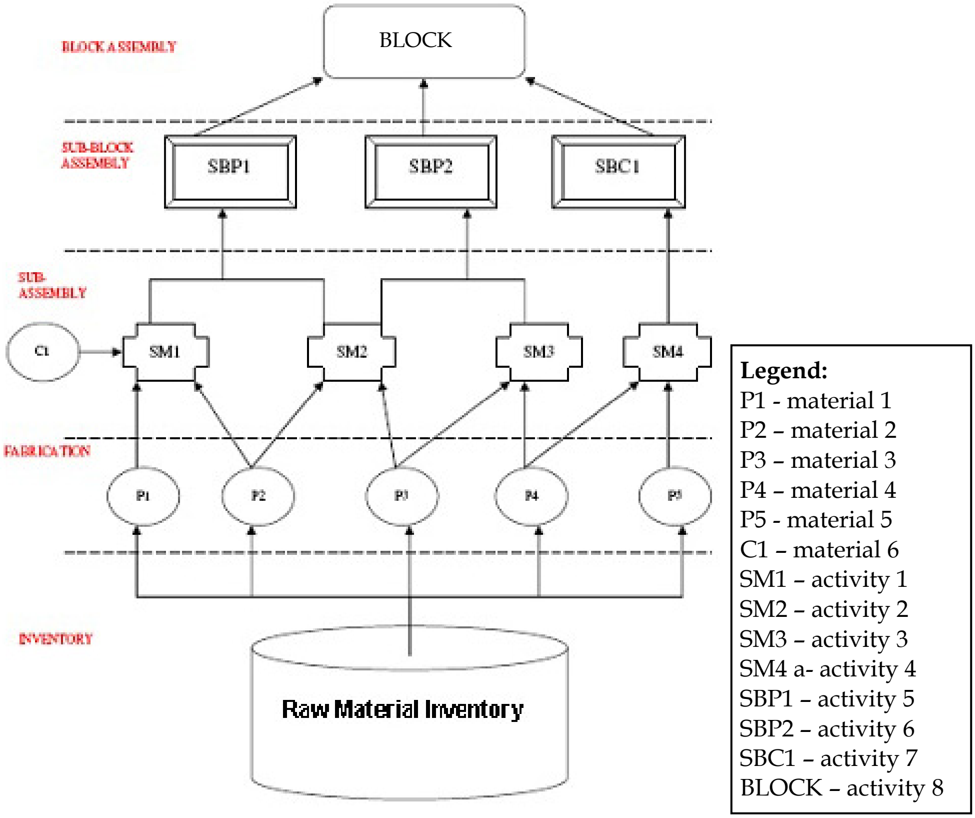


Fig. 1. Three – level assembly example.

4. Conclusions

In this paper, the relation between two problems, IPSS and MRCPSP, arising in manufacturing systems management, which have been formulated and solved by separate mathematical programming modeling approaches, have been shown to possess clear similarity features. This similarity has not been explored, and therefore there exists a clear potential for the interplay of methods and techniques between the two problems. These include novel model formulations and solution strategies for the IPPSP and MRCPSP variants, such as dynamic shop floor scheduling and project rescheduling (Ouelhadj & Petrovic, 2009), (Gerk & Qassim, 2008). This provides a rich field to be explored by future research work, a start having been made in a recent paper (Capek et al., 2011).

5. Acknowledgements

The author wishes to acknowledge financial support, in the form of a research fellowship, of the Brazilian National Research and Development council, CNPq. Recommendations and suggestions by an anonymous referee with a view to improving the paper are gratefully acknowledged.

6. References

Alfieri, A., Tolio, T., Urgo, M. (2010) A Project scheduling approach to production and material requirement planning in Manufacturing-to-Order environments. *Journal of Intelligent Manufacturing*, DOI 10.1007/s10845-010-0396-1.

Bang, J.-Y.; Kim, Y.-D. (2010). Hierarchical production planning for semiconductor water fabrication based on linear programming and discrete – event simulation. *IEEE Transactions on Automation Science and Engineering*, vol.7, no.2, pp,326-336.

Capek, R.; Sucha, P.; Hanzálek, Z. (2011). Production scheduling with alternative process plans. *European Journal of Operational Research*, doi: 10.1016/j.ejor.2011.09.018.

Dodin, B., Elimam, A.A. (2001) Integrated project scheduling and material planning with variable activity duration and rewards. *IIE Transactions*, vol.33, pp.1005-1018.

Gerk, J. E. V.; Qassim, R. Y. (2008) Project acceleration via activity crashing, overlapping, and substitution. *IEEE Transactions on Engineering Management*, vol.55, no.4, pp. 590-601.

Hartmann, S.; Briskorn, D. (2010). A survey of variants and extensions of the resource - constrained project scheduling problem. *European Journal of Operational Research*, vol.207, pp.1-14.

Kong, F.; Zuo, J.; Zha, G.; Zhang, J. (2011). Study and application of open die forging CAPP system based on process knowledge. *Journal of Advanced Manufacturing Systems*, vol.10, no.1, pp.45-52.

Laínez, J.M.; Reklaitis, G.V.; Puigjaner, L. (2010). Linking marketing and supply chain models for improved business strategic decision support. *Computers and Chemical Engineering*, vol.34, pp.2107-2117.

Li, X.; Shao, X.; Zhang, C.; Wang, C. (2010). Mathematical modeling and evolutionary – based approach for integrated process planning and scheduling. *Computers and Operations Research*, vol.37, pp.656-667.

Martínez, E. C., Duje, D., Pérez, G.A. (1997) On performance modeling of Project-oriented production. *Computers and Industrial Engineering*, vol.33, no.3, pp.509-527.

Martínez, E. C., Pérez, G. A. (1998) A project-oriented production model of batch plants. *Computers and Chemical Engineering*, vol.22, no.3, pp.391-414.

Monostori, L., Edos, G., Kádár, B., Kis, T., Kovács, A., Pffeifer, A., Váncza, A. (2010) digital enterprise solution for integrated production planning and control. *Computers in Industry*, vol.61, pp.112-126.

Munõz, E.; Capón-Garcia, E.; Moreno-Benito, M.; Espuña, A.; Puigjaner, L. (2011). Scheduling and control decision-making under an integrated information environment. *Computers and Chemical Engineering*, vol.35, pp.774-786.

Nicoletti, S., Nicoló, F. (1998). A concurrent engineering decision model : Management of the project activities information flow. *International Journal of Production Economics*, vol.54, pp.115-127.

Ouelhadj, D.; Petrovic, S. (2009). A survey of dynamic scheduling in manufacturing systems. *Journal of Scheduling*, vol.12, pp.417-431.

Sajadieh, M.S., Shadrkh, S., Hassanzadeh, F. (2009) Concurrent project scheduling and material planning: A genetic algorithm approach. *Scientia Iranica*, vol. 16, pp.91-99.

Shen, W.; Wang, L.; Hao, Q. (2006). Agent-based distributed manufacturing process planning and scheduling: A state – of – the – art survey. *IEEE Transactions on Systems, Man, and Cybernetics – Partt C: Applications and Reviews*, vol.36, no.4, pp.563-577.

Stray, J.; Fowler, J.W.; Carlyle, W.M.; Rastogi, A.P. (2006). Enterprise-wide semiconductor resource planning. *IEEE Transactions on Semiconductor Manufacturing*, vol.19, no.2, pp.259-268.

Yu-guang, Z., Kai, X., Yong, Z. (2011). Modeling and analysis of panel hull block assembly system through time colored petri net. *Marine Structures*, vol.24, no.4, pp.570-580.

16

Reliability Evaluation of Manufacturing Systems: Methods and Applications

Alberto Regattieri
DIEM - Department of Industrial and Mechanical Plants, University of Bologna
Italy

1. Introduction

The measurement and optimization of the efficiency level of a manufacturing system, and in general of a complex systems, is a very critical challenge, due to technical difficulties and to the significant impact towards the economic performance.

Production costs, maintenance costs, spare parts management costs force companies to analyse in a systematic and effective manner the performance of their manufacturing systems in term of availability and reliability (Manzini et al. 2004, 2006, 2008).

The reliability analysis of the critical components is the basic way to establish first and to improve after the efficiency of complex systems.

A number of methods (i.e. Direct Method, Rank Method, Product Limit Estimator, Maximum likelihood Estimation, and others (Manzini et. Al., 2009) all with reference to *RAMS* (Reliability, Availability, Maintainability and Safety) analysis, have been developed, and can bring a significant contribution to the performance improvement of both industrial and non-industrial complex systems.

Literature includes a huge number of interesting methods, linked for example to preventive maintenance models; these models can determine the best frequency of maintenance actions, or the optimization of spare parts consumption or the best management of their operating costs (Regattieri et al., 2005, Manzini et al., 2009).

Several studies (Ascher et al..1984, Battini et al., 2009, Louit et al., 2007, Persona et al. 2007) state that often these complex methodologies are applied using false assumptions such as constant failure rates, statistical independence among components, renewal processes and others. This common approach results in poor evaluations of the real reliability performance of components. All subsequent analysis may be compromised by an incorrect initial assessment relating to the failure process. A correct definition of the model describing the failure mode is a very critical issue and requires efforts which are often not sufficiently focused on.

In this chapter the author discusses the model selection failure process, from the fundamental initial data collection phase to the consistent methodologies used to estimate the reliability of components, also considering censored data.

This chapter introduces the basic analytical models and the statistical methods used to analyze the reliability of systems that constitute the basis for evaluation and prediction of the stochastic failure and repair behavior of complex manufacturing systems, assembled using a variety of components. Consequently, the first part of the chapter presents a general framework for components which describes the procedure for the solution of the complete Failure Process Modeling (FPM) problem, from data collection to final failure modeling, that, in particular, develops the fitting analysis in the renewal process and the contribution of censored data throughout the whole process. The chapter discusses the main methods provided in the proposed framework.

Applications, strictly derived from industrial case studies, are presented to show the capability and the usefulness of the framework and methods proposed.

2. Failure Process Modeling (FPM) framework

A robust reliability analysis requires an effective failure process investigation, normally based on non-trivial knowledge about the past performance of components or systems, in particular in terms of failure times. This data collection is a fundamental step. The introduction of a Computer Maintenance Management System - CMMS and of a Maintenance Remote Control System (Persona et al., 2007) can play an important role. Ferrari et al. (2003) demonstrate the risk due to a small data set or due to hasty hypothesis often considered (e.g. constant failure rate, independent identically distributed failure times, etc.).

Literature suggests different frameworks for the investigation of the failure process modeling of components and complex systems, generally focused on a particular feature of problem (e.g. trend tests in failure data, renewal or not renewal approach, etc.).

In this chapter a general framework is proposed considering all the FPM process from data collection to final failure modelling, also considering the contribution of censored data.

Figure 1 presents the proposed framework (Regattieri et al., 2010). Data collection is the first step of the procedure. This is a very important issue, since the robustness of analysis is strictly related to the collected data set. Both failure times and *censored* times are gathered. Times to failure are used for the failure process characterization and censored times finally enrich the data set used for the definition of the parameters of failure models, thus resulting in a more robust modeling.

In general, considering a population of components composed by m units, each specific failure (or inter-failure) time can be found. The result is represented by a set of times called Xi,j , where i^{th} represents the time of failure of the j^{th} unit: there is a complete data situation in this case, that is, all m unit failure times are available.

Unfortunately, frequently this is not a real situation, because a lot of time and information would be required. The real world test often ends before all units have failed, or several units have finished their work before data monitoring, so their real working times is unknown. These conditions are usually known as *censored data situations*.

Technically, censoring may be further categorized into:

1. Individual censored data
   All units have the same test time t*. A unit has either failed before t* or is still running (generating censored data);
2. Multiple censored data

Test times vary from unit to unit. Clearly, failure times differ but there are also different censoring times. Censored units are removed from the sample at different times, while units go into service at different times.

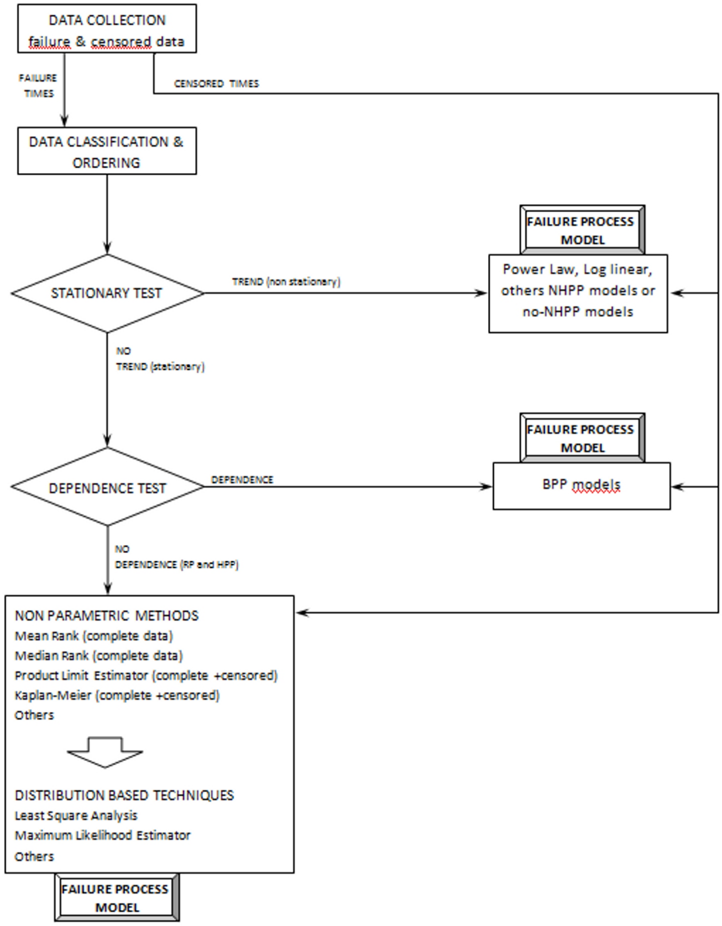


Fig. 1. Generalized framework for Failure Process Modeling (FPM)

In reference to Figure 1, let $X_{1,j} < X_{2,j} < \ldots < X_{i,j} < \ldots < X_{n,j}$ be the ordered set of failure or inter-failure times of item j; censored times (denoted $X_{ij}+$) are temporarily removed from the data set. The trend test applied to ordered failure times (graphical trend test, Mann test, etc.) determines if the process is stationary or not.

If the process presents a trend, the $X_{i,j}$ are not identically distributed and a non-stationary model must be fitted. The NHPP model is the most used form due to its simplicity and according to significant experimental evidence available (Coetzee, 1997). At this time the censored data must be reconsidered in the model. Their impact is discussed by Jiang et al. (2005).

If the failure process is trend free, the next step is to identify if inter-failure times are independent. There are a lot of tests for independence, but this check is usually skipped by practitioners, as stated by Ascher and Feingold, because of a lack of understanding of the relevance of this type of test. An effective way of testing the dependence is the serial correlation analysis discussed by Cox and Lewis (1966). Dependence between data involves a Branching Poisson Processes (BPP), which is also analyzed by Cox and Lewis (1966). Censored data also play an important role in the BPP model and must be considered during final modelling.

In real applications the failure process is frquently stationary and the failure data are independent: then a renewal process is involved. In spite of this, the proposed framework pays attention to the evaluation of reliability functions, in particular in presence of censored data.

More precisely, non parametric methods and distribution based techniques are suggested to find the reliability functions such as survival functions, hazard functions, etc. considering censored data.

The Product Limit Estimator method and Kaplan-Meier method for the first category and Least Square Analysis and Maximum Likelihood Estimator technique for the second category are robust and consistent approaches.

Regattieri et al. (2010), Manzini et al. (2009) and Ebeling (2005) discuss in details each method referred in the presented framework.

3. Applications

The proposed framework has been applied in several case studies. In this chapter, two applications are presented, in order to discuss methods, advantages and problems.

The first application deals with an important international manufacturer of oleo-dynamic valves. Using a reliability data set collected during the life of the manufacturing system, the effect of considering or not the censored data is discussed.

The second application involves the application of the complete framework in a light commercial vehicle manufacturing system. In particular, the estimation of the failure time distribution is discussed.

3.1 Application 1: The significant effect of censored data

During the production, the Company collects times to failure using the CMMS system. In particular, the performance of component *r.090.1768* is analysed; this is a very important electric motor: it is responsible of the movement of the transfer system of the valves assembly line. Figure 2 shows a sketch of the component.

The failure process can be considered as a Renewal Process; stationary test and dependence test are omitted for the sake of simplicity (for details see application 2).

Using non parametric methods, in particular the Kaplan Meyer method (Manzini et al. 2009, Ebeling, 2005) it is possible to evaluate the empirical form of reliability function (called $R(t_i)$).

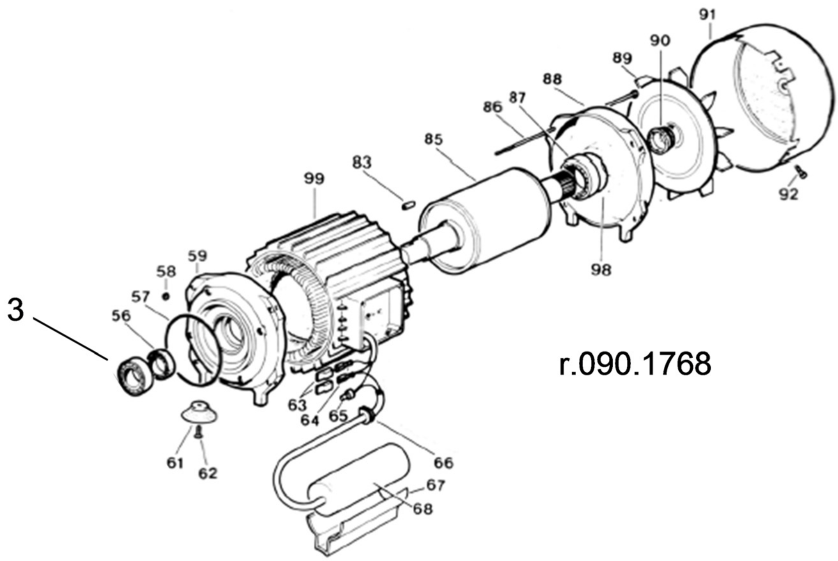


Fig. 2. Component r.090.1768

Table 1 presents all the available data, in terms of times to failures (t_i) and censored times (t_i+)

time to failure (h)	time to failure (h)	time to failure (h)
1124	667	2128
2785	700+	2500+
1642	2756	3467
800+	2489	2687
1974	1500+	1000+
2461	1945	1745
1300+	1478	1000+
2894	1500+	1348
3097	1246	2497
2674	2056	2500+

Table 1. r.090.1768 data set

Assuming *ti* as the ranked failure times and *ni* to be the number of components at risk, prior to the *ith* failure, the estimated reliability is calculated by:

$$\hat{R}(t_i) = \left(1 - \frac{1}{n_i}\right)^{\delta_i} \hat{R}(t_{i-1}) \quad (1)$$

where

$\delta_i = (1,0)$ (if failure occurs at time *ti* , if censoring occurs at time *ti*);

$$\hat{R}(0) = 1$$

The results of reliability analysis are summarized in Table 2 and Figure 3.

i	T_i		n_i	$(1-1/n_i)$	δ_i	R(ti)	
	0						1,000
1	667		30	0.967	1	R(667) = 0.967 R(0) =	0.967
2	700	+	29	0.966	0		
3	800	+	28	0.964	0		
4	1000	+	27	0.963	0		
5	1000	+	26	0.962	0		
6	1124		25	0.960	1	R(1124) = 0.960 R(667) =	0.928
7	1246		24	0.958	1	R(1246) = 0.958 R(1124) =	0.889
8	1300	+	23	0.957	0		
9	1348		22	0.955	1		0.849
10	1478		21	0.952	1		0.808
11	1500	+	20	0.950	0		
12	1500	+	19	0.947	0		
13	1642		18	0.944	1		0.764
14	1745		17	0.941	1		0.719
15	1945		16	0.938	1		0.674
16	1974		15	0.933	1		0.629
17	2056		14	0.929	1		0.584
18	2128		13	0.923	1		0.539
19	2461		12	0.917	1		0.494
20	2489		11	0.909	1		0.449
21	2497		10	0.900	1		0.404
22	2500	+	9	0.889	0		
23	2500	+	8	0.875	0		
24	2674		7	0.857	1		0.346
25	2687		6	0.833	1		0.289
26	2756		5	0.800	1		0.231
27	2785		4	0.750	1		0.173
28	2894		3	0.667	1		0.115
29	3097		2	0.500	1		0.058
30	3467		1	0.000	1		0.000

Table 2. Reliability evaluation using the Kaplan-Meier method (component r.090.1768)

Experimental evidences show that Companies often neglect censored data considering only the times to failure (i.e. the so called *complete* data set).

The use of the complete data set when several components are still working (i.e. there are censored data) introduces significant errors. Considering the component r.090.1768, Figure 4 shows a comparison between the Reliability functions obtained by Kaplan Meyer method applied to the set with censored data, and Improved Direct Method (Manzini et al., 2009) applied only to the failure times (complete data set).

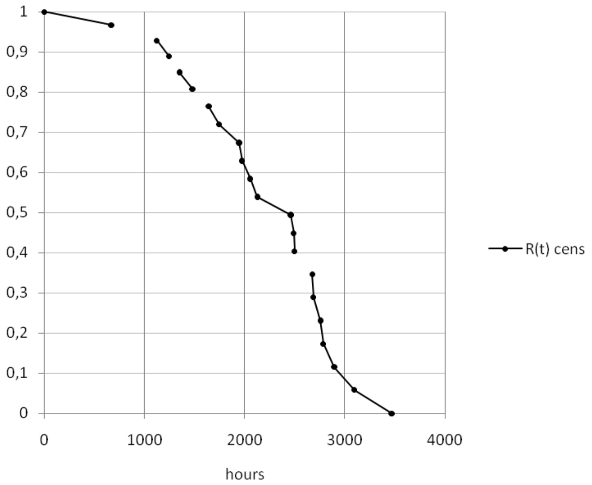


Fig. 3. Reliability Plot using Kaplan Meyer method (component r.090.1768)

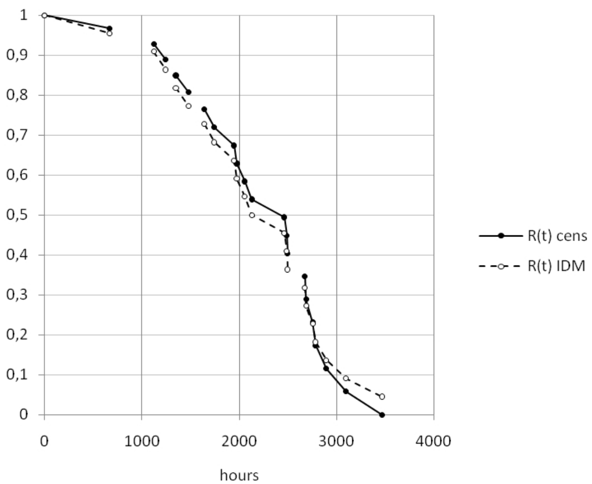


Fig. 4. Comparison of complete (using IDM) and censored data (using Kaplan Meyer)

The error is generally an under estimation of reliability. It depends on the percentage of units not considered and on the censoring times.

Anyway, if censored times are not considered, a significant error is introduced.

3.2 Application 2: A complete failure process modeling

In this application the FPM process has been applied to carry out the reliability analysis of several components of a light commercial vehicles manufacturer production system. The plant is composed by a lot of subsystems; for each of them a set of critical components is considered. Each component has a preferable failure mode (wear, mechanical crash, thermal crash, etc.); from now on, the generic expression *failure* is used considering the particular and prominent failure mode for each component.

Table 3 shows the subset of critical components, called S1, analyzed in the Chapter.

No.	Component	Code	Subsystem
1	tow equipment	KKL5699	welding robot wr1
2	brake shoes	BF4-45	welding robot wr1
3	primary hinge	IHF7598	input gate
4	penstock	KK1243	body feeder
5	skillet ball bearing	A9097	skillet

Table 3. Critical components and subsystems

Among them, the main welding robot named wr1 is very critical, and in particular its components named KKL5699, which are the main actuators, are considered to be mainly responsible for the poor reliability performance of the entire manufacturing system. For this reason, the Chapter presents the application of the proposed framework to the component KKL5699. Finally, the conclusions take into account all the critical components shown in Table 1.

3.2.1 Analysis of KKL5699 component

The tow system is composed by 9 identical repairable components KKL5699, working in 9 different positions, named with letters A,B,…, L, under the same operating conditions. For this reason, they are pooled in a single enhanced data set. The working time is 24 hours/day, 222 days/year.

The CMMS has collected failure data from initial installation (T_0 = 0). Table 4 reports the interfailure time X_{ij} (failure i of item j) and the cumulative failure times F_{ij} as shown in Figure 5.

The data are collected during 5 years of operating time, but FPM must be an iterative procedure applied at different instants of system service. The growth of the data set allows a more robust investigation of the failure process. In particular, the paper involves the results of analysis developed at the end of different time intervals [T_0,t] : 1,440, 3,696, 4,824, 6,720, 8,448, 11,472, 13,440, 15,560, 18,816 and 23,688 hours. For the generic time t,the analysis uses all the failure times collected, but also the existing censored times according to the components in service. For each instant of analysis, 9 suspended times are collected due to the working times from the last repair action of components and the time analysis. Table 5 reports the data set of failure times, the censored times and the relative working position available at the instant of analysis 3,956 hours.

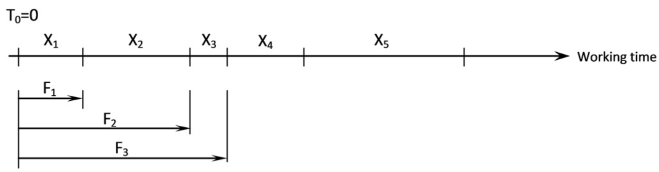


Fig. 5. Inter-failure time X_{ij} and cumulative failure times F_{ij}

Obviously, when the instant of analysis increases, the number of failure times increases too. Whereas the censored times are constantly 9, then the *Censoring Rate - CR*, given by (1), decreases:

$$CR = \frac{Nc(t)}{Ntot(t)} \tag{2}$$

where *Nc(t)* is the number of censored times available at time analysis *t*, and *Ntot(t)* is the number of times (failure and censored) available at time analysis *t*.

The different Censoring Rates involved in the analysis are presented in Table 6.

i	X_{iA} (h)	F_{iA} (h)	X_{iB} (h)	F_{iB} (h)	X_{iC} (h)	F_{iC} (h)	X_{iD} (h)	F_{iD} (h)	X_{iE} (h)	F_{iE} (h)
1	829	829	1.658	1.658	673	673	2.134	2.134	2.724	2.724
2	1.132	1.961	4.536	6.194	983	1.656	1.864	3.998	1.672	4.396
3	2.354	4.315	2.353	8.547	1.567	3.223	3.276	7.274	1.945	6.341
4	1.856	6.171	899	9.446	2.349	5.572	1.745	9.019	2.218	8.559
5	1.957	8.128	1.654	11.100	3.314	8.886	1.794	10.813	1.634	10.193
6	3.420	11.548	1.673	12.773	1.786	10.672	2.398	13.211	2.835	13.028
7	7.264	18.812	2.364	15.137	1.745	12.417	1.128	14.339	1.674	14.702
8	2.573	21.385	1.879	17.016	2.234	14.651	1.523	15.862	1.893	16.595
9	1.111	22.496	2.657	19.673	987	15.638	1.759	17.621	2.673	19.268
10					1.756	17.394	3.298	20.919	1.278	20.546
11					2.567	19.961	1.245	22.164		
12					2.163	22.124				

i	X_{iF} (h)	F_{iF} (h)	X_{iG} (h)	F_{iG} (h)	X_{iH} (h)	F_{iH} (h)	X_{iL} (h)	F_{iL} (h)
1	2.143	2.143	1.980	1.980	3.278	3.278	2.173	2.173
2	1.783	3.926	2.637	4.617	2.167	5.445	2.184	4.357
3	2.784	6.710	1.857	6.474	2.783	8.228	1.476	5.833
4	2.647	9.357	1.643	8.117	1.924	10.152	1.284	7.117
5	1.673	11.030	2.195	10.312	2.735	12.887	749	7.866
6	985	12.015	785	11.097	2.196	15.083	4.583	12.449
7	1.643	13.658	1.903	13.000	2.457	17.540	2.748	15.197
8	1.849	15.507	2.749	15.749	1.932	19.472	2.476	17.673
9	2.649	18.156	2.374	18.123	1.384	20.856	2.945	20.618
10	3.916	22.072	2.734	20.857	1.836	22.692	1.932	22.550
11	673	22.745	1.925	22.782				

Table 4. Inter-failure and failure time data set (KKL5699)

According to the proposed framework, the first test deals with the stationary condition. Figure 6, referring to all the pooled data set, presents the *cumulative failures vs time plot* graphs. No trend can be appreciated in the failure data. The Mann test, counting the number of reverse arrangements, confirms this belief, both for each component and for the pooled data set. The Laplace trend test leads to the same conclusions, in particular its test statistic is $u_L=0.55$ according to a p-value $p=0.580$. Comparing u_L^2 with the χ^2 distribution with 1 degree of freedom, there is no evidence of a trend with a significance level of 5% (Ansell and Philips, 1994).

interfailure times X_{ij} (h)	position	censored times X_{ij}^+ (h)	position
829	A	1.735	A
1.132	A	2.038	B
1.658	B	473	C
673	C	1.562	D
983	C	972	E
1.567	C	1.553	F
2.134	D	1.716	G
2.724	E	418	H
2.143	F	1.523	L
1.980	G		
3.278	H		
2.173	L		

Table 5. Data set at 3.956 working hours (inter-failure times and censored times) component KKL5699

Instant of analysis (hs)	1440	2410	3696	4824	6720	8448	11472	13440	16560	18816	23688
N. of interfailure times X_{ij}	2	9	12	18	26	32	46	55	66	75	93
N. of censored times X_{ij}^+	9	9	9	9	9	9	9	9	9	9	9
Censoring rate - CR	0,818	0,500	0,429	0,333	0,257	0,220	0,164	0,141	0,120	0,107	0,088

Table 6. Censoring rates (KKL5699)

A first assessment shows that the component failure process is stationary. The next step deals with the renewal process hypothesis evaluation.

The *serial correlation analysis* is used to reveal the independence of the analyzed data set. According to the *autocorrelation plot*, in Figure 7, we cannot reject the null hypothesis of no autocorrelation for any length of lag (5% is the significance level adopted).

The Durbin-Watson statistic confirms this belief, then the component failure process for component KKL5699 can be considered to be a renewal process (RP).

Considering the RP assumption, *non parametric* methods or *distribution based techniques* are available to define the failure process model.

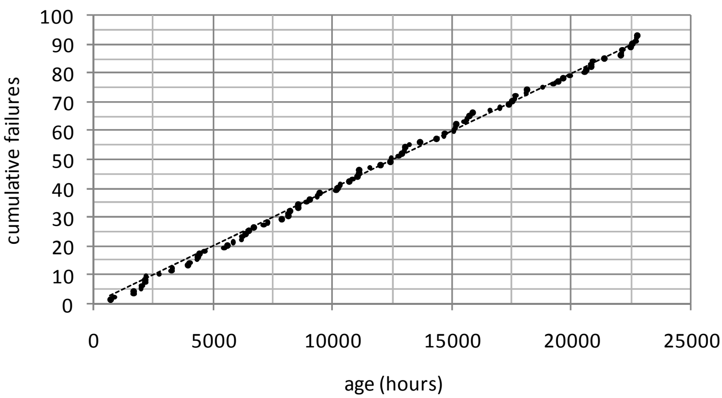


Fig. 6. Cumulative failures vs time plot (pooled data set KKL5699)

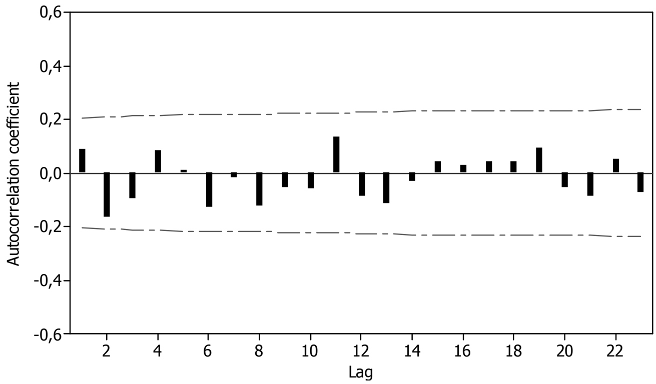


Fig. 7. *Correlogram* for pooled *Xij* with 5% significance level (KKL 5699)

It is important to take into consideration the role of censored data: as demonstrated in the previous application 1, their use enhances the data set and then increases the confidence of the model.

According to the censored data consideration, data are previously analyzed by the *Product Limit Estimator* method (Manzini et al. 2009, Ebeling, 2005), and then the best reliability distributions (i.e. survival function, hazard rate, etc.) are fitted by the *Least square analysis* method. This approach is applied for each time interval of the system service time (i.e. instant of analysis).

From now on, we will only consider the pooled data set; for this reason, the time notation X_{ij} collapses into X_z. The value of survival function - R(X_z), after a working time equal to X_z using the Product Limit Estimator method, is given by:

$$R(X_z) = \left(\frac{n+1-z}{n+2-z} \right)^{\delta_z} R(X_{z-1}) \tag{3}$$

where

δ_z = (1;0) (if failure occurs at time X_z ; if censoring occurs at time X_z);
R(0) = 1;
n number of failure and censored events available

All the service times are investigated. Figure 8 shows the survival empirical curves for component KKL5699, obtained at different instants of analysis, sometimes very spaced out one from each other (e.g. several months apart).

The reliability evaluation after 1,440 hours (roughly 3 months of service) appears very approximate, while after 23,688 hours (more than 4 years of service) it is very confident. Considering the data set, the survival function evaluations change in a significant way (up to 25%) along the life of component; the data collection and the maintenance of data collected is thus a very important issue.

An alternative approach is to identify a proper statistical distribution for the principal reliability functions, such as the survival function, *R(t)*, the failure cumulative probability function, *F(t)*, and the hazard function, *h(t)*, to evaluate its parameter(s), and perform a goodness-of fit test.

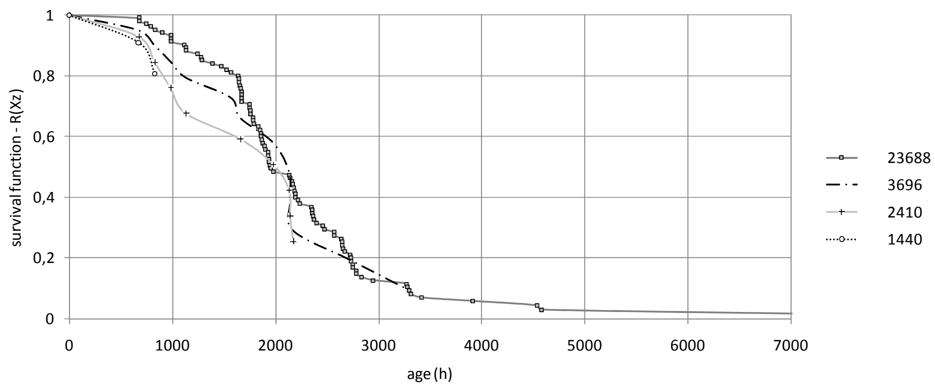


Fig. 8. Survival function at different service times (1,440, 2,410, 3696, 23,688 hours) for component KKL5699

In general, this approach is very interesting because in recent years a significant number of techniques, based on the knowledge of the reliability distributions, have been developed, and then they can provide an important contribution to the performance improvement of both industrial and non-industrial complex systems. These techniques are often referred to as *RAMS (Reliability, Availability, Maintainability and Safety)* analysis.

For example, scientific literature includes a huge number of interesting methods, some being more complex than others, linked to preventive maintenance models; these models support the determination of the best intervention interval, or the optimization of the procedures, that determine spare parts consumption or the best management of their operating costs.

The *2-parameters Weibull* distribution is one of the most commonly used distributions in reliability engineering because of the many failure processes it attains for various values of parameters. It can therefore model a great variety of data and life characteristics. The distribution parameters can be estimated using several methods: the Least Square method (LS), the Maximum Likelihood Estimator (MLE) and others (Ebeling, 2005). In the KKL5699 component case, the Least Square method is preferred for its simplicity and robustness.

Table 7 summarizes the results in terms of Weibull parameters and index of fit, according to different instants of analysis.

Instant of analysis (hs)	1.440	2.410	3.696	4.824	6.720	8.448	11.472	13.440	16.560	18.816	23.688
shape factor β	1,790	2,133	2,489	2,730	2,814	2,749	2,849	2,888	2,989	2,914	2,900
scale factor η	3.160,420	2.087,720	2.268,320	2.251,380	2.378,990	2.393,510	2.311,830	2.335,510	2.335,140	2.400,010	2.427,910
index of fit	0,9451	0,9642	0,9857	0,9871	0,9822	0,985	0,9769	0,9793	0,9788	0,9647	0,9723
Censoring rate - CR	81,8	50,0	42,9	33,3	25,7	22,0	16,4	14,1	12,0	10,7	8,8

Table 7. Parameters of Weibull distribution at different instants of analysis (KKL5699)

The parameters of Weibull distribution move toward steady values of about 3.0 for β and about 2.400 for η. Figure 9 shows graphically the trend of parameters according to different instants of analysis.

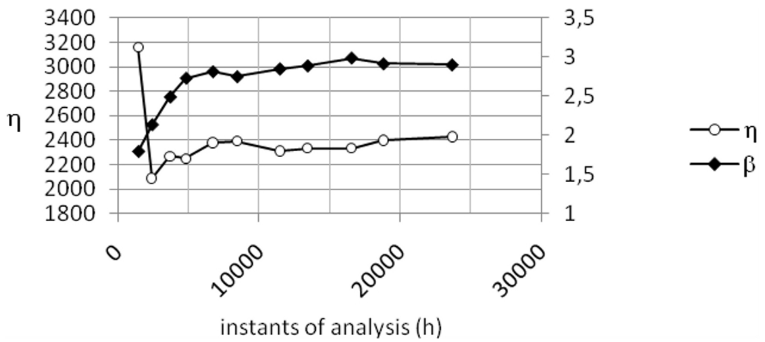


Fig. 9. Trend of Weibull parameters according to the different service times (KKL5699)

Another interesting analysis deals with the link between the estimate of Weibull parameters and the censoring rate. Figures 10 and 11 show the significant role played by *CR* on β and η paths. The different censoring rates are due to a fixed number of censored data (i.e. 9) and an increasing number of failure data.

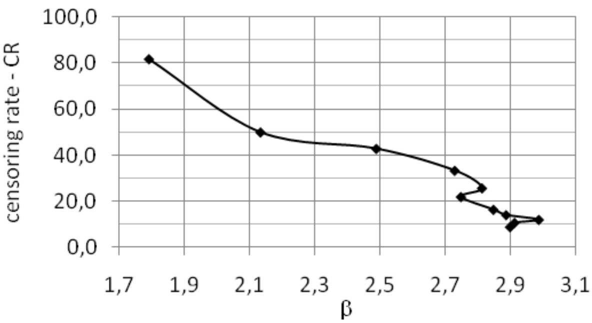


Fig. 10. Shape factor (β) according to the censoring rate (CR) (KKL5699)

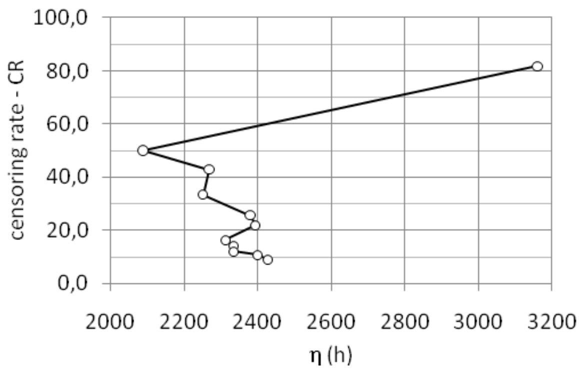


Fig. 11. Scale factor (η) according to the censoring rate (CR) (KKL5699)

Reliability functions, such as the survival function, the cumulative probability of failure and the hazard rate, can be directly derived by 2-parameters Weibull distribution using the estimated parameters of Table 7. Figure 12 shows the survival function R(t) of component KKL5699 for a generic *time interval t* estimated after 2,410, 4,824 and 23,688 hours of service.

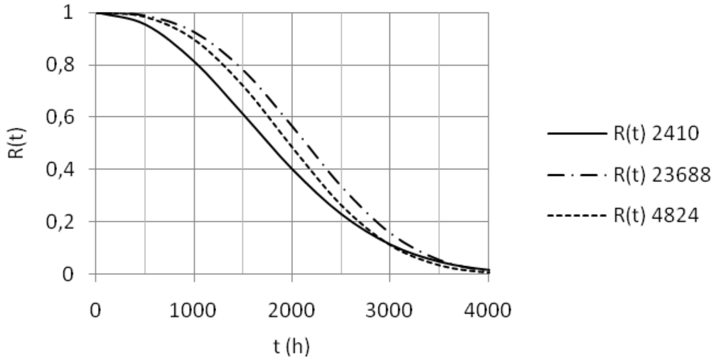


Fig. 12. Survival function (Weibull distribution) according to different instants of analysis (KKL5699)

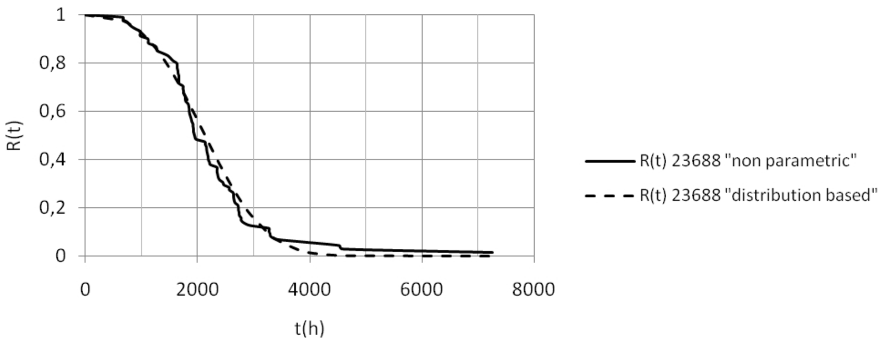


Fig. 13. Comparison between Survival functions obtained by non-parametric method and by 2-parameters Weibull distribution for component KKL5699 after 23,688 hours.

The proposed framework (Fig.1) provides both the *non parametric* and the *distributions based* approaches. Their use usually results in similar outcomes but, as stated before, the second one incorporates more information and is preferable when possible. Figure 13 shows the comparison between the estimates of the survival functions of component KKL5699 obtained after 23,688 hours of service adopting the two different approaches.

3.2.2 Analysis of the subset S1

The stationary and dependence tests of the times to failure highlight how all the components of subset S1 can be described as having a Renewal Process behavior. Failure and censored data are used to perform the Product Limit Estimator method to evaluate empirical values of the cumulative failure distributions. The *2-parameters Weibull* distribution is used to evaluate the reliability functions in an analytical manner; in particular the distribution parameters are estimated using the Least Square method for each instant of analysis. Table 8 summarizes the results in terms of Weibull parameters and in terms of index of fit after 23,688 operating hours.

No.	Component	Code	shape factor β	scale factor η	index of fit
1	tow equipment	KKL5699	2,900	2.427,910	0,9723
2	brake shoes	BF4-45	2,071	3.157,351	0,9885
3	primary hinge	IHF7598	3,541	1.879,467	0,9478
4	penstock	KK1243	2,761	4.578,568	0,9627
5	skillet ball bearing	A9097	2,684	3.125,457	0,9798

Table 8. Parameters of Weibull distribution of S1 components after 23.688 operating hours

Figure 14 shows the survival function of the critical components as calculated on the basis of the parameters reported in Table 8.

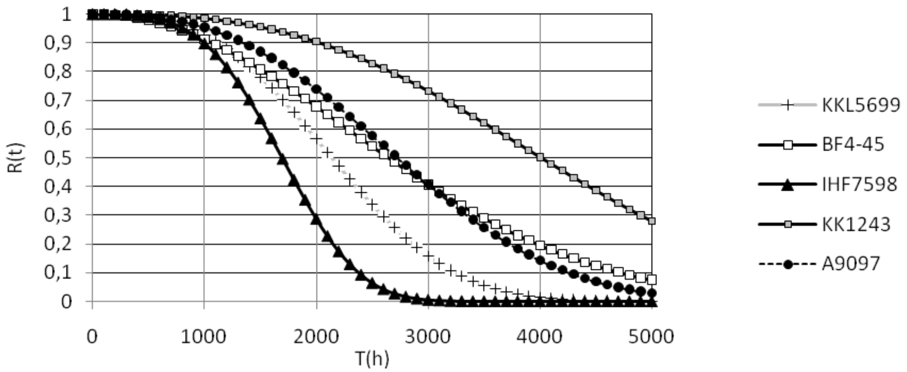


Fig. 14. Survival function of S1 components (2-parameters Weibull)

4. Conclusion

The Failure Process Modeling plays a fundamental role in reliability analysis of manufacturing systems. Complex methodologies are often applied using false assumptions such as constant failure rates, statistical independence between components, renewal processes and others. These misconceptions result in poor evaluation of the real reliability performance of components and systems. All complicated subsequent analysis may be compromised by an incorrect initial assessment relating to the failure mode process.

The experimental evidences show that a correct definition of the model describing the failure mode is a very critical issue and requires efforts often not sufficiently focused on by engineers.

The information collection of both failure data and censored data is a fundamental step. The CMMS method and a system automatically managing the alarms coming from the different sensors installed, can represent valid tools to improve this phase.

As demonstrated in the presented applications, the neglecting of the censored information results in significant errors in the evaluation of the reliability performance of the components.

The knowledge of a fitted analytical distribution is very interesting, because it allows several developments: for example, the determination of the best intervention frequency, or the optimization of the procedures that determine spare parts consumption or the best management of their operating costs.

The FPM procedure must also be maintained: during the service of systems, the reliability data set grows and a more robust estimation of reliability functions is allowed. FPM process, performed using the proposed framework (Fig. 1), must be an iterative procedure renewed during the life of systems.

5. References

Ansell J.I., Phillips M.J., *Practical methods for reliability data analysis*, Oxford University Press, New York, 1994.

Ascher H.E. Feingold H., *Repairable systems reliability. Modeling, inference, misconceptions and their causes*, Marcel Dekker, New York, 1984.

Battini D., Persona A., Regattieri A., Buffer size design linked to reliability performance: a simulative study, *Computers and Industrial Engineering* 56(4), pp. 1633-1641, 2009.

Coetzee J., The role of NHPP models in practical analysis of maintenance failure data, *Reliability Engineering & System Safety*, 56 (2), pp. 161-168, 1997.

Cox D.R., Lewis P.A., *The statistical analysis of series of events*, Methuen, London, 1966.

Ebeling C.E., *Reliability and maintainability engineering*, McGraw-Hill, New York, 2005.

Ferrari E., Gamberi M., Pareschi A., Persona A., Regattieri A., On importance of reliability data collection for failure rate modelling: Reliability Active Data Base – RADB. *9th International Conference on reliability and quality in design*. pp. 132-137. Honolulu – USA, 2003.

Jiang S.T., Landers T.L., Rhoads T.R., Semi-parametric proportional intensity models robustness for right-censored recurrent failure data, *Reliability Engineering & System Safety*, 90 (1), pp. 91-98, 2005.

Louit D.M., Pascual R., Jardine A.K.S., A practical procedure for the selection of time-to-failure models based on the assessment of trends in maintenance data, *Reliability Engineering & System Safety*, Elsevier, London, online.

Manzini R., Regattieri A., Pham H., Ferrari E., *Maintenance for industrial systems*, ISBN: 978-1-84882-574-1, Springer, London, 2009.

Manzini R., Gamberi M., Regattieri A., Persona A., Framework for designing a flexible cellular assembly system, *International journal of production research* 42 (17), pp. 3505-3528, 2004.

Manzini R., Gamberi M., Regattieri A., ,Applying mixed integer programming to the design of a distribution logistic network, *International journal of industrial engineering-theory applications and practice*, 13(2), pp. 207-218. 2006.

Manzini R., Gamberi M., Gebennini E., Regattieri A., An integrated approach to the design and management of a supply chain system, *International journal of advanced manufacturing technology*, 37(5-6), pp.625-640, 2008.

O'Connor P., *Practical reliability engineering*, ISBN:978-0-47084-463-9, Wiley, New York, 2002.

Persona A., Regattieri A., Pham H., Battini D., Remote control and maintenance outsourcing networks and its applications in supply chain management, *Journal of operations management*, 25(6), pp.1275-1291, 2007.

Regattieri A., Gamberi M., Gamberini R., Manzini R., Managing lumpy demand for aircraft spare parts, *Journal of air transport management*, 11(6), pp. 426-431, 2005.

Regattieri A., Manzini R., Battini D., Estimating reliability characteristics in presence of censored data: a case study in a light commercial vehicle manufacturing system, *Reliability Engineering & System Safety*, 95 (10), pp. 1093-1102, 2010.

17

Migrating from Manual to Automated Assembly of a Product Family: Procedural Guidelines and a Case Study

Michael A. Saliba* and Anthony Caruana
Department of Industrial and Manufacturing Engineering, University of Malta
Malta

1. Introduction

A challenge that is often faced by product manufacturers is that of migration from a manual to an automated production system. The need to embark on this migration may develop from a number of different scenarios. Typical triggers for the automation of a manual production process include a need to increase competitiveness (by reducing labour costs and increasing labour productivity), a need to meet higher quality demands from the customer, an increased awareness of health and safety issues leading to the need to move human workers away from hazardous tasks, and a need to improve production efficiency parameters (e.g. reduce manufacturing lead time, improve production capacity, improve production flexibility and agility). The general strategic and/or technical aspects pertaining to the implementation of automation have been widely addressed in the literature (e.g. Asfahl, 1992; Chan & Abhary, 1996; Groover, 2001; Säfsten et al., 2007). In this work the focus is on the compilation and application of a number of standard design tools and of production system evaluation tools to facilitate and support the migration to an automated production system, in a scenario that involves a certain degree of product variety. The results are presented in the form of a set of recommended procedural guidelines for the development of a conceptual solution to the migration process, and the implementation of the guidelines in a real industrial case study.

Specifically, this work addresses the situation where a manufacturer needs to investigate a potential manufacturing system migration for the assembly of a part family of products, where no or minimal product design changes are allowed. A list of procedural guidelines is proposed, in order to aid the manufacturer in analyzing the requirements for the transition, and in carrying out a conceptual design for a suitable automated manufacturing system. The guidelines are applied in the context of the industrial case study, where it is required to investigate and develop a migration plan for the assembly of three related product part families.

The case study involves a relatively large manufacturing plant (about 700 employees), that produces electromechanical switch assemblies for the automobile industry. The trigger for

* Corresponding Author

the migration process is a significant increase in projected order volumes over a four year period, thus necessitating an increase in production capacity, as well as providing a good opportunity to obtain substantial return on investment following the implementation of advanced manufacturing technologies. The three part families under consideration are referred to as the "single gang" switches consisting of 20 variants, the "old three gang" switches consisting of 11 variants, and the "new three gang" switches consisting of 9 variants. A representative member of each of these families is shown in the illustration in Figure 1. Switch assembly is currently carried out in a mainly manual manner, with the aid of pneumatic presses to provide the required clipping forces. The projected production volumes for the current year (Year 1) and for the subsequent years are summarized in Table 1.

The academic goals of this research are (i) to compile a set of guidelines for migration in a scenario of this type, and to apply the guidelines to this case study; (ii) to perform a study on assembly related similarities of the product families and to take advantage of these similarities in the automation process; (iii) to interpret the analytical results obtained from the feasibility analysis, so as to define the most suitable assembly line; and (iv) to utilize analytical tools in order to define the best possible concept at all stages of the automation. From an industrial perspective, the goals are (i) to perform a cost reduction exercise on the current assembly processes; (ii) to perform a feasibility analysis of the developed concepts with respect to a number of considerations such as the assembly line balancing, cycle time reduction, production capacity and maintenance requirements; and (iii) to plan the integration of ergonomic principles in all workstations and also develop and implement safety guidelines in the process development.

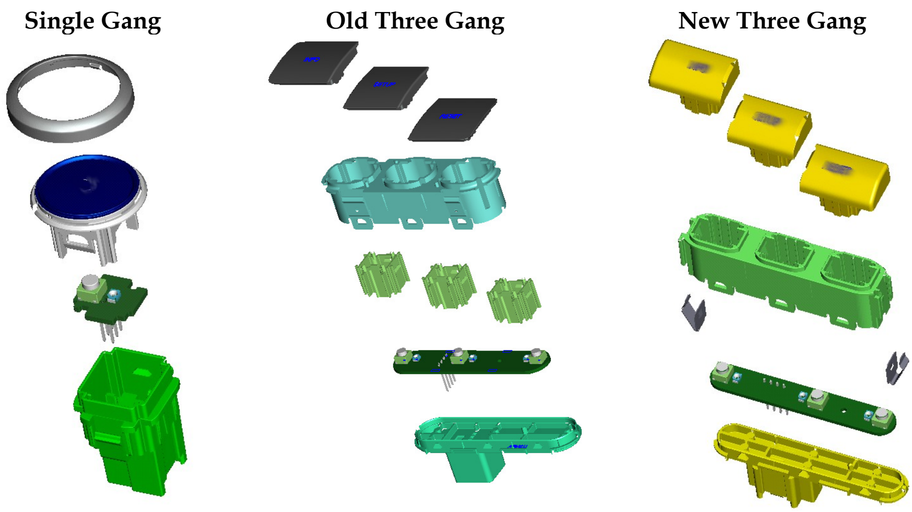


Fig. 1. An exploded view of a representative member of each product family

	Year 1	Year 2	Year 3	Year 4
Single gang switches	136,000	935,000	1,432,600	1,552,600
Old 3 gang switches	1,098,000	816,000	816,000	816,000
New 3 gang switches	0	150,000	865,000	1,495,000

Table 1. Projected annual production volumes

2. Literature review

In product manufacture, assembly needs to be given significant importance, since assembly related operations generally amount to 70% of the total product cost (Boothroyd et al., 2001). Therefore rationalisation of the assembly process is essential in order to optimize, mechanize and automate the activities performed, especially for assembly-intensive products. Where a new product is being developed, an *Integrated Product Development* (IPD) approach can be taken, whereby the marketing, design, and manufacture of the product family can be optimized and developed concurrently (Andreasen & Hein, 2000). Where the product family exists and is already in manual production, the options for design modifications may be minimal, and the development of the automated production system rests on effective categorization of the products through analytical methodologies such as group technology (e.g. Hyer & Wemmerlov, 2002), and the effective exploitation of the identified similarities.

Several approaches are found in the literature to address the migration problem. Asfahl (1992) identified five phases in the implementation of automation to a currently manual process: *planning*, *development*, *mock-up and test*, *installation*, and *production and follow-through*. He further highlighted a number of key activities to be carried out during the planning phase: the isolation of the potential application; the identification of the project objectives; the consideration of the drawbacks; the early planning of safety aspects; the detailed documentation of the current (manual) operation; the selection of fixed versus flexible automation; and the development of a proposed layout for the system. Chan and Abhary (1996) applied an analytic hierarchy process to compare three different potential automation strategies to the existing manual plant in a case study, using a simulation approach and several evaluation criteria. Kapp (1997) introduced the "USA Principle" – *U*nderstand the existing process, *S*implify the process, and *A*utomate the process, originally intended as a guide for the implementation of enterprise resource planning (ERP), but applicable as a straightforward approach to all automation projects. Groover (2001) suggested a three-phase process whereby the individual processing stations are automated first, followed by the integration of the systems through automated handling between stations. Baines (2004) recommended a nine step approach to the manufacturing technology acquisition process: technology profiling; establishment of technology requirements; identification of a technology solution; formation of an outline business case; selection of a technology source (which may include internal development of the technology); demonstration of the technology; confirmation of the business case; implementation of the technology; and post-investment audit. Säfsten et al. (2007) suggested that manufacturing strategy development could be based on function allocation, employing a system design process that allocates various functions to either humans or machines, and optimizing the level of automation in the plant. Winroth and Säfsten (2008) further suggested an automation strategy whereby the bottom-up activities (stemming from the internal need for improvement) and the top-down activities (stemming from the market requirements) are both taken into consideration in the optimization exercise.

Most of the technical literature on the development of automation systems focuses on the process that needs to be automated and/or on the product that needs to be manufactured, but less so on the production system as a product in its own right. Thus, a number of design methods that have become a mainstay in product design are not normally prescribed in a systematic way to the development of integrated automation systems. Design methods include tools and techniques such as product design specification (PDS), morphological charts, decision matrices, and failure modes and effects analyses (FMEA) (e.g. Dieter & Schmidt, 2009). This is the research gap that has been identified in this work, and the results reported in this chapter attempt to bridge this gap by drawing on these design methodologies, and prescribing them side by side with other conventional developmental steps, in order to optimize the conceptual design process for an automation system, in an environment of variety in the products that need to be manufactured. The contribution of this work is further extended to include a detailed illustration of the step by step application of the procedural guidelines to a complex industrial case study.

3. The procedural guidelines

A systematic approach to the conceptual design of a new, automated manufacturing system when migrating from manual assembly of a product part family, where little or no change in product design is allowed, is presented in Table 2. The proposed list of procedural guidelines and development tools given in the table has been compiled on the general basis of the discussion given in section 2, and the developmental steps are intended to be applied sequentially, for the case of a high production volume environment in the presence of product variety. The guidelines are intended to cover the early technical and feasibility studies. Thus it is pre-assumed that the company has already taken a strategic decision to analyze the selected manual process with a view to implementing automation (if feasible), and the guidelines lead to the end of the conceptual design phase but do not address any part of the embodiment design phase or of the development of test or prototype hardware. In the general literature, it has been estimated that about 75% of the product cost is normally already committed by the end of the conceptual design stage (Ullman, 1997), and in this work the research boundary has been set to address this critical phase of product development. It is emphasized once again that it is the *production equipment* that is being referred to and considered as the "product" in the context of the previous sentence, rather than the objects (products, or product part family) that will be manufactured by the equipment.

In the following section, the use and implementation of the procedural guidelines is illustrated in the context of the industrial case study, for the development of the conceptual design of a new manufacturing system for the assembly of the three families of automotive switches. The results for each step are summarized, presented and discussed.

4. Implementation of the procedural guidelines: A case study

4.1 Analysis of the product family designs

In a manufacturing environment, parts having similar geometric shapes and sizes or similar processing steps may be grouped into part families, in order to facilitate their design and/or their production. This manufacturing philosophy is referred to as *group technology*. Thus, parts within a particular part family will all be uniquely different, however they will have enough similarities to classify them together as one group (e.g. Groover, 2001). In the present case study, the product designs are fixed, and therefore the application of group technology principles is intended to facilitate the production of the parts. Parts classification systems are normally based either on similarities in design attributes, or on similarities in manufacturing attributes, or on similarities in both design and manufacturing attributes; however other types of similarity may also be used. Hyer and Wemmerlov (2002) identify nine criteria that may be used to classify parts, based on similarities in product type, market, customers, degree of customer contact, volume range, order stream, competitive basis, process type, and/or product characteristics.

1.	Analyze the current product family design, with a view to understanding clearly all similarities and variations between the members of the family and between their components.
2.	Analyze the current assembly processes and the existing assembly line(s), with a view to understanding the processes, and identifying drawbacks and opportunities for simplification.
3.	Perform a capacity analysis based on the current set-up, with a view to understanding and defining current capabilities and limitations.
4.	Draw up a product design specification (PDS) chart for the new production system, with a view to defining the requirements and wishes for the new system.
5.	Perform a group technology (GT) analysis, with a view to confirming/revising the parts classification in the context of automated manufacture.
6.	Create precedence diagrams for the process, with a view to understanding the various ways in which assembly operations can be carried out.
7.	Set up a morphological chart for the overall operation, with a view to identifying various alternatives for carrying out the various process steps.
8.	Draw up a number of different layouts at the conceptual level, with a view to identifying different alternatives for the assembly.
9.	Perform a provisional analytical study of each of the concepts, based on various criteria such as achievable cycle times, quality, shop floor area, and flexibility.
10.	Draw up a decision matrix to select the most suitable concept.
11.	Carry out a process failure modes and effects analysis (PFMEA), with a view to identifying and addressing failure mechanisms.
12.	Perform a safety analysis, with a view to identifying and addressing production hazards.
13.	Perform an ergonomic analysis, with a view to optimizing the production system with respect to interactions with human workers.
14.	Carry out a new capacity analysis for the new system, with a view to quantifying the achievable capabilities through automation.
15.	Perform a provisional return on investment analysis, with a view to quantifying provisionally the projected savings and break even times upon implementation of the new system.

Table 2. The procedural guidelines

In this case study, a preliminary analysis based on product design strongly indicated that the parts fell into three natural groupings as shown in Figure 1 above, based on the overall features of their geometries. All of the 20 variants of single gang switches included a socket, a printed circuit board (PCB), and a push button, as shown in Figure 2(a). The PCB included one or more coloured light emitting diodes (LEDs) as required. Some of these variants included one or more of three additional parts: a chrome ring to provide a different aesthetic finish (such as in the switch shown in Figure 2(b)), a light shield for variants that had two different graphics on their front face (to prevent light leakage between graphics), and a jewel

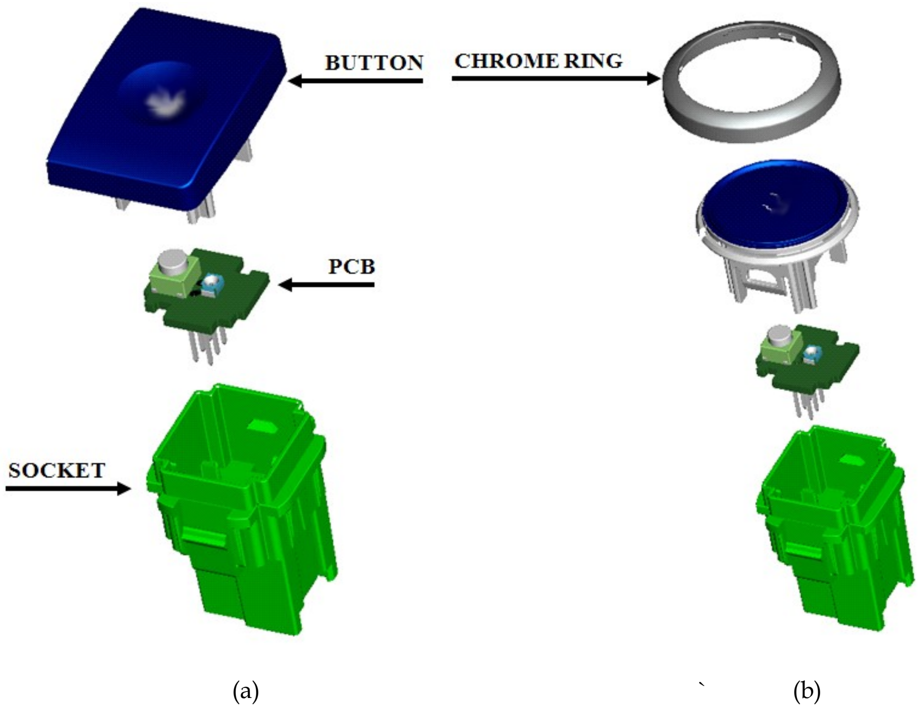


Fig. 2. Single gang switches. (a) a basic variant, (b) variant with a chrome ring

(press fit into the button) to transmit light from the LED to the surface of the button. All of the 11 variants of old three gang switches included a socket, a PCB, one or more sliders, a three unit housing, and three buttons as shown in Figure 3(a). The differences between the variants were defined by the types and combinations of buttons (functional, display, or blank). Functional buttons require a slider, in order to actuate a tact switch on the PCB, and also have a graphic display. Display buttons have only a graphic display (illuminated by an LED on the PCB), and blank buttons have no function or display. The nine variants of new three gang switches have a sleeker design, and use metal clips to attach to the dashboard of the vehicle (see Figure 3(b)).

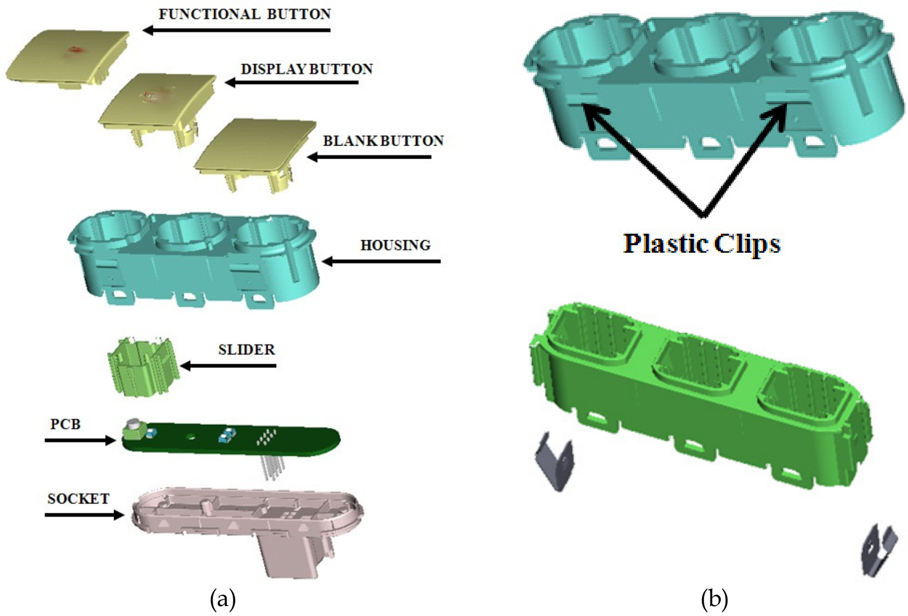


Fig. 3. (a) A variant of old three gang switch, (b) clipping mechanisms (top – old three gang; bottom – new three gang)

4.2 Analysis of the current manufacturing processes

The layouts of the existing production lines are illustrated in Figure 4, with the old three gang switches and some of the single gang switches manufactured in Cell 1, and the new three gang switches and the rest of the single gang switches manufactured in Cell 2. Due to limitations in the end of line testing steps, only one model of switch can be assembled on each cell at any one time, and substantial set-up times are associated with the change over between batches of different models of switch.

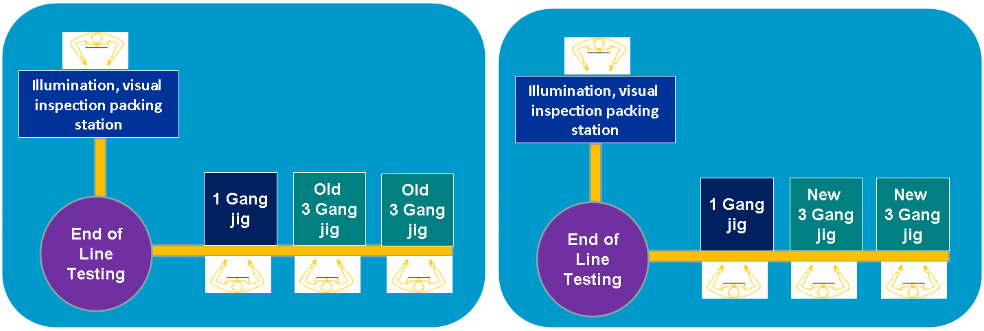


Fig. 4. (a) Layout of Cell 1, (b) layout of Cell 2

The labour intensive nature of the assembly process for the single gang switches is illustrated in Figure 5 and Figure 6. The operator reaches for one socket from the silo, and one PCB from the tray and places the socket over the PCB in cavity (1). If a chrome ring or light shield is required for the switch being assembled, the operator reaches for the chrome

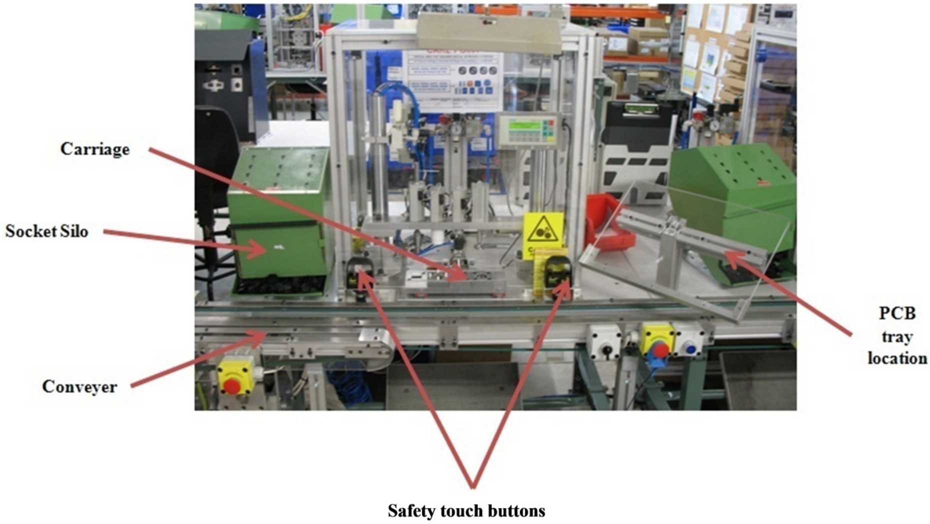


Fig. 5. Assembly jig for single gang switches

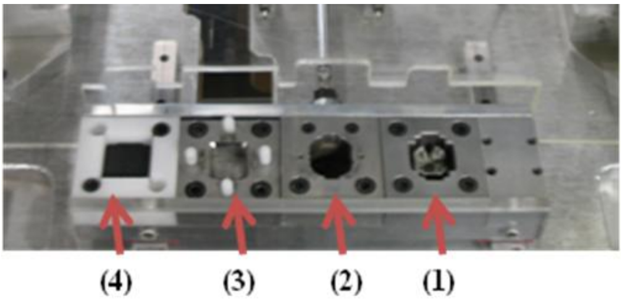


Fig. 6. The four cavities on the single gang switch assembly carriage

ring or light shield and for the button and performs manual alignment. A chrome ring and button sub-assembly is placed in cavity (3), while a light shield and button sub-assembly is placed in cavity (4). The operator then presses the two safety touch buttons placed on the sides of the assembly jig simultaneously. The carriage moves inside the jig and two pneumatic cylinders clip the socket and button sub-assemblies. Subsequently, the carriage moves outside the jig and the operator picks up the socket sub-assembly from cavity (1), rotates it and places it inside cavity (2), whilst placing the button sub-assembly on top of it. The operator also loads the parts for another socket and button sub-assemblies so that during the jig operation, three clipping processes are performed simultaneously. The operator then presses the two safety buttons simultaneously and the carriage moves inside the jig and a pneumatic cylinder clips the two sub-assemblies together. When the carriage moves outside the jig, the operator removes the switch and places it on the conveyor. For all the switches that require no chrome ring or light shield, the operator reaches for the button only when the socket sub-assembly has been placed in cavity (2).

Assembly of the three gang switches is somewhat more complex due to the greater number of parts, however the nature of the operations is similar.

The end of line testing is performed via a fully-automated four-station indexing table. The four stations are (i) a loading station which loads the assembled switch from the conveyor on to the station using a pneumatic pick and place device; (ii) a testing station where the switch is subjected to electrical and force testing, and (for the three gang switches) a barcode label is read; (iii) a camera and laser station where LED illumination intensity and graphic orientation is inspected, and where the customer part number and date code are engraved by laser; and (iv) an unloading station that transfers the switch, using a pick and place device, to a separate conveyor for final inspection, or onto a reject bin.

The final inspection and packaging workstation is fully manual. Here the operator ensures that no scratches, dents or other defects are present on the button's surfaces; checks the integrity of the clipping features; verifies that the terminals are not bent; and ensures that the correct laser marking and date code have been used. Conforming switches are subsequently packed in the respective packaging, whilst non-conforming switches are disposed of in the reject bin.

4.3 Capacity analysis for the current set-up

A capacity analysis was performed in order to quantify the number of switches that can be assembled and tested using the existing production lines. This was done by measuring the time required for every assembly process step of each switch variant, and by analyzing the cycle times and projected production volumes for each individual variant for the four years under consideration. In this respect, the *cycle time* is defined as the time interval for the completion of one complete production unit. In the case study considered, the cycle time was taken to be the longest time from among the three operations performed, i.e. *assembly cycle time* (the time taken to assemble one full switch), *testing cycle time* (the time taken to complete the longest testing step adding the indexing time of the table), and *finishing cycle time* (the time taken to inspect and pack). Equipment availability was assumed to run at 85%, which is due to (i) one product changeover of 15 minutes per shift, resulting in a 3.33% loss; (ii) an allowance of 30 minutes per shift for maintenance activities, including 15 minutes for breakdowns and 15 minutes for planned preventive maintenance, resulting in a 6.67% loss; (iii) personnel related stoppages of 10 minutes per shift resulting in a 2.22% loss; and (iv) process yield running at 97.5%. The number of shifts required to cater for these volumes could thus be calculated using an 85% equipment availability, with 7.5 operating hours per shift, for five days a week and 48 weeks per year.

The results of the capacity analysis are summarized in Table 3. Due to current layout constraints only two product families can be tested in parallel, and this means that the permissible total number of daily shifts is six. As can be seen in the table, during fiscal years 3 and 4, the total output cannot be reached because the number of daily shifts required is not achievable. In addition to this, during fiscal year 4 the number of daily shifts required to achieve the required new three gang switch volumes is 3.11 which is not achievable, since new three gang switches can only be assembled and tested on cell 2. These results pointed out the need of improving the current layouts so as to cater for the required volumes.

	Number of Daily Shifts Required			Total Number of Daily Shifts Required
Fiscal Year	Single Gang	Old Three Gang	New Three Gang	
Year 1	0.30	2.55	0.00	2.85
Year 2	2.28	1.97	0.32	4.57
Year 3	3.38	1.97	1.81	7.16
Year 4	3.75	1.97	3.11	8.83

Table 3. Number of shifts required to cater for the projected volumes using the existing production lines

4.4 Product design specification chart

A PDS chart contains a detailed list of *requirements* that the final product must fulfil, and is drawn up prior to starting the actual design. The aim of the PDS is to encompass all of the required information for a successful solution design and to ensure that the needs of the user are achieved. The PDS also lists a number of *wishes*, which are specifications that are not essential for the success of the project. These wishes however give the project a competitive edge and increase the potential benefits gained through its implementation.

A PDS chart was created for this project, listing all of the specifications that should be taken into account, when designing the required improvements on the switch manufacturing cells. The section of the chart dealing with the *performance* criterion of the production equipment is shown in Table 4. The other criteria that were considered were *target product cost*, *required service life*, *serviceability*, *safety*, *environment*, *size*, *ergonomics*, *materials*, *transportation*, *manufacturing facilities*, *appearance*, *quality and reliability*, *personnel requirements*, *product lifespan*, *documentation*, and *commissioning*.

Examples of design wishes (not shown in Table 4) include the minimization of shop floor space occupied by the equipment, the use of inexpensive (but reliable) materials, and the ability to manufacture the equipment in house.

Specification	Requirement	Need
Performance criterion		
Ability to assemble, test and finish the three switch families.	✓	
Ability to cater for the projected volumes, with an excess capacity of 15%.	✓	
An assembly cycle time for the new three gang switches of less than 10 seconds.	✓	
An assembly cycle time for the single gang switches of less than 8 seconds.	✓	
Ability to test single gang and three gang switches simultaneously.	✓	

Table 4. A section of the PDS chart

4.5 Group technology analysis

In this case study, the preferred criterion for parts classification was found to be that based on product characteristics, since the parts fell into three natural groupings as discussed in section 4.1. The characteristics and variations of each of the three part families were analyzed in detail, with a view to confirming this classification and to prepare for the detailed technical design phase of the project.

An analysis of the constituent parts of the single gang switches produced the following results:

Button – there are different types of button, due to different customer requirements, mainly in terms of graphic design, shape and the type of surface. However all the buttons in the switch family have common guiding and clipping features.

Socket – two types of sockets exist with the main geometric difference being the position of the foolproof feature as shown in Figure 7. A second difference is in the socket colour, where type A is black and type B is grey.

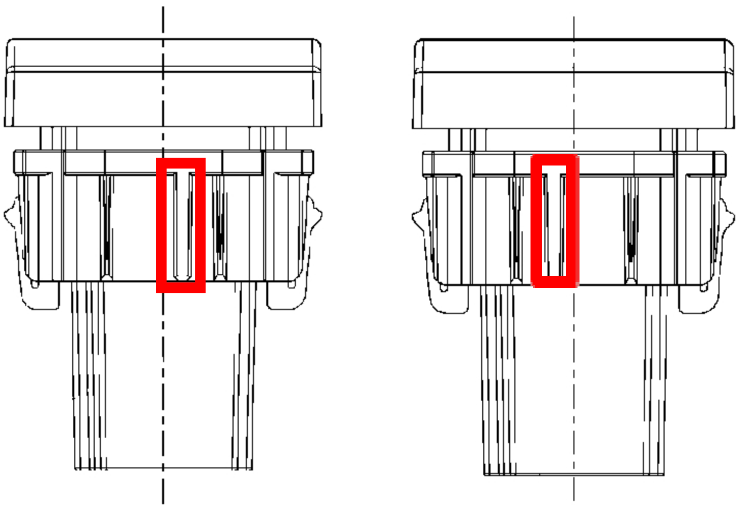
PCB – there are different types of PCB having different profiles and different location of the electrical components.

Chrome Ring – there is only one type of chrome ring.

Jewel – the jewel needs to be aligned with the surface so it must have the same shape as the surface of the button. Two types of jewel exist that correspond to two types of surface.

Light Shield – there is only one type of light shield.

Similar analyses were carried out for the old three gang and new three gang switches. The part variations associated with all three product families are summarized in Table 5.



(a) (b)

Fig. 7. The two types of socket for the single gang switch. (a) Type A, (b) Type B

Single gang switches		Old three gang switches		New three gang switches	
Constituent part	*No. of Variants*	*Constituent part*	*No. of Variants*	*Constituent part*	*No. of Variants*
Button	15	Button set	10	Button set	8
Socket	2	Socket	5	Socket	2
PCB	12	PCB	9	PCB	5
Chrome ring	1	Housing	1	Housing	1
Jewel	2	Slider	1	Metal clip set	1
Light shield	1				

Table 5. Part variations for the three product families

From the results, it can be seen that there are substantial differences between the three sets of products, in terms of the gross geometries and of the constituent parts. In particular, it is noted that a different cavity is required for each of the three families (the geometric differences between the different types of socket within each switch family are minor, and in each case can be catered for by the same cavity). At the same time, the components that constitute each product family allow for ease of automation, since there are only a small number of variations for the parts. The analysis therefore confirmed the classification of the switches into three distinct product families as indicated in Table 5.

4.6 Precedence diagrams

The generalized manufacturing process flow chart for each switch, as extracted from the description given in section 4.2, is illustrated in Figure 8(a), and consists of three major steps. Step 1 involves the assembly of the switch. Step 2 involves the testing of the switch (force, electrical, and illumination testing) and laser marking. Step 3 involves a visual inspection of the switch and final packaging.

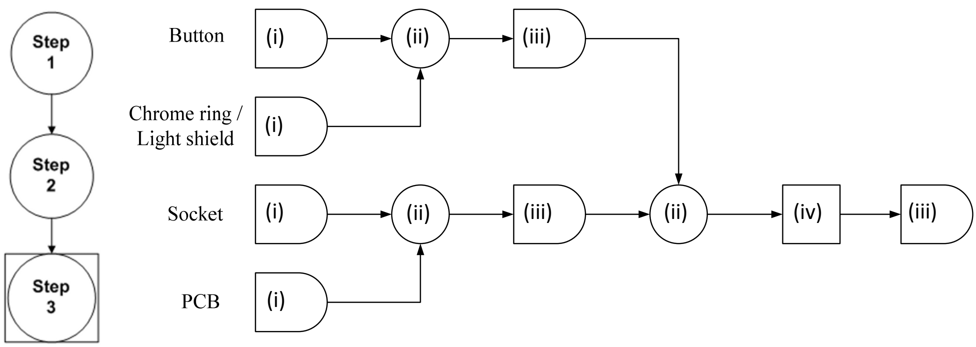


Fig. 8. (a) General process flow chart for each switch, (b) Precedence diagram for the assembly of a single gang switch

The precedence diagram for the assembly process (i.e. for Step 1) for the single gang switch is shown in Figure 8(b). In the figure, the shapes labelled (i) represent temporary storage, the circles labelled (ii) represent the complex operation “bring sub-components together, align, and clip", the shapes labelled (iii) represent delays, and the square labelled (iv) represents a brief visual verification that the clipping has been carried out correctly. For this particular case study it was not possible to simplify these steps any further.

The precedence diagrams for the old three gang and the new three gang switches were extracted in a similar manner. The three diagrams served to guide the development of the morphological chart, and the generation of alternative conceptual solutions, described in the next two sections.

4.7 Morphological chart

A morphological chart is an analytical tool which aims at finding all theoretically conceivable solutions to a problem (Roozenburg & Eekels, 1995). It provides a visual way of capturing the required product functions and of exploring possible different solutions that may exist for each product function. The chart facilitates the presentation of these solutions and provides a framework for considering alternative combinations of the individual function solutions.

The main functions associated with the problem at hand were identified to be the (i) transfer system between stations, (ii) part orienting mechanism, (iii) part feeding mechanism, (iv) handling mechanism, (v) gripping mechanism, (vi) part inspection system, and (vii) part packaging. The morphological chart is shown in Table 6.

Function	Option 1	Option 2	Option 3	Option 4
Transfer System	In-line indexing system	In-Line indexing system with return carriers in the vertical plane	Rotary indexing system	Pallet system
Part Orienting	Vibratory bowl feeder	Magnetic rotary feeder	Machine vision system coupled with a robotic arm	Manual
Part Feeding	Vibrating conveyor	Linear feeder	Horizontal belt conveyor with passive guides	Manual
Handling System	Pneumatic pick and place	Electric pick and place	Robotic arm	Manual
Part Gripping	Vacuum suction	Magnetic gripping	Pneumatic grippers: radial, 3-point and angular	Manual
Part Inspection Systems	Machine vision system	Colour sensor	Human visual inspection	
Packaging System	Robot based system	Customized automation	Manual	

Table 6. Morphological chart for the new manufacturing system. The selected solutions are highlighted.

Solution selection was made on the following bases:

Transfer system – The projected high production volumes necessitate a low indexing time, and if a pallet system is used this would only be achievable by using a very large number of pallets. Thus the manufacturing cost of the system would increase due to the large number of cavities required. A rotary indexing table reduces maintenance interventions, since maintenance requirements are less compared to that of an in-line indexing system.

Part orienting – A vibratory bowl feeder can provide the required output, and is the cheapest and most reliable solution among the four options considered.

Part feeding – High part feeding accuracy is required in order to ensure correct operation of the system and this accuracy can be achieved through the use of linear feeders coupled with vibratory bowl feeders. The use of vibration conveyors or of passive guides cannot achieve the required accuracy. The manual option is expensive.

Handling system – The transportation of the part between two fixed positions can easily be achieved by a pneumatic handling system, which is the cheapest alternative among the four considered.

Part gripping – A system based on vacuum suction could be used, however for its full implementation a number of intricate vacuum heads would need to be designed, thus substantially increasing the cost of the system. A magnetic system cannot be used for all components, since most of the components are made of plastic. A manual system is an expensive option and therefore pneumatic grippers with specifically designed jaws were selected.

Part inspection system – Visual inspection is required at the end of line testing stage (which is already automated, and uses a machine vision system) and at the final inspection station. Due to the complex nature of the final inspection it was determined that this could only be carried out reliably by human operators.

Packaging system – Since the final inspection is manual, the preferred option would be to have the human operator package the completed switch after inspection.

4.8 Concept generation

4.8.1 Overview of proposed concepts

Four different concept layouts were generated to address the problem. The first concept involves automation of the single gang switch assembly, and relocation of all assembly of this switch to Cell 1. Assembly of the new three gang switch would also be automated, and Cell 2 would be dedicated to this process. The second concept involves the retention of the present, labour intensive, assembly processes, but with the incorporation of an additional station to Cell 2 for the assembly of new three gang switches, to meet the projected production volumes. The third concept is a compromise between the approaches of the first and second concepts, and involves the retention of the present, labour intensive, assembly processes for the single gang switch and for the old three gang switch, and the transfer of the single gang switch assembly station from Cell 2 to Cell 1. Cell 2 would be dedicated to the automated production of the new three gang switches as in Concept 1. The fourth concept involves the combination of all production processes into a single cell, and automating the assembly of the single gang and of the new three gang switches. It is noted that due to the fact that the production volume of the old three gang switches is expected to decrease, automation of the assembly process for these switches is not recommended under any of the proposed concepts. The four concepts are presented in greater detail in the following sections.

4.8.2 Concept 1

The proposed layout for Cell 1 under this approach is shown in Figure 9. The cell consists of an indexing table used for the assembly of the single gang switches, two manual jigs used for the assembly of the old three gang switches and a testing indexing table which can test the two different switches simultaneously. Linear conveyors transfer the switches from the assembly stations to the testing station. An operator loads the PCB and button on Station 1 of the loading indexing table. The work carrier of this indexing table will be sub-divided into two sections, one holding the PCB and one the button. These two parts are bought-in parts presented to assembly in painting jigs or trays and therefore automation of the loading function would require a tray changing mechanism and an x-y-z pick and place device. The

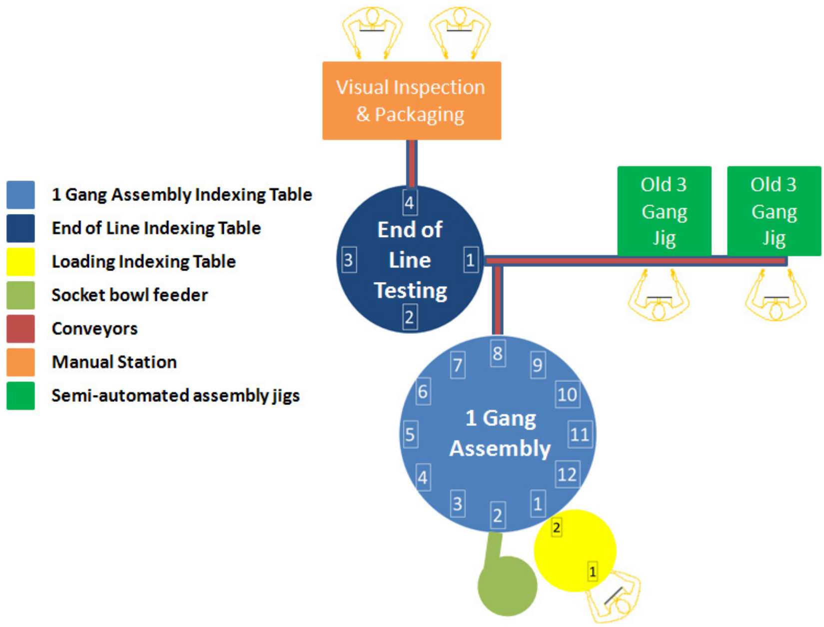


Fig. 9. Proposed Layout for Cell 1 (Concept 1)

initial cost required to create these subsystems would be much higher than the operational cost of one operator who would still be required to attend the machine, and their implementation is therefore not recommended. The parts are then automatically transferred onto a twelve station indexing table. The work carrier on this indexing table will be sub-divided into three sections, namely cavities 1, 2 and 3. The proposed stations for the indexing table are listed in Table 7. The twelve station indexing table was chosen because one is already available at the company, thus reducing the initial cost required. This results in five free stations which can be utilized for future improvements of the layout. The fully assembled single gang and old three gang switches are unloaded onto conveyors which transfer them to the end of line indexing table, based on the current automated system. Table 8 lists the four stations of the fully automated indexing table. The loading station will either pick up one switch type or both switch types, depending upon the switch being available at the conveyor, since old three gang and single gang assemblies would not be synchronized. A new anti-mixing part inspection system, based on machine vision, would need to be incorporated into the end of line testing station, to distinguish between the two switch families.

Station	Description
1	Loading of button onto cavity 3 and PCB onto cavity 1
2	Loading of socket onto PCB
3	Clipping of socket to PCB (sub-assembly 1)
4	Turning of socket and placing onto cavity 2
5	Free station
6	Loading of button onto sub-assembly 1
7	Clipping of button with sub-assembly 1
8	Unloading
9	Free station
10	Free station
11	Free station
12	Free station

Table 7. Single gang switch assembly stations (Concept 1, Cell 1)

Station	Description
1	Loading
2	Force and electrical testing
3	Camera test and laser mark
4	Unloading

Table 8. End of line testing stations (all concepts, all cells)

The proposed layout for Cell 2 consists of a semi-automated twelve station indexing table, used for the assembly of new three gang switches, as shown in Figure 10. The work carrier on this indexing table will be sub-divided into three sections, namely cavities 1, 2 and 3. The socket, clips and housing are oriented via vibratory bowl feeders and automatically loaded on the respective stations. The PCB is manually loaded on a conveyor which is then automatically transferred onto the socket in station 2. Table 9 lists the proposed operations to be performed by each station. The PCB and buttons are manually loaded as in Cell 1. In station 3, the sub-assembled components are unloaded onto a conveyor and the same operator loading the PCB, adds a label to the sub-assembly. Subsequently a second operator adds the three buttons in their corresponding position and places the switch onto a third conveyor. The clipping of the buttons is performed via an automatic clipping and pre-actuation station and finally the switch is loaded onto the testing indexing table. The end of line indexing table is similar to the one on Cell 1.

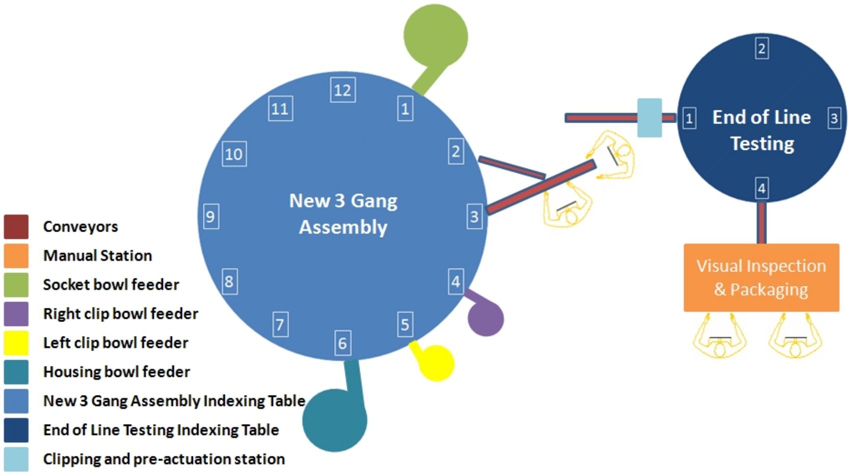


Fig. 10. Proposed layout for Cell 2 (Concept 1)

Station	Description
1	Loading of Socket in cavity 1
2	Loading of PCB onto socket in cavity 1 - (sub-assembly 1)
3	Unloading of sub-assembly 3 onto conveyor from cavity 2
4	Loading of right clip in cavity 3
5	Loading of left clip in cavity 3
6	Loading of housing onto clips in cavity 3
7	Clipping of housing with clips – (sub-assembly 2)
8	Loading of sub-assembly 2 onto sub-assembly 1 in cavity 1
9	Clipping of sub-assembly 2 onto sub-assembly 1 - (sub-assembly 3)
10	Transfer of sub-assembly 3 onto cavity 2
11	Free Station
12	Free Station

Table 9. New three gang switch assembly stations (Concept 1, Cell 2)

4.8.3 Concept 2

The proposed layouts under this approach are shown in Figure 11. This concept entails the incorporation of an additional manual assembly jig to Cell 2 that would be used for the assembly of new three gang switches. This is achieved by modifying the conveyor currently used to transfer the assembled parts from the jigs to the testing indexing table, so as to cater for the additional jig. Changes are also proposed to the testing program of both end of line testers, so as to reduce the testing time required. A new operator is required for the visual inspection and packaging station of Cell 2, so as to cater for all the switches being assembled. Two operators would thus be dedicated to new three gang switches, and one to single gang switches. This concept is a labour intensive concept which however requires less initial investment due to the fact that only minor modifications are required to the existing structure. The projected production volumes can however still be met through this layout.

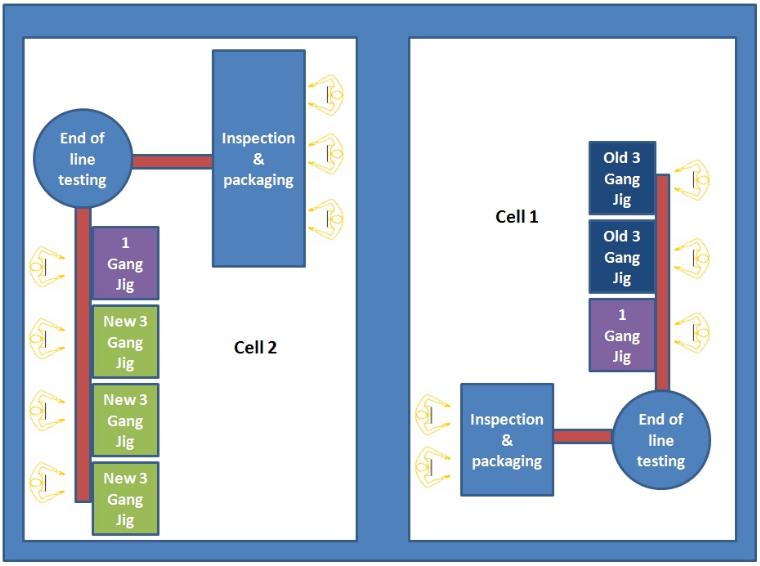


Fig. 11. Proposed layouts for Cell 1 and Cell 2 (Concept 2)

4.8.4 Concept 3

The layouts of the cells under this approach, as described in section 4.8.1, are shown in Figure 12. Modifications are required to the testing program of both cells so as to reduce the testing cycle time.

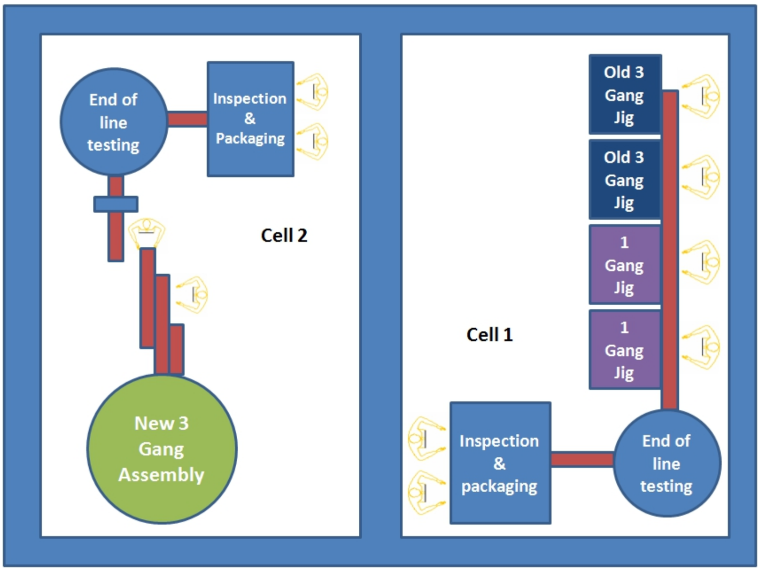


Fig. 12. Proposed layouts for Cell 1 and Cell 2 (Concept 3)

4.8.5 Concept 4

The fourth concept consists of one indexing table that is used for the assembly of both the single gang and the new three gang switches, as shown in Figure 13. A 20-station indexing table is used having work carriers divided into six sections, where three sections (cavities 1, 2 and 3) are used for new three gang switches and the other three (cavities 4, 5 and 6) are used for single gang switches. The main difference between this layout and the one proposed in Concept 1 involves the combination of the two assembly lines. The orienting, feeding, loading and clipping mechanisms are the same as those proposed in Concept 1. The proposed stations are listed in Table 10, where it can be seen that the first twelve stations are dedicated to the new three gang switches and the remaining eight stations to the single gang switches. There would be a total of three free stations. The end of line testing systems are the same as those proposed in Concept 1.

Switch	Station	Description
New Three Gang Switch Assembly	1	Loading of socket in cavity 1
	2	Loading of PCB onto socket in cavity 1 - (sub-assembly 1)
	3	Unloading of sub-assembly 1 onto conveyor
	4	Loading of right clip in cavity 3
	5	Loading of left clip in cavity 3
	6	Loading of housing onto clips in cavity 3
	7	Clipping of housing with clips – (sub-assembly 2)
	8	Loading of sub-assembly 2 onto sub-assembly 1 in cavity 1
	9	Clipping of sub-assembly 2 onto sub-assembly 1 - (sub-assembly 3)
	10	Transfer of sub-assembly 3 onto cavity 2
	11	Free station
	12	Free station
Single Gang Switch Assembly	13	Loading of button onto cavity 6 and PCB onto cavity 4
	14	Loading of socket onto PCB in cavity 4
	15	Clipping of socket to PCB (sub-assembly 4)
	16	Turning of sub-assembly 4 and placing onto cavity 5
	17	Free station
	18	Loading of button onto sub-assembly 4
	19	Clipping of button with sub-assembly 4
	20	Unloading

Table 10. Assembly stations for the single gang and new three gang switches (Concept 4)

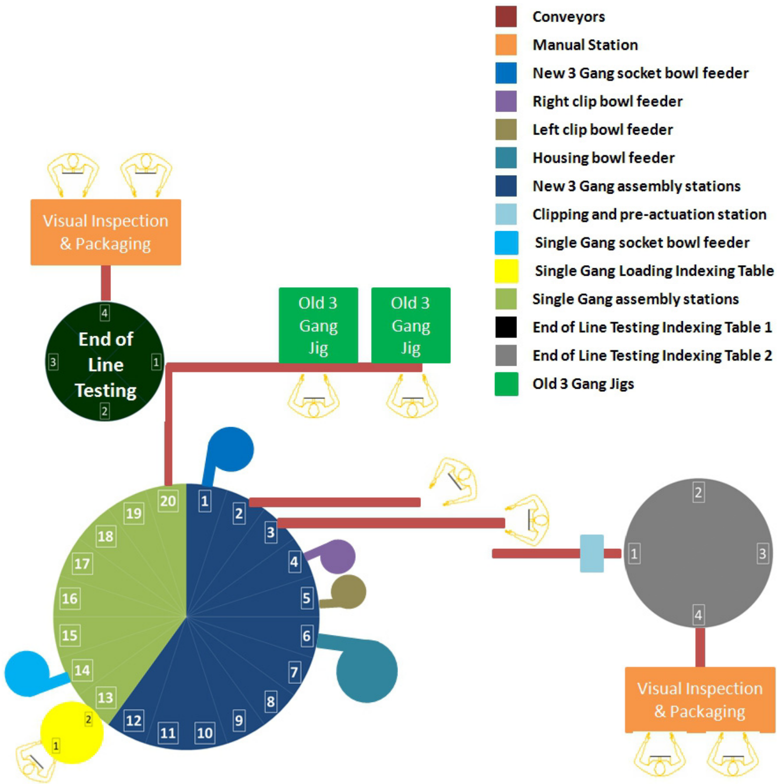


Fig. 13. Proposed layout for combined assembly in a single cell (Concept 4)

4.9 Provisional analytical studies

4.9.1 Overview of analysis

The proposed concepts were analyzed with respect to a number of parameters, namely *cycle time, initial investment cost, labour requirements, line balancing, final product quality, shop floor area consumed, lead time to manufacturing the equipment, maintenance requirements, knowledge transfer availability*, and *flexibility*.

4.9.2 Cycle time

The production of a switch consists of three main operations, namely assembly, testing and finishing. The production cycle time corresponds to the longest cycle time from among these three operations. In order to derive the individual cycle time of each operation, the expected duration of every manufacturing step for each of the proposed concepts was estimated, and the operation cycle time is then given by the longest duration from among its individual stations, taking into account also the indexing time where applicable. It was found that the bottleneck of the production line for all four concepts is the assembly operation. The results of the analysis are summarized in Table 11.

	Single gang	Old three gang	New three gang
Concept 1	6.0s	15.0s	6.0s
Concept 2	8.5s	15.0s	10.8s
Concept 3	8.5s	15.0s	6.0s
Concept 4	6.0s	15.0s	6.0s

Table 11. Production cycle times for each switch family under each concept

4.9.3 Initial investment costs

The initial costs associated with each concept were estimated. It was found that Concept 4 would be the most expensive to implement, followed by Concept 1, Concept 3, and Concept 2 respectively.

4.9.4 Labour requirements

The total number of human operators required for each conceptual approach was determined. Based on the descriptions given in section 4.8, Concept 1 would require nine operators, Concept 2 would require twelve operators, Concept 3 would require ten operators, and Concept 4 would require nine operators.

4.9.5 Line balancing

The *balance efficiency* of a production line is a measure of the time used for productive work at each station as compared to the total available time (e.g. Groover, 2001). In a perfectly balanced line (i.e. 100% balance efficiency) the durations of the jobs carried out at each individual station (be they manual or automated) would be exactly equal to each other (and equal to the cycle time), and there would be no idle time at any station. In practice this ideal situation is very difficult to obtain since it is unlikely that the total work required can be broken down into discrete steps of exactly the same duration, however it remains important to strive for as high an efficiency as possible. The achievable line balance efficiencies for the four concepts described in section 4.8 are given in Table 12.

	Single gang	New three gang
Concept 1	63%	68%
Concept 2	60%	59%
Concept 3	60%	68%
Concept 4	68%	

Table 12. Balance efficiencies for the single gang and new three gang switch families under each concept

4.9.6 Final product quality

The increase in production rate should not be achieved at the expense of reduced product quality and therefore all concepts considered were developed with great concern towards maintaining high quality standards. Thus for example, the proposed automatic handling of buttons would be performed without any part coming into contact with the surface of the button. Given that a well designed automation system is often capable of greater consistency than a system based on human operators, it would be expected that product quality would increase through greater use of automation.

4.9.7 Shop floor area consumed

Due to limitations in the shop floor area, the space consumed by the manufacturing lines would need to be minimized. It was estimated that approximately 55 m^2 of floor area would be consumed by the cells under Concept 1, 50 m^2 under Concept 2, 50 m^2 under Concept 3, and 90 m^2 under Concept 4.

4.9.8 Lead time to manufacturing the equipment

The development of a new production line involves the mechanical and electrical systems design, manufacture of the required components, wiring and programming of the system, and assembly, tuning, and testing of the system. It was estimated that for Concept 1 the total lead time would be approximately 28 weeks, for Concept 2 it would be 2.5 weeks, for Concept 3 it would be 23 weeks, and for Concept 4 it would be 29 weeks.

4.9.9 Maintenance requirements

The selected system would require both preventive and corrective maintenance tasks in order to function correctly over a period of time. The time required to perform such maintenance translates into lost production time and therefore the maintenance requirements of the developed concept need to be minimized so as to maximise productivity and efficiency. The maintenance requirements increase with the number of mechanical, pneumatic and electrical components in the system. Thus Concept 1 and Concept 4, which contain indexing tables, automated stations and vibratory bowl feeders would have higher maintenance requirements than would Concept 2 which contains only mechanized jigs.

4.9.10 Knowledge transfer availability

In today's competitive market, cost reduction and shorter lead time to market are very important. Knowledge transfer aims at achieving these goals, through sharing of the knowledge learnt from previous projects, especially in terms of technologies and procedures. The four conceptual solutions are based on layout styles that are already widely applied at the company and therefore personnel are already experienced with similar equipment. This results in a reduction of the lead time to implement the concept and an improvement in operation and troubleshooting efficiency.

4.9.11 Flexibility

The number of distinct members of the three switch families is expected to increase in the next few years. All four concepts that have been generated allow for these expected new variations, however future customer requirements are difficult to forecast with precision. The *automation* concepts that have been proposed are based on fixed automation systems and thus would not allow for major variations in switch design. An increased flexibility is however supplied by the *manual* assembly jigs since in the associated concepts, potential future changes to the switch design can be more easily catered for by the human operators.

4.10 The decision matrix

Concept selection was based on a decision matrix. The ten criteria discussed in section 4.9 were ranked in order of importance, in consultation with experienced company personnel, and were subsequently given a weighting ranging from 10 for the most important criterion, to 1 for the least important. Each concept was then assigned an individual score for each criterion, based on the analysis of section 4.9. The total score for each concept was then obtained from the weighted sum of the individual scores. The complete decision matrix is shown in Table 13. Based on this result, the selected solution was based on Concept 1.

Selection criterion	Weighting	Concept 1	Concept 2	Concept 3	Concept 4
Labour requirements	**10**	7	1	3	7
Cycle time	**9**	7	1	3	7
Final product quality	**8**	7	1	3	7
Initial investment cost	**7**	3	7	5	1
Line balancing	**6**	5	1	3	7
Knowledge transfer availability	**5**	3	3	3	3
Flexibility	**4**	1	5	3	1
Lead time to manufacture	**3**	3	7	5	1
Shop floor area consumed	**2**	5	7	7	1
Maintenance requirements	**1**	3	7	5	3
TOTAL		281	159	195	265

Table 13. The decision matrix

4.11 Process failure modes and effects analysis

A process failure modes and effects analysis (PFMEA) is a detailed analysis of the errors and malfunctions that can occur during an engineering process, including assessment of the severity, probability of occurrence, and effects of the potential malfunctions, with a view to improve the process design and reliability. An extensive PFMEA was carried out on the selected concept, searching for and assessing various potential failure modes at every station of both production cells. In addition to the various specific process design provisions that were made to address each failure mode that was identified and evaluated through the PFMEA, a number of general conclusions could be drawn from the qualitative and quantitative results of the exercise. Firstly, it was noted that various mistakes can be made by the human operators at the manual stations, and that these mistakes can be minimized by providing clear and concise working instructions to the operators. In this respect all necessary training must also be given. Secondly, it was noted that malfunction of the grippers and air supply, and errors in alignment and settings, can have significant but avoidable detrimental effects on the production process. During the PFMEA a high severity rating was assigned to all of the pick and place operations, to motivate special attention to all associated production line components during commissioning. These ratings would later need to be revised so as to reflect better the final conditions of the line. Thirdly, due to the fact that numerous variants exist for the parts being assembled, the risk of product misidentification and mixing is high. Therefore inspection tests need to be performed in order to detect this failure mode, and this can be achieved through the addition of colour sensors on the linear feeder, a camera inspection on the end of line testing, and an automatic bar code scanner. Fourthly, it is noted that as a final measure, all potential failure modes can be detected by the end of line tests being performed, thus ensuring high reliability of the final product being delivered to the customer.

4.12 Safety analysis

In order to ensure that all safety considerations are integrated within the project as early as possible in the design process, a safety analysis was performed on the selected concept. The analysis followed the five step approach recommended by Bahr (1997) – Step 1: Define the system; Step 2: Identify the hazards; Step 3: Evaluate the hazards; Step 4: Resolve the hazards; and Step 5: Carry out follow-up activity. The system was defined (Step 1) to encompass the two production lines (Cell 1 and Cell 2). The results of Step 2 through Step 4 of the safety analysis are summarized in Table 14. Step 5 can be realized through continued regular checks to ensure: effectiveness of all safety modules; correct functionality of all emergency stop buttons; presence of all protective covers and that all covers are tightly fixed; no cutting edges have been created by wear and tear of the machine; presence of all required grounding systems; and effectiveness of the extraction system and regular filter replacement.

4.13 Ergonomic analysis

In order to improve worker interaction with the system being operated, ergonomic principles were applied to the system design, so as to accommodate human needs. This improves operator performance and well-being, resulting in an increase in overall system performance and efficiency. The analysis has been performed on the manual workstations to ensure that the most ergonomic design is chosen. The ergonomic design specifications are based on the recommendations in Kanawaty (1992) and Wojcikiewicz (2003).

The height of the seated workbench is to be set at approximately 0.72 m so as to ensure that the worker's arms are below the shoulders. The leg clearance should be approximately 0.4 m at knee level and 0.6 m for the feet, without any obstructions such as drawers, between the legs. Height adjustable chairs are to be utilised for the all manned workstations, so as to ensure that the back and neck are not inclined more than 30°. A foot rest should also be available for operators if required. All silos and trays containing assembly parts should be placed within the maximum reach area of the operator, whereas the cavity should be placed within the optimum reach area. Movement of the eyes should be minimized since it takes approximately three seconds for the eyes to rotate and re-focus. Therefore the buttons and the work-piece should be placed within the 15° view angle, on either side of the centreline, since this angle requires no eye movement to allow for the grabbing of the parts. Part silos and the label printer should be placed within the 35° view angle. Correct lighting should be available since this helps to reduce errors and thus improve productivity. The light intensity requirement for the operations to be performed in this case study should be about 500 lux, where one lux is given by the illumination of a surface placed one meter away from a single candle. The light should be uniformly distributed so as to avoid pronounced shadows and excessive contrasts.

Category	Hazard description	Potential causal factors	Sev.	Occ.	Hazard Resolution
Mechanical	Crushing of body part	Unexpected movement of pneumatic cylinders	II	A	Safety guards with interlocks
		Unexpected movement of electric motors on testing station	II	A	
	Operator cuts a body part	Sharp Edges on equipment	III	B	Chamfers and edge deburring
	Operator catches a body part in a pinch point	Pulleys controlling conveyor movement	II	B	Protective covers for all pulleys
	Impact	Unexpected movement of pneumatic cylinders	II	B	Safety guards with interlocks
		Unexpected movement of indexing table	II	B	
	Wrap Points	Entanglement of clothing and accessories with conveyor	III	C	Protective covers for all pulleys; emergency stops
Electrical / Electronic	Energized equipment resulting in electric shock	Improper electrical connections and wiring	I	C	Include fuses, circuit breakers, and electrical grounding; use electrical safety checklist with double-checking; enclose wiring in control box; emergency stop switches
		Poor insulation	I	D	
		Insufficient grounding	I	D	
		Inadvertent activation	I	B	

Severity key: I-Catastrophic; II-Critical; III-Marginal; IV-Negligible.
Occurrence key: A-Frequent; B-Probable; C-Occasional; D-Remote; E-Improbable

Table 14. (first part) Safety analysis: identification, evaluation, and resolution of hazard

Category	Hazard description	Potential causal factors	Sev.	Occ.	Hazard Resolution
Noise / Vibration	Permanent damage to hearing	Environmental sound level exceeds 80dBA	II	B	Pneumatic cylinders equipped with silencers.
	Personnel fatigue	Excessive vibrations to operator's workstation	III	A	No vibratory or linear feeders placed in proximity to operators' workstations.
Lasers	Eye exposure	Collimated beam direct from the laser head into the operator's eyes	II	C	Laser systems enclosed by safety guards with interlocks
	Burning of operator hands	Collimated beam direct from the laser head over the operator's hand	II	C	
	Operator inhales toxic fumes	Toxic fumes arising from burning of plastics by laser marking.	III	A	Fume extraction system

Severity key: I-Catastrophic; II-Critical; III-Marginal; IV-Negligible.
Occurrence key: A-Frequent; B-Probable; C-Occasional; D-Remote; E-Improbable

Table 14. (continued) Safety analysis: identification, evaluation, and resolution of hazards

4.14 New capacity analysis

A detailed capacity analysis was carried out on the proposed production system, based on the assumptions made in section 4.3. The results of this analysis are summarized in Table 15. The total required output can be reached easily using the proposed system, and even in the most demanding year (Year 4) there is a substantial reserve capacity.

	Number of Daily Shifts Required			**Total Number of Daily Shifts Required**
Fiscal Year	**Single Gang**	**Old Three Gang**	**New Three Gang**	
Year 1	0.15	2.55	0.00	2.70
Year 2	1.02	1.97	0.16	3.15
Year 3	1.54	1.97	0.94	4.45
Year 4	1.67	1.97	1.63	5.27

Table 15. Number of shifts required to cater for the projected volumes using the proposed production lines

4.15 Provisional return on investment analysis

In order to estimate the financial benefits that would be gained by the company upon the implementation of the proposed layouts, a provisional return on investment analysis was carried out. The operational cost savings were calculated by comparing labour costs under the present and the proposed layouts. The labour seconds required to manufacture each switch was calculated by multiplying the cycle time by the number of operators required to operate the cell. These calculations indicate substantial cost savings over the four year period, with return on the initial investment achieved in less than three years.

5. Conceptual drawings

While not included among the more critical procedural guidelines proposed in section 3 above, the generation of three-dimensional renditions of the conceptual design of a system helps the design team visualize the overall concept and may aid in the optimization of the spatial layout.

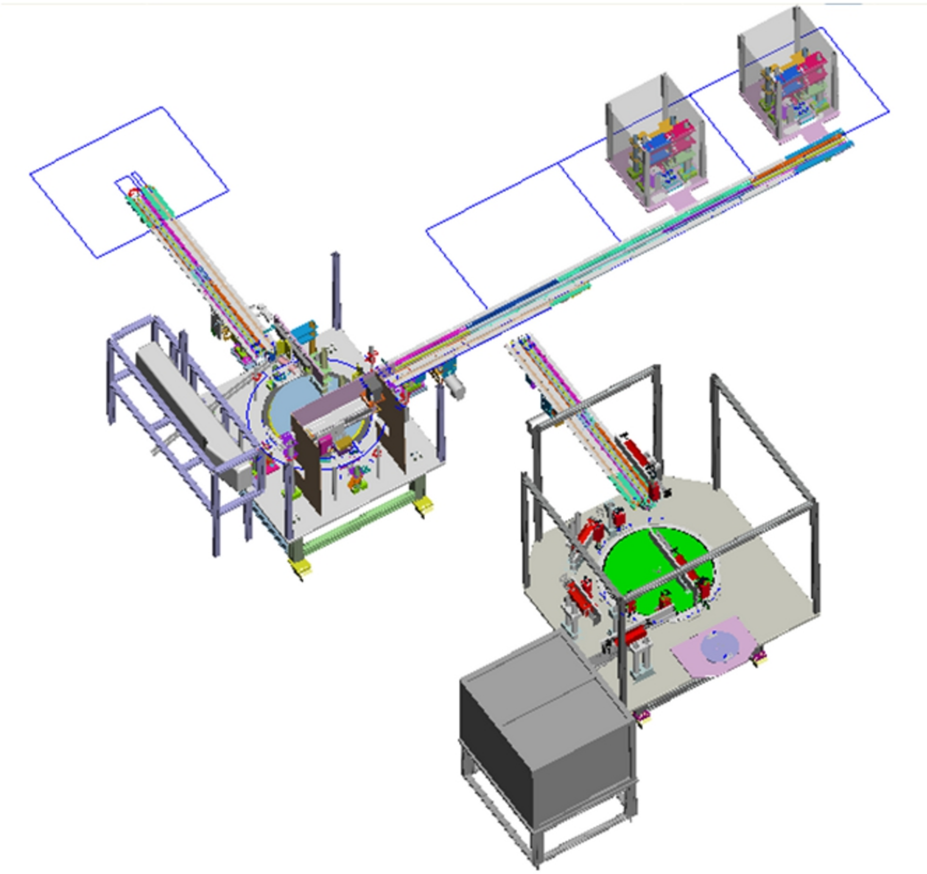


Fig. 14. A 3-D rendition of the proposed layout for Cell 1

A three-dimensional conceptual drawing of the proposed Cell 1 for this case study is given in Figure 14, and shows the two mechanized old three gang switch assembly jigs, the indexing table for single gang switch assembly, and the end of line testing module. The drawing for the proposed Cell 2 is given in Figure 15, and shows the indexing module for the assembly of the new three gang switches and the end of line testing module. These drawings were generated using Pro/ENGINEER (PTC, 2008).

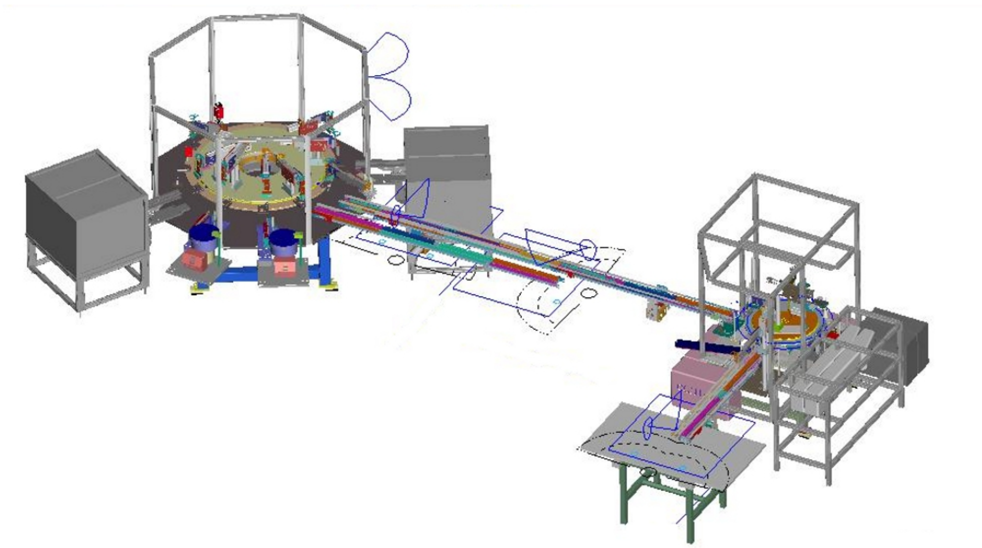


Fig. 15. A 3-D rendition of the proposed layout for Cell 2

6. Conclusion

The procedural guidelines that have been presented in this work contribute an important planning and implementation approach for the development of a conceptual design for a new manufacturing system, when migrating from manual to automated assembly of a part family of products. The novelty in the approach presented here is in the fusion of the conventional guidelines for the development of production automation systems, with a product design approach to the manufacturing system. The detailed case study that is presented in this work serves to demonstrate the application of the guidelines, and will serve as a useful reference tool for future projects of this nature. Future research in this area can include an extension of this approach to the embodiment and detailed design stages of the production system development.

In the case study it is shown that the new automated manufacturing system will result in a cycle time reduction of six seconds for the single gang switches, and of nine seconds for the new three gang switches. This will result in a corresponding increase in production capacity, thus also improving the flexibility of the company since it will be able to react to new customer orders more quickly. The new layout also results in a reduction in the manufacturing lead time, allowing the forecasted customer requirements to be catered for with over 40% excess capacity. The initial investment that is required is justified, since a significant reduction in labour costs is experienced, resulting in a return on investment of less than three years.

7. References

Andreasen, M.M. & Hein, L. (1987). *Integrated Product Development*, IFS (Publications), ISBN 0948507217, Bedford, UK.

Asfahl, C.R. (1992). *Robots and Manufacturing Automation* (Second Edition), John Wiley & Sons, Inc., ISBN 0471553913, NY, USA.

Bahr, N.J. (1997) *System Safety Engineering and Risk Assessment: A Practical Approach*, CRC Press, ISBN 1560324163, London, UK.

Baines, T. (2004). An integrated process for forming manufacturing technology acquisition decisions. *International Journal of Operations & Production Management*, Vol. 24, No. 5, pp. 447–467, ISSN 0144-3577.

Boothroyd, G., Dewhurst, P. & Knight, W. (2001). *Product Design for Manufacture and Assembly* (Second Edition, Revised and Expanded), CRC Press, ISBN 082470584X, NY, USA.

Chan, F.T.S & Abhary, K. (1996). Design and evaluation of automated cellular manufacturing systems with simulation modelling and AHP approach: a case study. *Integrated Manufacturing Systems* Vol. 7, No. 6, pp. 39–52, ISSN 0957-6061.

Dieter, G.E. & Schmidt, L.C. (2009). *Engineering Design* (Fourth Edition), McGraw-Hill, ISBN 0072837039, NY, USA.

Groover, M.P. (2001). *Automation, Production Systems, and Computer-Integrated Manufacturing* (Second Edition), Prentice-Hall, ISBN 0130889784, NJ, USA.

Hyer, N. & Wemmerlov, U. (2002). *Reorganizing the Factory: Competing through Cellular Manufacturing*. Productivity Press, ISBN 1563272288, OR, USA.

Kanawaty, G. (ed.) (1992). *Introduction to Work Study* (Fourth Edition), International Labour Organization, ISBN 9221071081, Geneva, Switzerland.

Kapp, K.M. (1997). The USA Principle: The Key to ERP Implementation Success. *APICS~The Performance Advantage*, Vol. 12, pp. 62–66.

PTC (2008). *Pro/ENGINEER design software*, Parametric Technology Corporation, MA, USA.

Roozenburg, N.F.M. & Eekels, J. (1995). *Product Design: Fundamentals and Methods*, John Wiley & Sons, Inc., ISBN 0471954659, NY, USA.

Säfsten, K., Winroth, M. & Stahre, J. (2007). The content and process of automation strategies. *International Journal of Production Economics*, Vol. 110, pp. 25–38, ISSN 0925-5273.

Ullman, D.G. (1997). *The Mechanical Design Process*. McGraw-Hill, ISBN 0070657564, NY, USA.

Winroth, M. & Säfsten, K. (2008). Automation Strategies - Implications on strategy process from refinement of manufacturing strategy content. *Proceedings of the 19th Annual Conference of the Production and Operations Management Society*, 008-0708, La Jolla, CA, USA, May 2008.

Wojcikiewicz, K. (2003). Seven Key Factors for Ergonomic Workstation Design. *Manufacturing Engineering*, Vol.131, No.1, pp. 45–50.

18

Stochastic Multi-Stage Manufacturing Supply Chain Design Considering Layered Mini-Cellular System Concept

Jing Huang and Gürsel A. Süer*
Ohio University
USA

1. Introduction

Supply chain design attempts to deploy resources to synchronize product flow through multiple tiers of the network and eventually fulfill customers' requirements (Lee, 2000). Traditionally, due to limited sales information and trade barriers, consumers chose local products. But nowadays consumers access the ever opening global market much easier with the help of the globalization. Though overall demand keeps increasing, the market competition becomes fierce as more and more competitors emerge. The highly dynamic market results in difficulties in predicting the demand for companies' products. Furthermore, the demand uncertainties exacerbate the challenge for synchronizing production. Companies start to build more safety stock and hold excess capacity, resulting in decrease of system efficiency. Thus, resource allocation becomes a critical part in the supply chain design. Three levels of resource allocation in supply chain design are summarized in Figure 1. This is only an attempt to present multiple perspectives of resource allocation problems without guaranteeing the coverage of all industrial circumstances. The exceptions and variations in this framework could always be found in the real world applications.

As the market boosts globally, where to manufacture, store and sell various products become the first decision in resource allocation. "Where" could refer to the market such as North American and East Asian (market and production allocation). In this case, managerial decisions related to marketing strategy and business practice are involved. Some auto manufacturers such as Toyota and Honda open local manufacturing facilities in every market they enter. They take the advantage of local resources to increase responsiveness to local market. Furthermore, the risks caused by demand fluctuation are limited to individual markets and does not adversely affect other operations located throughout the world.

"Where" could also mean a specific geographical location for a specific facility in the supply chain network (facility location). By keeping products manufactured or stocked in one central place, company could benefit from economics of scale and increase its efficiency; however, they might reduce responsiveness. Locating resources dispersedly but close to the consumers could improve customer's satisfaction level, but this increases complexity in

* Corresponding Author

coordination of product flow. Though Honda and Toyota open facilities for each local market, their local manufacturing facilities are clustered within a certain area to facilitate just-in-time system thereby reducing leadtime tremendously. On the contrary, Seven-Eleven builds its supply facilities close to its convenience stores in Japan. Each store could efficiently manage its inventory by using the Total Information System, where each order is tracked and recorded by the scanner terminal. Distribution centers receive food from manufacturing plants and directly transfer it to the trucks instead of carrying any inventory for fast food. Thus, Seven-Eleven is able to provide fresh product such as lunch box, sandwiches, bakery and bread and improve its responsiveness (Chopra & Meindl, 2007).

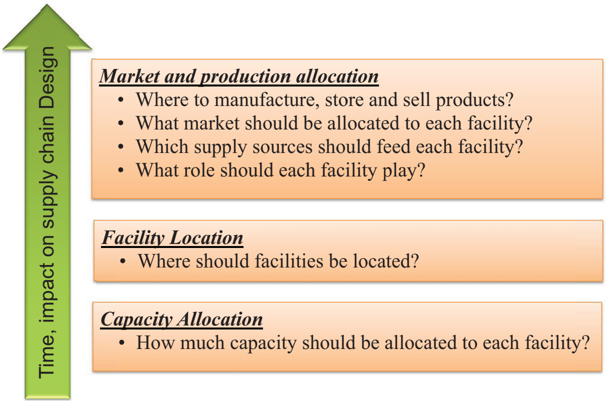


Fig. 1. Three main issues in supply chain design

In the supply chain operational phase, "where" refers to a specific capacity in the shop floor, warehouse, or transportation (capacity allocation). Quantitative models are applied to study machine capacity planning and transportation planning problems where machine or vehicle utilization could be optimized to meet the demand on time under various demand patterns.

Market and production allocation was discussed by Dicken and summarized as: "globally concentrated production", "host market production", "product-specialization for a global or regional market", and "transnational vertical integration" (Dicken, 1992). Globally concentrated production holds production in one base and ships products worldwide. No doubt, production cost could be reduced if production is located in a low-labor cost country; however, the risks of delayed response to the market change arise. Host market production eliminates this risk by dispersing production to each individual market without allowing sales across market boundaries. A better understanding of local customers could be developed and sensitivity to market change could be maintained in each individual market. In the third type of production location strategy, each of the markets manufactures only one product group that is sold to other markets as well. This strategy creates a large-scale and highly specialized manufacturing environment where production cost decreases but transportation cost increases. Transnational vertical integration strategy assigns components or semi-finished products to each of the markets based on manufacturing process, and eventually assemble finished products in one market. It takes advantage of geographical variation of production cost, especially the labor cost. For example, producing low-tech components in developing countries but core components in developed countries could minimize the cost while maintaining product quality. However, additional transportation cost is added if finished products are sold back to the market where components are manufactured.

Managerial decisions related to marketing strategy and business practice are involved in making appropriate market and production allocation decisions. Thus, this chapter assumes that "transnational vertical integration" strategy is adopted, where components are produced in various geographical locations and sold in North American market.

Facility location and capacity allocation decisions are then addressed and solved by a quantitative model in this chapter. Facility location and capacity allocation determines the location, allocation, and production/delivery volume of the flow of goods in a supply chain. Efficient use of all resources to handle supply chain uncertainties is important to facilitate supply chain coordination thereby improving companies' competitiveness (Lee, 2005).

Demand uncertainty is one of the main obstacles in making appropriate decisions. To satisfy demand, supply chain designer tends to reserve extra capacity; however this results in low utilization and therefore higher production cost. Reserving too little capacity results in demand shortage and low responsiveness.

In addition, supply chain design and manufacturing system design are traditionally two sequential steps (Rao & Monhanty, 2003; Cosner, 2008; Schaller, 2008). Roughly estimated capacity requirements are used to locate facility and allocate the production. Various manufacturing systems are then formed within each selected facility. Inaccuracy of estimated capacity requirements results in unsuitable supply chain design and further decreases the manufacturing performance in each facility.

We are proposing a four-phase approach to design and implement the layered mini-cellular system for a multi-stage manufacturing supply chain. In this study, the manufacturing system design and supply chain design are integrated into one scenario. A layered mini-cellular manufacturing system is adopted in the production facility, which is discussed in detail later. Mini-cells are first formed based on probabilistic demand. The mini-cell formation results then serve as inputs to a capacitated plant location model to determine which potential manufacturing plant is selected and how much capacity is allocated to this plant for each manufacturing stage. To continue studying on supply chain operational decisions, a production planning model is proposed to help decide detailed production quantity in each mini-cell for each manufacturing stage in each period.

The remainder of the chapter is organized as follows. The proposed layered mini-cellular system is introduced in section 2. In section 3, solution methodologies are discussed in detail. Experimentation results of proposed system are reported in section 4. In section 5, the performance of layered design is investigated and compared with a classical system. The conclusion is drawn in section 6.

2. Proposed layered mini-cellular system

The manufacturing systems can be categorized into fixed, product, process, and cellular layout in terms of its production layout. Fixed layout is particularly designed for heavy or fragile products such as airplanes, submarines and trains. The product stays in a fixed position, and machines are moved around the product to finish tasks. Product layout is usually adopted by the system with low product diversity but high volume. Each product line is designed for a specific product and performs very efficient production with short throughput time and low work-in-process inventory. On the other hand, process layout is appropriate for a system with high product variety but low volume. Similar processes/machines are grouped and shared by different products, therefore increasing utilization. However, the multidirectional production flow brings challenge in shop floor control. In a cellular layout, products are grouped into families based on the process similarities first and then produced in their own cells. Cellular layout integrates the essentials of both the product layout (product dedication) and process layout (process similarity) into one scenario. It is able to deal with high product variation, in the meanwhile, still maintain a relatively synchronized flow within each cell. The further advantages of cellular system include shorter set-up times, shorter leadtimes, less work-in-process inventory, and fewer defects.

In a classical cellular manufacturing system, each cell is dedicated to only one product family. The cell requirements may vary significantly under a highly fluctuating demand situation, which results in poor utilization of resources. Süer (Süer et al., 2010) brought more flexibility into the cellular system by introducing shared and remainder cells. Assume that the capacity requirements are computed based on the normally distributed demand and processing times (more detailed discussion is in section 3.1) as represented in Table 1. Both expected utilization of X^{th} cell and accumulated demand coverage by X cells are reported in Table 1. For instance, 0.009/0.999 implies that the 4^{th} cell of family 1 is utilized 0.9% of the time, and four cells together could cover demand of family 1 99.9% of the time. To be able to cover the production demand 99.9% of the time, 4+3=7 cells are required for product families 1 and 2. It is clear that the 4^{th} cell of family 1 and the 3^{rd} cell of family 2 are rarely utilized. The demand of family 1 is still covered 99% of the time even without the 4^{th} cell, thus, we may eliminate this cell. However, the 3^{rd} cell of family 2 could not be avoided; otherwise the demand of family 2 will be only covered 86% of the time. In this case, we may group the 3^{rd} cell of family 2 with the 3^{rd} cell of family 1. The capacity requirement is reduced, in the mean time, the desired demand coverage (99%) for each product family is also guaranteed.

Expected Utilization / Demand Coverage	Family 1	Family 2
1st Cell	0.98/0.02	0.99/0.3
2nd Cell	0.9/0.39	0.84/0.86
3rd Cell	0.75/0.99	0.05/0.993
4th Cell	0.009/0.999	

Table 1. An example of capacity requirements

In the layered cellular system, a cell with poor utilization might be combined with another cell. In a shared cell, two product families can be processed. A remainder cell can handle more than two product families. Troubles caused by unstable demand are limited to 'shared' and 'remainder' mini-cells, and demand compensation effect among various product families could help to stabilize demand. The layered system is illustrated in Figure 2, where the production flow is assumed to be unidirectional in each cell. Considering that the chapter mainly studies the supply chain design, the batch production is assumed to simplify the capacity computations.

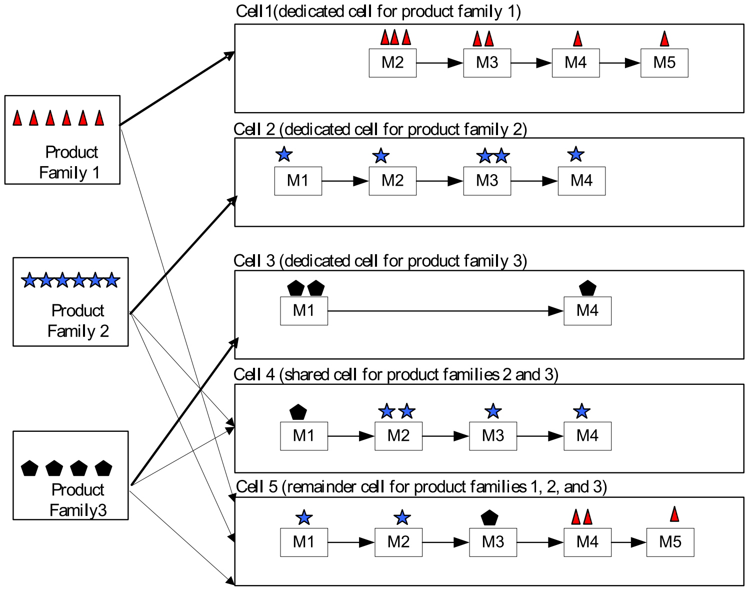


Fig. 2. Three main issues in supply chain design (adopted from Süer et al., 2010)

The layered cellular system proposed by Süer (Süer et al., 2010) assumes a single-stage manufacturing system. In the real-world applications, a multi-stage manufacturing system is usually involved in the manufacturing tier of a supply chain, where each manufacturing facility only performs partial production. In this chapter, 'cell' concept is evolved to a 'mini-cell' concept. A cell performs full package production, while a mini-cell performs operations in a specific manufacturing stage.

3. Stochastic multi-stage manufacturing supply chain design

A four-phase approach is proposed to design and implement the layered mini-cellular system for solving resource management problem in a multi-stage manufacturing system. Table 2 summarizes phases and methodologies used in each phase.

Phase	Objective	Solution Method
1. Expected mini-cell utilization determination	Computing the number of mini-cells and their expected utilizations for each manufacturing stage	Probability Theory
2. Mini-cell formation	Grouping mini-cells for each manufacturing stage	Heuristic Procedure
3. Supply chain network design	Selecting production facilities and allocating mini-cells to the selected facilities	Mixed Integer Linear Programming
4. Simulation	Multi-period production planning	Mixed Integer Linear Programming

Table 2. Summary of stochastic multi-stage manufacturing supply chain design

3.1 Determining expected mini-cell utilization

In this phase, mini-cell requirement for each manufacturing stage is computed based on product demand and processing times. Mean and standard deviation of capacity requirements (in hours) for product family *i* is determined by the processing time (in minutes) of bottleneck operation in manufacturing stage *j* shown in Equations 1 and 2, where N_i is the number of parts in product family *i*, $\mu_{Deamand_{in}}$ is the demand mean of part *n* in product family *i*, $\sigma_{Deamand_{in}}$ is the demand standard deviation of part *n* in product family *i*, and $PTBottleneck_{ijn}$ is the processing time of bottleneck machine for part *n* in product family *i* at stage *j*.

$$\mu_{CRij} = \sum_{n=1}^{N_i} (\mu_{Demand_{in}} \times PTBottleneck_{ijn} / 60) \tag{1}$$

$$\sigma_{CRij} = \sqrt{\sum_{n=1}^{N_i} \left[(\sigma_{Demand_{in}} \times PTBottleneck_{ijn} / 60)^2 \right]} \tag{2}$$

Each mini-cell is assumed to work 40 hours per week. Thus, the probability of covering the demand by a mini-cell is computed as shown in Equation 3, where $X^{th}C_{ij}$ implies the X^{th} mini-cell required by product family *i* in stage *j*.

$$P(X^{th}C_{ij}) = cdf_{Normal}((40 \times X - \mu_{CRij}) / \sigma_{CRij}) \tag{3}$$

The expected utilization of the X^{th} mini-cell for product family *i* in stage *j* is given in Equation 4. $P(NCR_{ij}>X)$ indicates the probability that the number of mini-cells required by product *i* in stage *j* is greater than *X* as given in Equation 5, while $P(X-1<=NCR_{ij}<=X)$ means the probability that mini-cells required is between *X-1* and *X* as given in Equation 6. PU_1 is the utilization of X^{th} mini-cell when mini-cell requirement is greater than *X*, therefore it is fixed as 1. PU_2 is computed as given in Equation 7, where μ_i is the mean demand of product family *i*, and σ_i is the standard deviation of the demand.

$$E(X^{th}C_{ij}) = P(NCR_{ij} > X) \times PU_1 + P(X-1 \le NCR_{ij} \le X) \times PU_2 \tag{4}$$

$$P(NCR_{ij} > X) = 1 - P(X^{th}C_{ij}) \tag{5}$$

$$P(X-1 \le NCR_{ij} \le X) = \begin{cases} P(X^{th}C_{ij}) & X = 1 \\ P(X^{th}C_{ij}) - P((X-1)^{th}C_{ij}) & X \ne 1 \end{cases} \tag{6}$$

$$PU_2 = \int_{40(X-1)}^{40X} \frac{y \times \frac{1}{\sigma_i\sqrt{2\pi}} e^{-(y-\mu_i)^2 \frac{1}{2\sigma_i^2}}}{40 \times P(X-1 \le NCR_{ij} \le X)} dy - (X-1) \tag{7}$$

An example result of mini-cell capacity estimation for manufacturing stage *j* is shown in Table 3. The results imply that 4+4+3=11 mini-cells are required to cover the demand of these three product families. The expected utilization and demand coverage of each mini-cell are also given in the same table. Please note that, we will continue to use this small example to illustrate the procedures that will be discussed in the following sections.

Expected Utilization / Demand Coverage	Family 1	Family 2	Family 3
1st Mini-Cell	0.99/0.002	0.99/0.5	0.98/0.6
2nd Mini-Cell	0.87/0.39	0.79/0.86	0.55/0.91
3rd Mini-Cell	0.32/0.96	0.1/0.98	0.02/0.999
4th Mini-Cell	0.03/0.999	0.003/0.999	

Table 3. An example of mini-Cell utilization and demand coverage

3.2 Grouping mini-cells

In Table 3, obviously, several mini-cells are rarely utilized (e.g. 0.3% utilization of fourth mini-cell for family 2). A heuristic procedure is introduced in this section to reduce the number of required mini-cells by grouping mini-cell segments. The grouping process is implemented based on process similarities among product families. Another important criterion of the grouping process is demand coverage. For example, in Table 3, three mini-cells are able to cover the demand for product family 1 96% of the time, therefore, the fourth mini-cell might not be required.

Figure 3 illustrates the heuristic procedure, where XC_i implies X^{th} mini-cell for product family i, XCU_i is the utilization of this mini-cell, C_j is the newly formed mini-cell j, and LC_j is the leftover utilization for newly formed mini-cell j. Heuristic procedure attempts to form dedicated, shared and remainder mini-cells with the objective of reducing the total number of mini-cell requirements. In the meantime, it prefers to group product families with similar manufacturing operations in order to avoid increasing machine/workforce numbers and operational complexities within a mini-cell. This heuristic procedure is repeated for each manufacturing stage.

An example result of grouping 11 dedicated mini-cells (see Table 3) for manufacturing stage j is shown in Figure 4. After the grouping procedure is applied, four dedicated mini-cells stay, and two dedicated mini-cells are grouped into a shared mini-cell. The other three mini-cells originally dedicated to product families 1, 2, and 3 are grouped into a single remainder mini-cell. Since three mini-cells are able to cover the demand for product family 1 96% of the time, the fourth mini-cell required by product family 1 (noted in the red block) is abandoned during the grouping process. The same procedure is applied to the fourth mini-cell required by product family 2. It is observed that 11 mini-cells cover the demand all the time, and the grouping process reduces the number of mini-cells to six still covering demand 96% of the time.

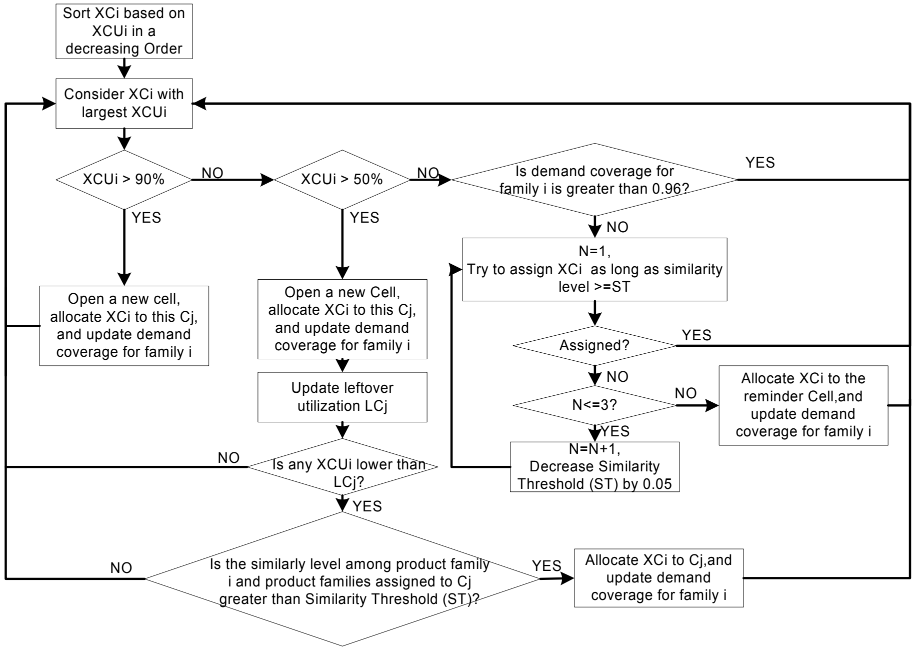


Fig. 3. Flowchart of heuristic procedure

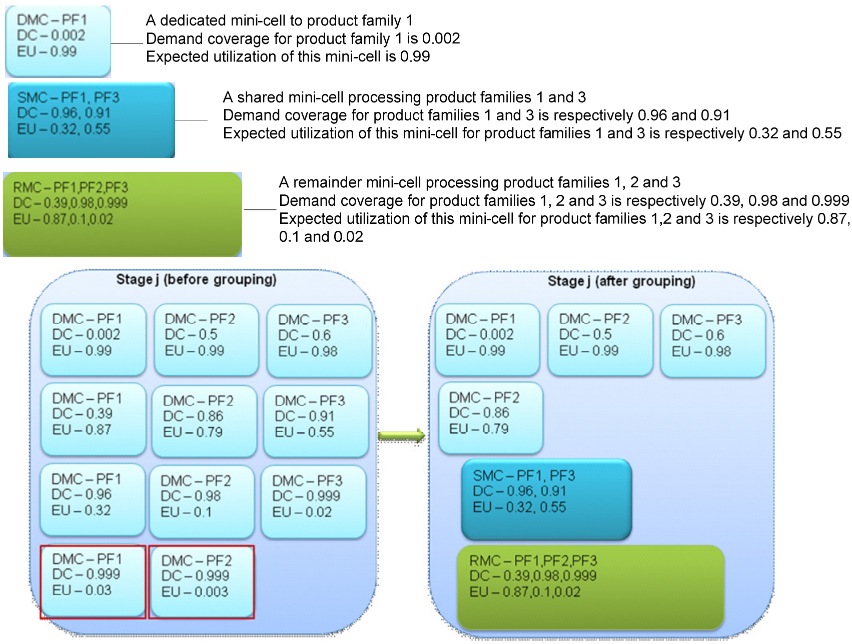


Fig. 4. An example of grouping mini-cells

3.3 Locating facility and allocating mini-cells

A capacitated plant location mathematical model is built to allocate mini-cells to the candidate plants. The objective of this model is to minimize the total cost including production cost, investment cost, and transportation cost as given in Equation 9. Candidate plants are located in different areas with limited capacities, and various production and investment costs. Transportation costs include the costs of transporting products between two consecutive manufacturing stages, and also the costs of shipping products to the market. A single market is assumed. Equation 10 guarantees that the capacity allocated to a plant does not exceed the maximum available capacity. A mini-cell can be assigned to only one facility as presented in Equation 11. Equations 12 and 13 maintain transportation balance, in other words, the quantity of products shipped into a plant should match product quantity shipped out of this plant. Equation 14 enforces investment cost of opening a manufacturing stage in a plant. Parameter EQ_{ik} is roughly estimated as given in Equation 8, where EU_{ik} is the expected utilization of product family in in mini-cell k, and PT_{ij} is the processing time (in minutes) of bottleneck operation in stage j for family i.

$$Q_{ik} = 40 \times 60 \times EU_{ik} / PT_{ij} \tag{8}$$

Indices:

i Product family index
j Manufacturing stage index
k Mini-cell index
m Plant index

Parameters:

I Number of product families
J Number of manufacturing stages
K Number of mini-cells required
M Number of potential plants
NM_k Number of machines/workforce in mini-cell k
U_{jk} 1, if mini-cell k performs operations in manufacturing stage j; 0, otherwise.
EQ_{ik} Estimated quantity of product family i produced in mini-cell k
LOI_{ij} Previous stage index of stage j for product family i. 0 implies stage j is the first stage for family i or family i does not require manufacturing stage j
$MAXC_{jm}$ Available number of mini-cells for stage j in plant m
IC_{jm} Weekly equivalent investment cost for stage j in plant m
PC_{jm} Production cost for stage j in plant m ($/40hour)
UTC Unit transportation cost ($/mile/unit)
D_{mn} Distance from plant n to plant m
DM_m Distance from plant m to market
M Big value

Decision variables:

X_{km} 1, if mini-cell k is allocated to plant m; 0, otherwise

W_{jm} 1, if stage *j* is opened in plant *m*; 0, otherwise
TQ_{ijmn} Transportation quantity of family *i* from plant *n* to plant *m* from stage *j-1* to stage *j*

Objective Function:

$$\min \quad Z = \sum_{j=1}^{J}\sum_{m=1}^{M}(PC_{jm} \times \sum_{k=1}^{K}(NM_k \times U_{jk} \times X_{km}) + IC_{jm} \times W_{jm}) + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{n=1}^{M}\sum_{m=1}^{M}(TQ_{ijmn} \times D_{mn} \times UTC) + \sum_{m=1}^{M}\sum_{i=1}^{I}\sum_{k=1}^{K}(X_{km} \times U_{Jk} \times EQ_{ik} \times DM_m \times UTC) \tag{9}$$

Subject to:

$$\sum_{k=1}^{K}(X_{km} \times U_{jk}) \le MAXC_{jm} \text{ for } j = 1,\cdots,J \ \& \ m = 1,\cdots,M \tag{10}$$

$$\sum_{m=1}^{M} X_{km} = 1 \text{ for } k = 1,\cdots,K \tag{11}$$

$$\left.\sum_{n=1}^{M} TQ_{ijmn} = 0\right|_{LOI_{ij}=0} \quad \left.\sum_{n=1}^{M} TQ_{ijmn} = \sum_{k=1}^{K}(X_{km} \times U_{jk} \times EQ_{ik})\right|_{LOI_{ij}\neq 0} \text{ for } i = 1,\cdots,I \ \& \ j = 1,\cdots,J \ \& \ m = 1,\cdots,M \tag{12}$$

$$\left.\sum_{m=1}^{M} TQ_{ijmn} = 0\right|_{LOI_{ij}=0} \quad \left.\sum_{m=1}^{M} TQ_{ijmn} = \sum_{k=1}^{K}(X_{km} \times U_{LOI_{ij}k} \times EQ_{ik})\right|_{LOI_{ij}\neq 0} \text{ for } i = 1,\cdots,I \ \& \ j = 1,\cdots,J \ \& \ n = 1,\cdots,M \tag{13}$$

$$M \times W_{jm} \ge \sum_{k=1}^{K}(X_{km} \times U_{jk}) \text{ for } j = 1,\cdots,J \ \& \ m = 1,\cdots,M \tag{14}$$

An example of allocation process is demonstrated in Figure 5. The results indicate that four mini-cells including three dedicated mini-cells and one reminder mini-cell are allocated to plant 3. One dedicated mini-cell and one shared mini-cell are to plant 4. Plant 1 is not able to implement any operation of manufacturing stage *j*, thus, there is no mini-cell allocated to this plant. Plant 2 is not chosen either based on various factors such as capacity, production cost, and distances.

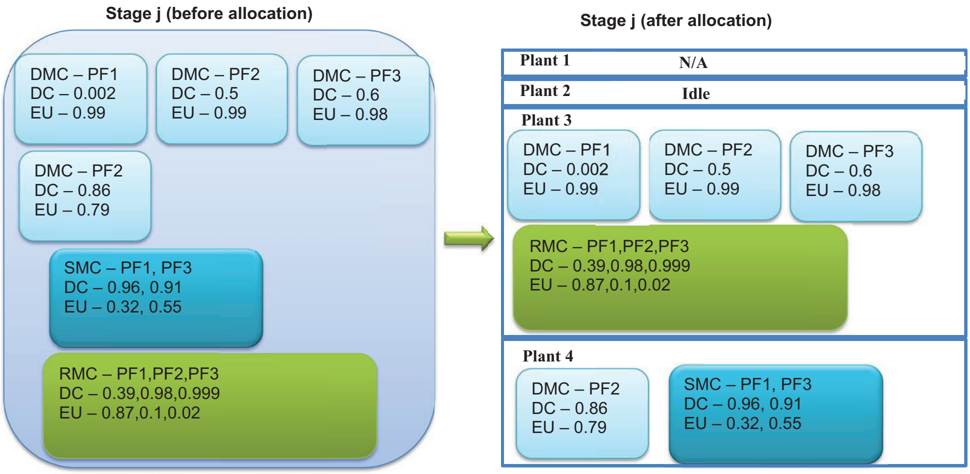


Fig. 5. An example of facility location and mini-cell allocation

3.4 Multi-period production planning

The capacitated plant location model selects plants and determines how many mini-cells should be built in each plant for each manufacturing stage. To continue studying on supply chain operational decisions, a production planning model is proposed to help decide production quantity in each mini-cell for manufacturing stage j in each week. The objective is to minimize demand shortage as given in Equation 15. Equation 16 computes demand shortage for each product family. Equation 17 guarantees that the product family is not assigned to a mini-cell which does not handle that family. Each cell only functions up to 40 hours per week as given in Equation 18. The boundary of decision variable is defined in Equation 19.

Indices:

i Product family index
k Mini-cell index

Parameters:

I Number of product families
K Number of mini-cells required
D_i Demand of product family i for current week
CU_{ik} Expected utilization of family i in mini-cell k
PT_i Processing time (in minutes) of bottleneck operation of family i
M Big value

Decision variables:

Q_{ik} Quantity of product family i produced in mini-cell k
DS_i Demand shortage of product family i

Objective Function:

$$\min \qquad Z = \sum_{i=1}^{I} DS_i \tag{15}$$

Subject to:

$$DS_i = \max\left\{0, D_i - \sum_{k=1}^{K} Q_{ik}\right\} \text{ for } i = 1,\cdots,I \tag{16}$$

$$Q_{ik} \le M \times CU_{ik} \text{ for } i = 1,\cdots,I \;\&\; k = 1,\cdots,K \tag{17}$$

$$\sum_{i=1}^{I} (PT_{ik} \times Q_{ik}) \le 40 \times 60 \text{ for } k = 1,\cdots,K \tag{18}$$

$$Q_{ik} \ge 0 \tag{19}$$

An example of production plan giving detailed production quantity of each mini-cell for manufacturing stage *j* in week *n* is shown in Figure 6. For example, in plant 3, a dedicated mini-cell to product family 1 needs to produce 35 units of product family 1 in week *n*.

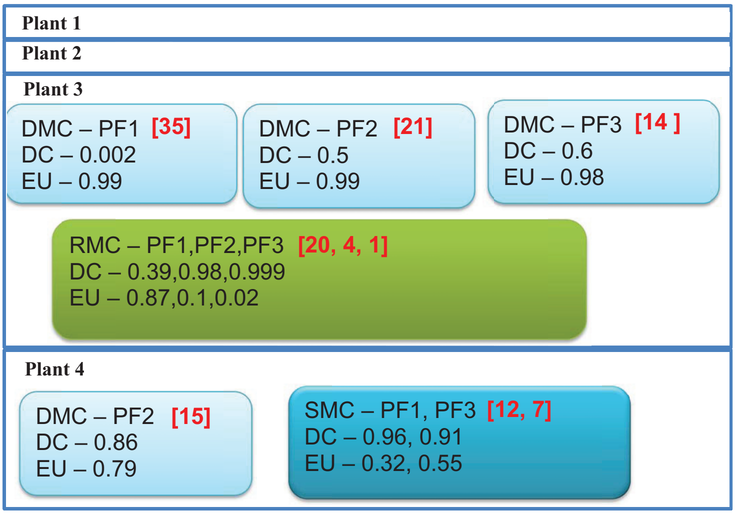


Fig. 6. An example of weekly production planning result

4. The system studied and the preliminary results

An example of supply chain involving a three-stage manufacturing system is studied in this section. This system was originally inspired from a global jewelry manufacturing company. The candidate production plants are mainly located in Caribbean, East Asia, and South East Asia. The jewelry products change along the fashion trend therefore resulting in highly fluctuating demand. It is very challenging to manage capacity to satisfy demand without reserving too much capacity.

The system studied consists of 12 product families with normally distributed weekly demand. The standard deviation is 25% of the average demand. Up to eight operations are required to manufacture a product, and they are grouped into three operation groups since some production facilities are not able to perform some of the operations (e.g. plating operation). All products go through three stages in the same order. In the jewelry manufacturing process, different parts within one product family always require the same operations and processing times are very close, thus, Table 4 shows the processing times and weekly demand for each product family instead of each part. This chapter focuses on supply chain design of manufacturing tier, thus, only one market is assumed.

	Stage	**Manufacturing Operations**	**PF1**	**PF2**	**PF3**	**PF4**	**PF5**	**PF6**	**PF7**	**PF8**	**PF9**	**PF10**	**PF11**	**PF12**
Processing Times (Minutes)	1	Findings (F)		10.6			10.8							11.1
		Casting (C)	10.2		9.7	10.9		13.3	12.6	10.1	10.8	12.5	9.9	
		Tumbling(T)	10.8	11	9.9	12.6	11.2	13.9	13.5	12.3	11.2	13	13.4	12.2
	2	Plating (PL)	9.6	10.9	11.2	9.3	10.1	10.5	9.2	9.2	8.5	12	12.5	11.7
	3	Stone Setting(SS)	9.1	4.6			6.2	4.1					9.3	
		Enameling(E)				5.8	7.5		7.1	7.2		5.6		8.5
		Oven (O)				8.7	6.0		4.9	9.2		8.9		5.5
		Packaging(PA)	12.7	12.8	11.2	11.2	11.4	10.9	10.8	12.7	11.5	12.1	12.3	12.5
Weekly Demand (Units)		mean	1890	795	478	2127	722	520	2582	981	1773	1150	966	474
		SD	472	199	120	532	180	130	646	245	443	288	242	118

Table 4. Processing times and demand

Experimentation of supply chain design on the studied system is implemented by using the proposed four-phase approach. For a three-stage manufacturing system with 12 product families, the mini-cell requirements and expected mini-cell utilizations are computed. Cells consist of multiple machines and equipment. Each machine/equipment requires an operator as well. Table 5 summarizes the number of mini-cells required for each product family as well as the total number of machines/workforce for each stage. The results indicate that a total of

381 mini-cells are required. These 381 mini-cells require a total of 708 machines/equipment and also 708 operators. The number of machines/equipment and operators needed for stages 1, 2, and 3 are computed as 272, 113 and 323, respectively. Due to the space limit, Table 6 only shows partial utilization results for stage 1 for product families which require no more than 6 mini-cells. It is important to notice that many cells are rarely utilized as shown in Table 6 such as 4th, 6th, 6th, and 5th mini-cells for product families 3, 5, 6, and 12, respectively.

	PF	Stage 1	Stage 2	Stage 3	Total Number
Number of Mini-Cells	1	16	14	18	381
	2	7	7	8	
	3	4	4	4	
	4	20	15	18	
	5	6	6	7	
	6	6	5	5	
	7	26	18	21	
	8	9	7	10	
	9	15	12	16	
	10	12	11	11	
	11	10	9	9	
	12	5	5	5	
Number of Machines/Workforce		272	113	323	708

Table 5. Summary of capacity requirement results

Mini-cell	PF3	PF5	PF6	PF12
1st	0.995367	0.999409	0.999131	0.998108
2nd	0.792689	0.98242	0.969661	0.91366
3rd	0.180275	0.835848	0.736627	0.446252
4th	0.003406	0.441633	0.272882	0.050731
5th		0.101697	0.032402	0.000743
6th		0.008137	0.000961	

Table 6. Partial results of mini-cell utilization for stage 1

Heuristic procedure is applied to form dedicated, shared and remainder mini-cells. The results of mini-cell formation are summarized in Table 7. Demand coverage is set to be 0.96 for the heuristic procedure. In other words, product demand will be covered 96% of the time. It is observed that majority of mini-cells are dedicated mini-cells, thus, the operational complexity is limited. By grouping product families into shared and remainder cells, mini-cell requirement is reduced from 381 to 210. Total number of machines/workforce is reduced from 708 to 414.

		Stage 1	Stage 2	Stage 3	Total Number
Number of Mini-Cells	Dedicated	50	40	49	210
	Shared	18	10	17	
	Remainder	8	11	7	
Number of Machines/Workforce		158	61	195	414

Table 7. Summary of mini-cell formation results

There are seven potential production facilities performing operations for different manufacturing stages as shown in Table 8, where 0 implies that the plant doesn't perform operations in this manufacturing stage. For example, plant 4 only performs the plating operation. Distance matrix is given in Table 9. Production costs vary from 32 to 850 representing huge gap of labor costs between developed areas and developing areas.

The math model is solved by ILOG OPL software. The allocation of capacity to production facilities is determined as shown in Figure 7. Plants 1 and 3 are not chosen to perform any operation due to their high production costs. However, production costs are not the only criteria of making decisions. For example, for manufacturing stage 3, production costs in plants 5 and 6 are very low compared with plant 2. But they are not chosen considering the high transportation costs.

	Stage 1	Stage 2	Stage 3
Plant 1	50/800/30000	40/835/30000	50/840/60000
Plant 2	30/600/5000	0	30/680/5000
Plant 3	60/650/10000	50/850/50000	75/720/30000
Plant 4	0	50/520/5000	0
Plant 5	0	0	50/32/25000
Plant 6	50/76/5000	0	50/61/5000
Plant 7	0	30/52/5000	50/72/5000

Table 8. Capacity (in 40 hours)/production cost ($40 hours) /investment cost

Plant	1	2	3	4	5	6	7
1	0	30	50	300	9000	8300	8500
2	30	0	40	260	8960	8280	8530
3	50	40	0	280	9030	8300	8500
4	300	260	280	0	10200	9300	9200
5	9000	8960	9030	10200	0	500	540
6	8300	8280	8300	9300	500	0	70
7	8500	8530	8500	9200	540	70	0
Market	300	350	310	300	7500	5600	5500

Table 9. Distance Matrix

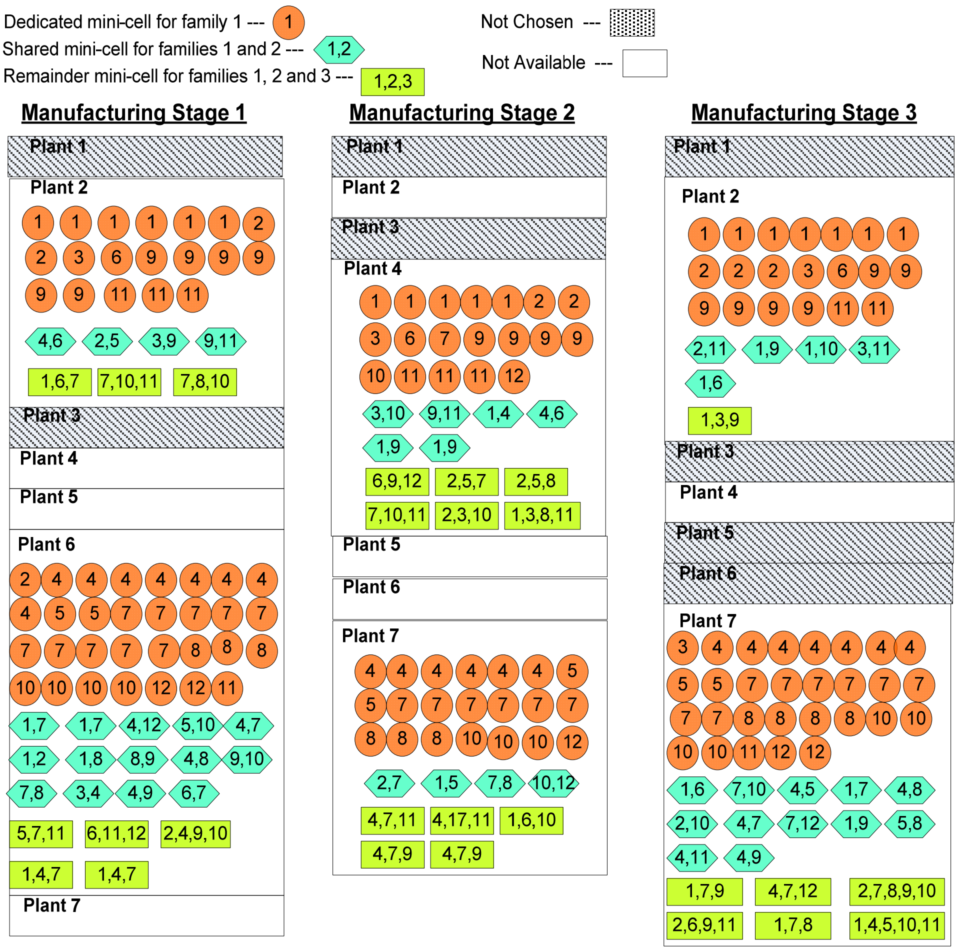


Fig. 7. Capacity allocation and plant location results

For multi-period production planning, the experimentation runs for a year (50 weeks). In each week, demand is randomly generated based on normal distribution. The proposed production model under such demand is solved by using OPL software, the demand shortage for each stage is recorded, and next, it is compared to the results of classical cellular system.

5. Comparison with classical mini-cellular system

In this section, performance of the proposed layered mini-cellular system is compared to that of a classical cellular system. Machine/workforce requirements and demand shortage are used as performance measures to evaluate these two systems. The proposed capacitated plant location model is also capable of solving resource management problems when classical cellular design is adopted in the production facility. However, the model parameters such as NM_k, U_{jk}, and EQ_{ik} are computed differently, since each mini-cell is dedicated to a single product family in the classical cellular system. The number of mini-cells required by product family *i* in stage *j* is computed in Equation 20, where μ_i is demand mean of family *i*, and PT_{ij} is the processing time (in minutes) of bottleneck operation for family *i* in stage *j*. RCU is the reserved cell utilization in order to handle high demand situation. Obviously, the value of RCU affects both performance measures: machine/workforce requirements and demand shortage. Reserving too little capacity leads to high demand shortage; while reserving too much capacity results in redundant machine/workforce therefore increasing production costs. In this section, the preliminary experimentation is carried out to illustrate the procedure, thus, RCU is set to be 10%. In the future, experimentation with various levels of RCU will be implemented and results will be studied.

$$NC_{ij} = \mu_i \times PT_{ij} / (40 \times 60 \times (1 - RCU)) \quad (20)$$

The capacity requirement is computed for each product family at each stage, and the results are summarized in Table 10. It is observed that classical system requires 176, 73, and 204 machines/workforce for manufacturing stages 1, 2, and 3, respectively; while layered system only requires 158, 61, and 195 machines/workforce.

	PF	Stage 1	Stage 2	Stage 3	Total Number
Number of Mini-Cells	1	9	12	10	245
	2	5	5	5	
	3	3	3	3	
	4	10	12	13	
	5	4	4	4	
	6	3	3	4	
	7	11	13	17	
	8	5	6	6	
	9	7	10	10	
	10	7	7	7	
	11	6	6	6	
	12	3	3	3	
Number of Machines/Workforce		176	73	204	453

Table 10. Summary of capacity requirement results of classical design

The performance of the proposed layered mini-cellular system in terms of handling fluctuating demand is investigated in this section. Fifty demand sets are randomly generated based on normal distribution given in Table 4. The service level for each period is computed based on the demand supplied from the facility network divided by total demand. The results are obtained by using both the layered design and classical cellular design with 10% reserved cell utilization (as shown in Figure 8). For manufacturing stage 1, it can be observed that layered mini-cellular system leads to a high service level (>90%) most of the time. There are only six out of 50 periods when the service level is below 90%. Under most conditions, layered system leads to a higher service level. There are only seven exceptions where classical design leads to a higher service level. The similar pattern could be also observed for manufacturing stages 2 and 3. It is clearly observed that, compared to the classical system, the layered system model requires less number of mini-cells and machines/workforce while still dealing with high demand fluctuation more effectively as evidenced by higher service levels.

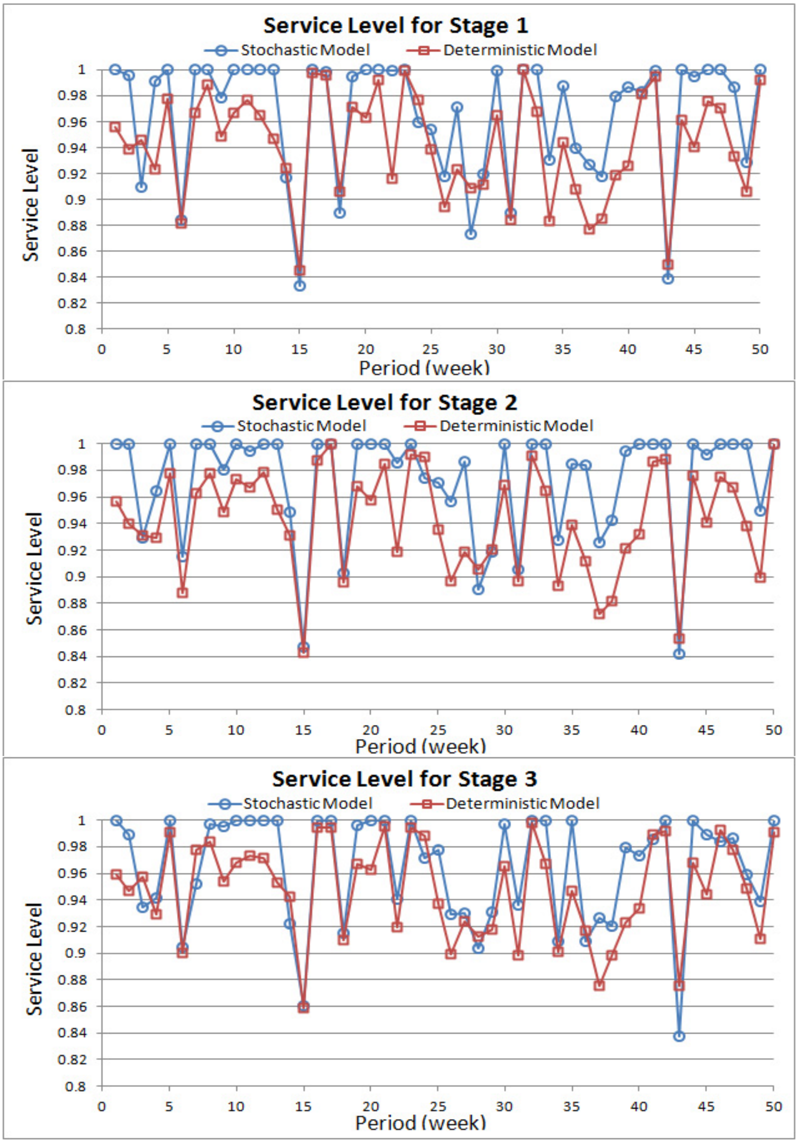


Fig. 8. Demand shortage for each manufacturing stage

6. Conclusion

This chapter studies the design of a supply chain involving multi-stage manufacturing operations with probabilistic product demand. Three levels of supply chain issues are first discussed. In the strategic level, a 'transnational vertical integration' market and production location strategy is taken, where the multi-stage manufacturing system is across various geographical locations and finished goods are sold in North American market. The study then mainly focuses on determining how much capacity should be allocated to which production facility for each manufacturing stage.

Manufacturing configuration in each individual facility is also taken into account in this chapter. The chapter integrates manufacturing system design with supply chain design by proposing a layered mini-cellular system. Each mini-cell is assumed to operate one manufacturing stage with maximum 40 hours weekly capacity. In the classical cellular system, a cell is dedicated to one product family. A layered system not only consists of dedicated mini-cells but also shared and remainder mini-cells. Mini-cell requirements and utilization are first estimated by using probability equations. Mini-cells are then grouped based on operation similarities among product families, and eventually, dedicated, shared and remainder mini-cells are formed. The latter two types of mini-cells deal with more than one product family so that resources are shared and demand fluctuation could be neutralized to a certain level.

A capacitated plant location math model is proposed to form supply chain network as well as allocate mini-cells to each facility for each manufacturing stage. Both mini-cell components and transportation costs are taken into account this model. Next, a capacity planning model determines detailed production quantity based on a specific weekly demand.

Experimentation is conducted and the results indicate the selection of production facilities and allocation of capacity. The performance of layered system is compared with the results of the classical cellular manufacturing system. It is important to notice that despite the lower number of machines/workforce was required by the layered system; the layered system satisfies demand better compared to the classical system. The results indicate that this study provides a complementary analytical model that explores the efficient way to locate and allocate inbound resources so that a certain level of supply chain efficiency and responsiveness could be achieved.

7. References

Chopra, S., & Meindl, P. (2007). Supply *Chain Management-Strategy: Planning, & Operation* (third ed.). New Jersey: Prentice-Hall, Inc., ISBN 978-013-6080-40-4, New Jersey

Cosner, J. D. (2008). Technology, Location, Price, and System Design Decisions for a Global Manufacturing Company. Master's Thesis, Industrial and Systems Engineering, Ohio University, Available from http://etd.ohiolink.edu/view.cgi/Cosner%20Jeremy%20D.pdf?ohiou1225542925

Dicken, P. (1992). *Global Shift: The Internationalization of Economic Activity* (second ed.). London: Paul Chapman Publishing, ISBN 978-185-3961-42-7, London

Lee, H. L. (2000). Creating Value through Supply Chain Integration, *Supply Chain Management Review,* Vol.6, pp. 30-37, ISSN 1521-9747

Lee, H. L. (2005). The Triple-A Supply Chain, *Harvard Business Review*, Vol. 683, No. 1, pp. 112-113, ISSN 0017-8012

Rao, P. P., & Monhanty, R. P. (2003). Impact of cellular manufacturing on supply chain management: exploration of interrelationships between design issues. *International Journal of Manufacturing Technology Management,* Vol. 5, pp. 507-520, ISSN 1368-2148

Schaller, J. (2008). Incorporating cellular manufacturing into supply chain design, *International Journal of Production Research,* Vol. 46, No. 17, pp. 4925-4945, ISSN: 0020-7543

Suer, G. A., Huang, J., & Maddisetty, S. (2010). Design of dedicated, shared and remainder cells in a probabilistic demand environment. *International Journal of Production Research,* Vol. 48, No. 19, pp. 5613-5646, ISSN: 0020-7543

19

Stochastic Capacitated Cellular Manufacturing System Design with Hybrid Similarity Coefficient

Gökhan Eğilmez[1], Gürsel A. Süer[1,*] and Orhan Özgüner[2]
[1]*Ohio University*
[2]*Johns Hopkins University*
USA

1. Introduction

Manufacturing system design is one of the most crucial steps of business processes. Several approaches have been proposed and implemented to increase productivity and profitability due to the change in customer characteristics, market condition and economy. Cellular Manufacturing (CM) is one of the approaches that emerged as an application of Group Technology in the late 70s due to the increase in product variety and demand variance. Group Technology (GT) is a product-oriented manufacturing approach to group similar products for smaller batch size production. As an application of GT, CM is the physical or virtual division of manufacturing facilities into manufacturing cells. A manufacturing cell is a small group of machines and/or workers ideally arranged in a flow layout to produce "similar items", in other words "product families".

Production volume and product variety have significant impact on the design of manufacturing system. Layout of the shop floor is generally used to classify the manufacturing systems. There are four well-known layout types; namely: fixed layout, product layout, process layout and cellular layout. A fixed layout consists of fixed parts and non-fixed resources which travel to parts to perform the operations. In product layout, resources are arranged based on the sequence of operations. This layout is very efficient to meet high volume demand when product variety is low. On the other hand, in process layout, similar resources thus processes are grouped together to meet low and medium volume demand and high product variety. Product layout is more efficient in terms of material flow, whereas process layout is more flexible to deal with high product variety. Cellular layout is a hybrid layout which includes the advantages of both product and process layouts. Cellular layout improves the manufacturing system performance from many aspects such as reduction in material handling, lead times, work-in-process inventory (WIP), re-work, scrap and efficient floor space usage (Wemmerlov U. & Johnson D. J., 1997).

Even though most of the cellular manufacturing system (CMS) design approaches work with deterministic data, uncertainty indeed significantly influences the CMS performance,

---

* Corresponding Author

especially in labor-intensive manufacturing cells. Therefore, the impact of such design parameters as variance in demand and variance in processing times should be taken into consideration during the CMS design. Moreover, most of the works in literature deals with the cell and family formation problem from only route-based similarity point of view. However, since the demand variance can have ruining impact on system performance, demand-based similarity should be also taken into consideration when building similarity matrix. In this chapter, a hybrid similarity matrix, which incorporates route-based and demand-based similarities, is proposed and a stochastic non-linear mathematical model is developed to design CMS considering uncertain demand and processing times. To validate the proposed model, simulation experiments are carried out. Finally, a Genetic Algorithm approach is proposed to deal with large problems.

2. Literature review

The literature is abundant with the works that include optimization methods. In addition to mathematical models, heuristics and meta-heuristics are used to tackle larger problems. The majority of works in literature address deterministic CMS design. However, uncertainty in some parameters such as demand and processing times brings probabilistic nature to design problems. While most of the studies in the literature have addressed the deterministic CMS design problem, less attention is paid to the problems that consider the probabilistic demand and processing times. The literature is reviewed in two sections, namely: deterministic design and stochastic design.

2.1 Deterministic CMS design

In deterministic case, mathematical optimization techniques are used to solve the cell formation problem. As a preliminary work, Purcheck (1974) developed a mathematical classification for the systematic analysis of process routes to group technology and cell formation problem (Purcheck, 1974). Kusiak (1987) provided a comparison of matrix and integer programming models, and discussed the impact of the models on the quality of process families and machine cells (Kusiak, 1987). Shtup (1989) proved the equivalency of cell formation problem to the Generalized Assignment Problem (GAP) (Shtubt, 1989). Rajamani, Singh and Aneja (1990) studied the impact of alternative process plans on the resource utilization and developed three integer programming models to analyze the effect of alternative process plans and simultaneous formation of part families and machine groups (Rajamani, Singh, & Aneja, 1990). Wei and Gaither (1990) developed an integer programming model for cell formation problem (Wei & Gaither, 1990). The objective was to minimize the cost of manufacturing exceptional parts outside the cellular system, subject to machine capacity constraints. Shafer and Rogers (1991) proposed a goal programming model to CMS design problem with the objectives: reducing setup times, minimizing intercellular movements of products and the investment in new equipment, and maintaining acceptable machine utilization levels (Shafer & Rogers, 1991). Kamrani, Parsaei and Leep (1995) developed a mathematical model and tested the performance of the model with simulation in four phases, namely: coding of parts, family formation, resource optimization, simulation (Kamrani, Parsaei, & Leep, 1995). Heragu and Chen (1998) applied a mathematical model to cell formation problem by considering three aspects; resource utilization, alternate routings, and practical constraints (Heragu & J. Chen, 1998). Chen (1998) worked on designing a sustainable cellular manufacturing system in a dynamic environment and developed an integer programming model to minimize material handling and machine costs as well as cell reconfiguration cost for a multi-period planning horizon (Mingyuan Chen, 1998). Wang (1998) formulated a linear assignment model to the group formation problem (Wang, 1998). Sofianopolou (1999) proposed a mathematical model and a two-phased simulated annealing algorithm to solve the problem of grouping machines into cells and selecting a unique product process plan for each product to be produced (Sofianopoulou, 1999). The manufacturing systems considered include such features as replicate machines and several design requirements as well as operation sequence constraints. Akturk and Turkcan (2000) developed an integrated algorithm that considers the cell layout, part-family and cell formation problems simultaneously (Akturk M.S. & Turkcan A., 2000). Albadawi, Bashir and Mingyuan (2005) proposed a two-phased mathematical model for cell the formation problem (Albadawi, Bashir, & Mingyuan Chen, 2005). In the first phase, factor analysis is used to build similarity matrix and machine cells are identified. In the second phase, parts are assigned to the identified machine cells with an integer programming model.

Metaheuristics have also been used to deal with larger cell formation problems. The most commonly used ones are Genetic Algorithms (GA), Simulated Annealing (SA) and Tabu Search (TS). Genetic Algorithms is a random search technique which generates solutions by using techniques inspired by natural evolution. Simulated Annealing is another search based optimization technique which evolves by replacing the current solution by a random "nearby" solution to reach a near global optimum. Tabu Search is a local neighborhood search technique which improves the solution quality by modifying the neighborhood structure of each solution as the search progresses. Moon, Gen and Süer (1999) developed a GA model to minimize additional capital investment in manufacturing cell design (Moon, Gen, & Suer, 1999). Asokan, Prabhakaran and Kumar (2001) proposed two metaheuristics, GA and SA, for the cell formation problem with the objective of minimizing the total moves and minimizing the cell load variation (Asokan, Prabhakaran, & Satheesh Kumar, 2001). Süer, Pena and Vazquez (2003) developed an evolutionary algorithm and applied to three different problems with seven different cost schemes with the objective of minimizing the total machine investment cost (Suer, Pena, & Vazques, 2003). Cao and Chen (2004) formulated an integrated methodology, which consists of a mixed integer non-linear programming model and a TS algorithm for the NP-Hard Problems (Cao & M. Chen, 2004). Jayaswal and Adil (2004) added simulated annealing and local search heuristics to minimize the sum of costs of inter-cell moves, machine investment and machine operating costs (Jayaswal & Adil, 2004). Solimanpur, Vrat and Shankar (2004) modeled a multi-objective integer programming and GA with multiple fitness functions to the design of cellular manufacturing systems with independent cells (Solimanpur, Vrat, & Shankar, 2004).

2.2 Stochastic CMS design

In contrast to abundant literature on deterministic CMS design, only a handful of works dealt with uncertainty. Such stochastic parameters as demand, processing time, capacity requirements are the driver of uncertainty in manufacturing environment. Seifoddini (1990) dealt with the uncertainty of the product mix and developed a probabilistic model to minimize the expected intercell material handling costs of the system (Seifoddini, 1990). Harhalakis, Nagi and Proth (1998) studied minimizing the expected inter-cell material handling cost over the entire design horizon and developed a two-stage heuristic approach (Harhalakis, Nagi, & Proth, 1998). In the first stage, the production volumes are determined with respect to the joint probabilities for every feasible production mix; in the second stage, the cell formation is obtained via the heuristic method. Wicks and Reasor (1999) employed forecasting methods to determine product mix and the demand for products and solved the multi-period cell formation problem with GA (Wicks & Reasor, 1999). Saad (2003) addressed reconfiguration of manufacturing systems and developed following sub-modules; configuration and reconfiguration module, loading module, and simulation-based scheduling module (Saad, 2003).

Queuing theory is also applied to cell formation problem (Mehrabad and Ghezavati, 2009). Each machine is considered as server and each product is assumed as customer. The objective is to minimize the idleness costs for machines, the total cost of sub-contracting for exceptional elements and the cost of resource underutilization. Süer et al. (2010) proposed both deterministic and stochastic approaches in CMS design. Their stochastic approach considered uncertainty in both product demand and production rates (Suer, Huang, & Sripathi, 2010). In this approach, a layered cellular design concept is introduced to cell formation problem. Cells are identified as dedicated, shared and remainder cell to deal with the uncertainty and a product family can be assigned to more than one Cell. In their study, the generalized p-median model by (Kusiak, 1987) is modified to meet objectives as maximizing the utilization of cells and forming the most similar parts as families. However, cell formation model considers the capacity requirements as deterministic even though there is uncertainty in demand and production rates thus capacity requirements.

In this chapter, a stochastic capacitated p-median model is developed to deal with the probabilistic demand and production rates, thus capacity requirements based on the Süer et al.'s (2010) deterministic approach. A new similarity coefficient is defined to combine the demand and process similarity. A new Genetic Algorithm (GA) model is developed for the larger problems. The obtained cell configurations from stochastic mathematical model and GA are simulated with Arena Simulation Software.

3. The manufacturing system studied

The problem is derived from a jewelry company. There are thirty products and eighteen machines in the system. Each product has to be processed on several machines depending on its process route. Since each product's route represents a unidirectional flow, the cell configuration is flow shop. The machine with the maximum processing time among all machines on process route is the bottleneck machine.

Each cell in the system is allocated to only one product family. In other words, cells are independent and dedicated to one product family. Hence, inter-cell transfer of products is not allowed. Inter-cell transfer restrictions have been also used in several manufacturing systems such as pharmaceutical, medical device and food manufacturing. In some of these industries, independent cell configuration is inevitable since potential product mix up may cause serious problems. Each product can only be assigned to one cell (no product splitting is allowed among cells). Since machine setup times are negligible, they are assumed to be zero in this study. Annual production capacity is taken as 2000 hours (50 weeks/yr * 40 hours/week). The annual demand and processing time for each product are random and follow normal distribution. The problem is the identification of product families and corresponding dedicated cells considering stochastic demand, stochastic processing times and hybrid similarity coefficient.

4. The proposed solution methodology: Stochastic CMS design

The proposed solution methodology is a hierarchical one and it consists of five steps, namely: identification of similarity coefficients, determining the bottleneck machine, and determining the probabilistic capacity requirements, stochastic non-linear mathematical model, and simulation. An example problem is solved to explain the methodology used.

4.1 Identification of similarities

In this section, identification of similarities is explained. Three types of similarity coefficients are used, namely: route-based, demand-based and hybrid similarity. Route- based similarity coefficient only considers the processing similarities of products in the manufacturing system. Demand-based similarity only considers the demand variation among products. Hybrid similarity is the combination of both similarity coefficients. Both of the similarity coefficients are explained in detail in the following sections.

4.1.1 Route-based similarity

The route-based similarity matrix is constructed based on the route similarities among products. Süer et al. (2010) modified the McAuley's (1972) similarity coefficient definition to find the similarities among products. The similarity coefficients are calculated via the suggested equation by Süer et al. (2010) as shown in equation 1. The route-based similarity (RB_{ij}) between products *i* and *j* is the ratio of number of common machines to total number of machines required.

$$RB_{ij} = \frac{No.of\ \ machines\ processing\ both\ parts\ i\ and\ j}{No.of\ \ machines\ processing\ parts\ either\ i\ or\ j} \tag{1}$$

4.1.2 Demand-based similarity

The main motivation of this similarity measure is to identify stable and unstable products. Stable products have lower variability and unstable products have higher variability. Assigning stable and unstable products to the same cell can cause turbulence in the cell. Even a single unstable product can complicate the operation control issues in a cell. Therefore stable products and unstable products are separated and allocated to different cells. By doing this; the turbulence in CMS is restricted to cells with unstable products only.

In this similarity coefficient, products' similarities are calculated based on the variability in demand (Equation 2). The variability in demand for product *i* (Vd_i) is obtained dividing mean demand (μd_i) by the variance of demand ($\sigma^2 d_i$) as shown in Equation 2 (Silver & Peterson, 1985). Firstly, the demand variability is calculated for all products.

$$Vd_i = \frac{\mu d_i}{\sigma^2 d_i} \quad (2)$$

Secondly, the absolute difference between each pair of products' variability values is obtained and entered in the difference matrix. Thirdly, the obtained difference values are scaled from 0 to 1 to be converted to demand-based similarity coefficients of pairs. In other words, the variability difference matrix is converted to variability dissimilarity matrix to be used as demand-based similarity matrix. The maximum difference that a pair has in difference matrix is assumed as the greatest dissimilarity. The scaling is applied assuming the maximum difference as 1. Fourthly, the dissimilarity matrix is converted to similarity matrix by subtracting dissimilarity values from 1.

4.1.3 Hybrid similarity

There is a need to strike a balance between route-based similarity vs demand-based similarity. Hybrid similarity coefficient is developed to cover both of previously explained similarities. Equation 3 represents the calculation of similarity coefficient. Beta (β) and (1- β) are the proportional impacts of route-based and demand-based similarities on the hybrid similarity coefficient, respectively. In this study, Hybrid Similarity Coefficient is used in CMS design.

$$H_{ij} = \beta * RB_{ij} + (1-\beta) * DB_{ij} \quad (3)$$

4.1.4 Hybrid similarity example

An example is derived to illustrate how the similarity concept is applied. Assume that there are five products with the following route, probabilistic demand and demand variability information shown in Table 1. According to the route information given in Table 1, route-based similarities are calculated by using Equation 1. The route-based similarity matrix is shown in Table 2. Demand-based similarity is the second step to calculate the hybrid similarity.

To build demand-based similarity matrix, first of all the difference between the mean demand/variance of demand ratios (Vd_i) are calculated for all pairs and shown in Table 3. The maximum difference is 4.757 and the minimum difference is 0. These values are scaled to 0-1 range as shown in Table 4. These values are then subtracted from 1 and thus the dissimilarity matrix given in Table 4 is converted to demand-based similarity matrix given in Table 5.

Product	Opr. 1	Opr. 2	Opr. 3	Mean Demand	Variance of Demand	Vd_i
1	A	B	C	2999	1284	2.336
2	A	C		4297	2604	1.650
3	A	C	D	2217	346	6.408
4	B	C	E	1255	359	3.496
5	D	F		2463	454	5.425

Table 1. Operational routes and demand information

Product \ Product	1	2	3	4	5
1	-	0.67	0.5	0.5	0
2	0.67	-	0.67	0.25	0
3	0.5	0.67	-	0.2	0.25
4	0.5	0.25	0.2	-	0
5	0	0	0.25	0	-

Table 2. Route-based similarity matrix

Product \ Product	1	2	3	4	5
1	0.000	0.686	4.072	1.160	3.089
2	0.686	0.000	4.757	1.846	3.775
3	4.072	4.757	0.000	2.912	0.982
4	1.160	1.846	2.912	0.000	1.929
5	3.089	3.775	0.982	1.929	0.000

Table 3. Vd_i Difference Matrix

Scaled Matrix	1	2	3	4	5
1	-	0.144	0.856	0.244	0.649
2	0.144	-	1	0.388	0.793
3	0.856	1	-	0.612	0.207
4	0.244	0.388	0.612	-	0.406
5	0.649	0.793	0.207	0.612	-

Table 4. Scaled Difference Matrix

Product \ Product	1	2	3	4	5
1	-	0.856	0.144	0.756	0.351
2	0.856	-	0.000	0.612	0.207
3	0.144	0	-	0.388	0.793
4	0.756	0.612	0.388	-	0.594
5	0.351	0.207	0.793	0.594	-

Table 5. Demand-based similarity matrix

After both route-based and demand-based similarity matrices are built, hybrid similarity matrix is developed by using Equation 3. In this example, β is taken as 0.5. The developed hybrid similarity matrix is shown in Table 6.

Product \ Product	1	2	3	4	5
1	-	0.763	0.322	0.628	0.175
2	0.763	-	0.335	0.431	0.103
3	0.322	0.335	-	0.294	0.522
4	0.628	0.431	0.294	-	0.297
5	0.175	0.103	0.522	0.297	-

Table 6. Hybrid similarity matrix

4.2 Bottleneck machine identification

In this case, the definition of bottleneck machine is modified since processing times are probabilistic. An example is given in Table 7 to illustrate the situation for product *i*.

Product *i*	Opr. – 1 on M/C 1	Opr. – 2 on M/C 2
Mean (μ)	5 min	4 min
Standard Deviation (σ)	1.2	1.6
Process Time Estimate based on 2 Sigma (ε=2)	7.4 min	7.2 min
Process Time Estimate based on 3 Sigma (ε=3)	8.6 min	8.8 min

Table 7. Bottleneck machine identification

Assume that product *i* requires two operations and the mean processing times for operations 1 and 2 are 5 min and 4 min, respectively. Also assume that standard deviation of processing time for operation 1 is 1.2 and for operation 2 is 1.6 minutes. If only mean values are to be considered, machine 1 would be regarded as the bottleneck machine. If processing times are estimated based on 2 sigma (ε= 2) using Equation (4), then processing time estimates will be 7.4 min and 7.2 min, respectively and machine 1 will still be the bottleneck machine. However, if the processing times are estimated based on 3 sigma (ε= 3), then the bottleneck operation will shift to machine 2. In this paper, we have considered the processing time estimate based on 3-sigma level as the basis for the bottleneck machine identification. The reason for this is that the probability that actual processing time will exceed the estimate based on 3-sigma value is very small.

$$p_{ik}^{e} = \mu_{ik} + \varepsilon * \sigma_{ik} \tag{4}$$

where, p_{ik}^{e} is the processing time estimate, μ_{ik} is the mean processing time, σ_{ik} is the standard deviation of processing time for operation k of product *i* and ε is the coefficient of standard deviation.

4.3 Capacity requirements in the presence of stochastic demand and processing times

In the deterministic case, the capacity requirement of a product is calculated via multiplying its demand with processing time. However, in the stochastic case, since both demand and processing time are probabilistic, the product of these two random variables becomes probabilistic and requires statistical analysis to find the probability density functions (pdf) of the capacity requirements. To find the fitted distribution (pdf) of the capacity requirement of product i, statistical analysis is performed with Arena Input Analyzer software.

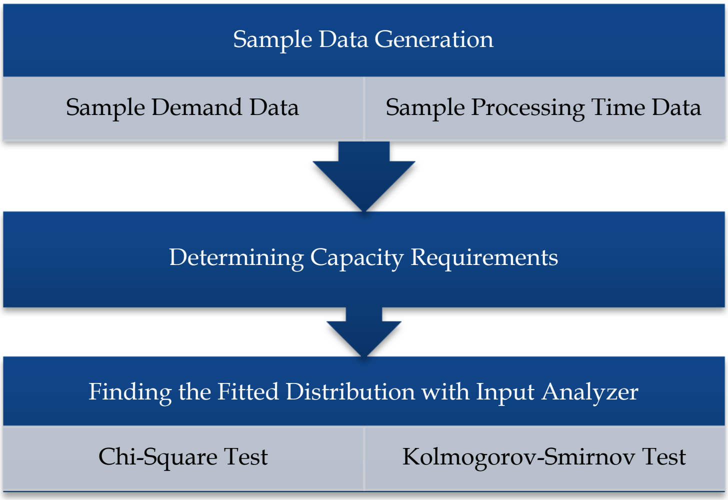


Fig. 1. The Framework of Input Analysis

The framework of the analysis is illustrated in Figure 1. Sample probabilistic demand and processing time data are generated. Capacity requirements are determined from the samples. The fitted distributions for capacity requirements are obtained from Input Analyzer software with respect to test results.

4.4 Stochastic capacitated non-linear cell formation

A stochastic non-linear mathematical model is developed by modifying Süer et al.'s (2010) deterministic model. The proposed model considers the variation of capacity requirements along with the mean capacity requirements. Product families and cell formations are determined with respect to available cell capacity and similarity coefficients. The indices, parameters and decision variables are listed as follows.

Indices:

i	Product index
j	Product index and family/cell index

Parameters:

S_{ij}	Similarity coefficient between product i and j
μ_{CR_i}	Mean capacity requirement for product i
${\sigma_{CRi}}^2$	Variance of capacity requirement for product i
n	Number of products
TU	Upper limit for cell capacity
α	Design factor

Decision Variables:

X_{ij} 1, if product i is assigned to family j ; 0, otherwise

Objective Function:

$$\max Z = \sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij} * X_{ij} - \sum_{j=1}^{n} X_{jj} \tag{5}$$

Subject to:

$$p\left(Z_{nj} \leq \frac{\left(TU - \sum_{i=1}^{n} \mu_{CR_i} * X_{ij} \right)}{\sqrt{\sum_{i=1}^{n} \sigma_{CRi}^{2} * X_{ij}}} \right) \geq (1-\alpha) \quad \text{j} = 1,2,\ldots,\text{n} \tag{6}$$

$$\sum_{j=1}^{n} X_{ij} = 1 \qquad i = 1,\ldots,n \tag{7}$$

$$X_{ij} \leq X_{jj} \qquad j = 1,\ldots,n \; and \; i = 1,\ldots,n \tag{8}$$

$$X_{ij} \in [0,1]; X_{ij} \; integer \tag{9}$$

The objective function is shown in equation 5. It maximizes the total similarity among products that are formed as families to be produced in dedicated cells, while minimizing the total number of cells. Equation 6 is the non-linear constraint which limits the cell utilization up to the cell capacity by considering mean and variance of capacity requirements based on a factor, α, which indicates the maximum acceptable probability that capacity requirements will exceed the capacity available. Equation 7 forces all products to be assigned to a cell. Equation 8 guarantees the assignment of each product to only one of the cells that are open. Equation 9 determines whether product *i* is assigned to cell *j* or not.

4.5 Simulation

In this study, the proposed solution methodology is validated with a simulation model. Even though, the CMS design literature is abundant with several mathematical models, model validation is considered in only a handful of works. Indeed, model validation is one of the significant requirements of any model-based solution methodology. Especially in a system where demand, processing times and capacity requirements are probabilistic, it is a must to validate the proposed approach. The type of model proposed in the study is white-box (causal descriptive) according to the Barlas's classification (Barlas, 1996). Therefore, it is expected that the model reproduces the behavior of the system studied. The behavior of the system is analyzed with respect to four measures, namely: cell utilization, WIP, waiting time and the number waiting. The hierarchical framework followed through the validation is shown in Figure 2.

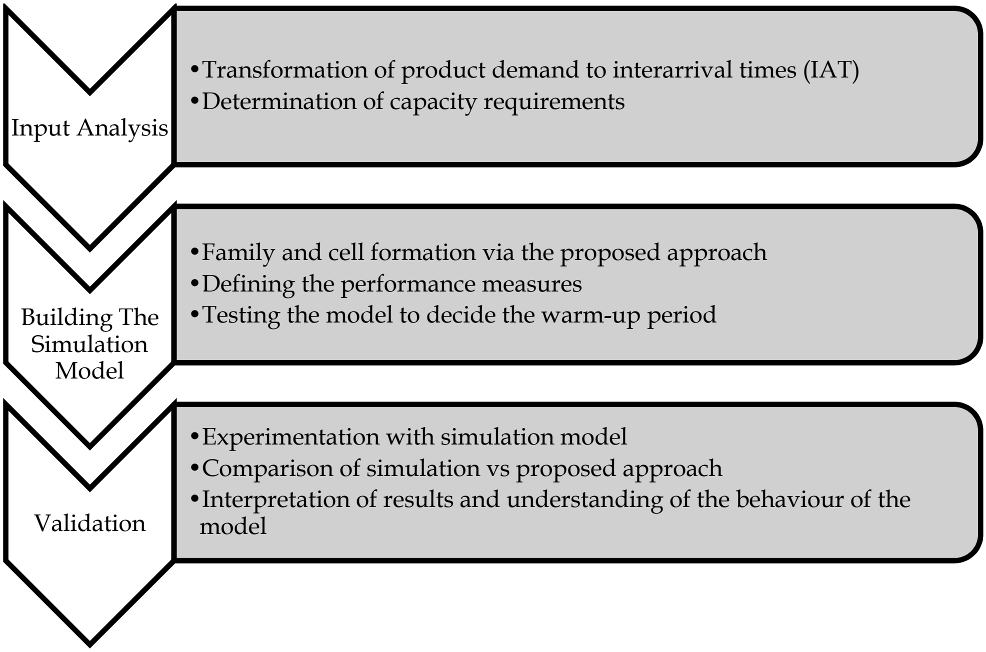


Fig. 2. The hierarchical framework

4.6 Example problem

An example problem is derived to explain the solution methodology (Please refer to the first 10 parts of part-machine matrix in Süer et al., 2010). There are 10 products in the system. The hybrid similarity matrix is shown in Table 8 and probabilistic capacity requirements are shown in Table 9. The capacity of a cell is considered as 800 hours for the example problem.

Part \ Part	1	2	3	4	5	6	7	8	9	10
1	1	0.808	0.765	0.555	0.849	0.561	0.75	0.68	0.712	0.636
2	0.808	1	0.685	0.718	0.747	0.59	0.677	0.775	0.631	0.597
3	0.765	0.685	1	0.545	0.794	0.569	0.703	0.597	0.751	0.627
4	0.555	0.718	0.545	1	0.539	0.562	0.545	0.681	0.477	0.636
5	0.849	0.747	0.794	0.539	1	0.605	0.692	0.65	0.656	0.611
6	0.561	0.59	0.569	0.562	0.605	1	0.499	0.69	0.507	0.656
7	0.75	0.677	0.703	0.545	0.692	0.499	1	0.583	0.855	0.542
8	0.68	0.775	0.597	0.681	0.65	0.69	0.583	1	0.534	0.715
9	0.712	0.631	0.751	0.477	0.656	0.507	0.855	0.534	1	0.499
10	0.636	0.597	0.627	0.636	0.611	0.656	0.542	0.715	0.499	1

Table 8. Hybrid similarity matrix

Part	1	2	3	4	5	6	7	8	9	10
Mean	154	153	494	93.9	106	500	138	45.6	439	135
Variance	231.04	262.44	5745.64	88.92	136.89	5343.61	187.69	25.1	6052.84	222.01

Table 9. Probabilistic capacity requirements (hrs)

Cell Formation	Cell 1 / Family 1	Cell 2 / Family 2	Cell 3 / Family 3	Cell 4 / Family 4
Product Family	Products (1,2,4,8)	Products (3, 5)	Products (6, 10)	Products (7, 9)
Resource Requirements	M/C (1-7,9,10,18)	M/C (1-4,8,10-12,18)	M/C (4,6,9,10,12,14,16, 18)	M/C (1,2,3,11,13,17, 18)
Expected Utilization	447	600	635	577
Util. w.r.t. 10 % risk	478 (60 %)	698 (87 %)	730 (91 %)	678 (85 %)

Table 10. Solution of example problem

The example problem is solved with the proposed non-linear mathematical model. The results are shown in Table 10. Products are formed as four families thus four cells are required where each is dedicated to one product family. Cell and family formations are the same since independent dedicated cells are assumed and intercell movements are not allowed.

5. Alternative solution methodology: Genetic algorithms

Genetic Algorithms (GA) is one of the most powerful metaheuristics used to solve NP-hard problems. It is usually used for solving large size problems where mathematical models run into computational and memory problems. The framework of GA is shown in Figure 3. This framework also represents the one cycle evolutionary process of GA.

GA consists of following steps (Süer, Mese, & Eğilmez, 2011):

1. Initial population of n chromosomes is formed randomly.
2. Mates are determined using the mating strategy to perform crossover.
3. The crossover and mutation operations are performed to generate offspring.
4. For selection, parents are added to the selection pool along with offspring.
5. The next generation is selected from this pool based on their fitness function values.
6. These steps are repeated until the number of the generations specified by the user is reached.
7. Finally, the best chromosome obtained during the entire evolutionary process is taken as the final solution.

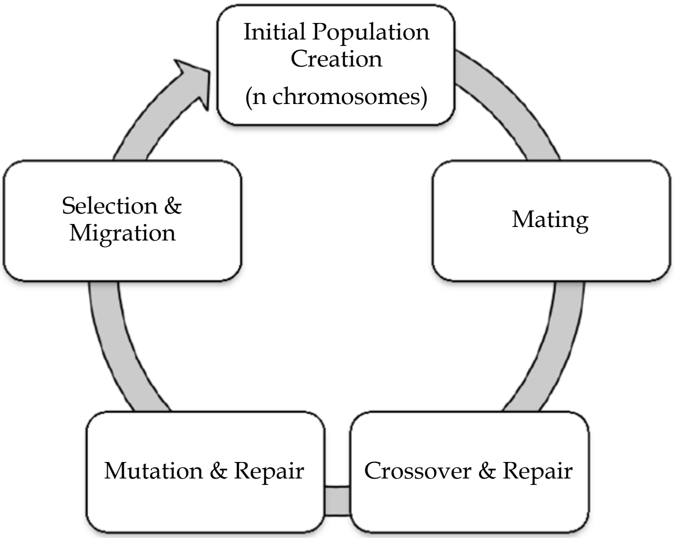


Fig. 3. Illustration of one generation

5.1 Initial population generation

Initial population is randomly generated based on the pre-defined number of chromosomes to form the population. Chromosome represents a candidate solution to the problem. In the proposed GA, the chromosome representation is designed to include product and cell numbers. An example chromosome is shown in Figures 4a and 4b. Each gene carries two types of information: part number and cell number. When generating chromosome, the product numbers are randomly assigned to genes as shown in Figure 4a. Once the assignment is finished the allocation of cells is done with respect to available capacity as shown in Figure 4b. The final chromosome representation is given in Figure 4b.

Cells are matched with products from left to right. First of all, cell 1 is opened and products are assigned to cell 1 as long as the cell capacity is available. If a product is going to result in overutilization, a new cell is opened. Product and cell allocation are illustrated in Table 11.

For example, after the assignment of products 1, 2 and 4 to cell 1, the total expected utilization of cell 1 is 400.9 hours and the standard deviation is 24.13. Under 10 % of risk (z=1.28), the upper bound of utilization is 400.9 + 1.28*24.13 = 431.79. If part 3 were to be assigned to cell 1, the total expected utilization would be 894.9 and variance would be 6328.04. The utilization under 10 % is equal to 894.9 + 1.28 * 79.55 = 996.72 hours > 800 hours. Since the cell is over utilized, a new cell is opened and product 3 is assigned to cell 2. The fitness function of a chromosome is the total similarity of cells. The similarity of cell is calculated via summing the similarities of products within the cell. The resulted cell formation and utilizations are shown in Table 11.

1	2	4	3	5	10	7	9	6	8

Fig. 4a. Example chromosome after product assignment

1,1	2,1	4,1	3,2	5,2	10,2	7,3	9,3	6,4	8,4

Fig. 4b. Example chromosome after cell allocation

	Cell 1			Cell 2	
Part	μ	σ^2	Part	μ	σ^2
1	154	231.04	3	494	5745.64
2	153	262.44	5	106	136.89
4	93.9	88.92	10	135	222.01
Total	400.9	582.40	Total	735	6104.54
	Cell 3			Cell 4	
Part	μ	σ^2	Part	μ	σ^2
7	138	187.69	6	500	5343.61
9	439	6052.84	8	45.6	25.10
Total	577	6240.53	Total	545.6	5368.71

Table 11. Cell utilization (hrs)

5.2 Mating strategy

Random Mating Strategy is used in mating. Firstly, the reproduction probabilities of the chromosomes are calculated according to their fitness function. Each chromosome in the population is mated with a randomly selected partner and they produce one offspring. The partner is selected using reproduction probability based on Roulette Wheel approach.

5.3 Crossover operation and repair

A modified order crossover method is employed for the problem studied. In the order crossover as represented in Figure 5, a random number between 1 and the maximum number of cells (4, in the example given in Figure 5) is drawn. Assume that 2 is drawn. This is to decide how many cells from parent 1 are going to be kept in offspring. Then, the two cells are randomly selected and copied to offspring. The remaining genes are filled from parent 2 based on the order of remaining products in parent 2's chromosome. Finally, cell numbers of products that are obtained from parent 2 are assigned based on the availability of cell capacity. As long as the cell capacity is available, products from left to the right are assigned to the same cell. If a capacity violation occurs, a new cell is opened.

5.4 Mutation operation and repair

Mutation is only applied to part numbers and after the mutation operation, same repair operation as in crossover strategy is employed to re-identify cell formation based on available cell capacity. Each gene is mutated subject to a mutation probability. If a mutation occurs in a gene, a random part number is replaced with an existing part number in a particular cell.

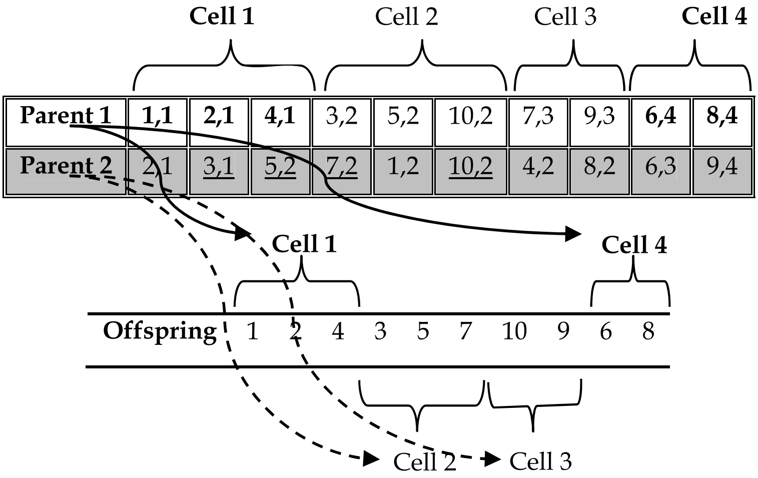


Fig. 5. Cell-Based Order Crossover (CBOC)

5.5 Selection and migration

Selection pool consists of all offspring and parents. The best chromosome is selected as the best solution for the particular generation. After each generation, a predetermined portion of existing generation is advanced to the next generation. Assume that x is the desired percentage of selection pool advancing to the next generation. X is basically an experimentation parameter. When generating next generation, the remaining positions are filled with randomly generated chromosomes (thus the term migration).

6. Experimentation and results

Experimentation consists of two sections. In the first section, the proposed stochastic non-linear mathematical model is compared with Süer et al.'s (2010) deterministic model. The results of both approaches are compared with respect to the obtained cell formation and simulation results. In the second section, proposed alternative solution methodology (GA) is compared with the non-linear mathematical model. Five different problems are considered, namely: 10, 20, 30, 45 and 60 parts. In the first section, the 30-part problem is used for the comparison and simulation runs. For the comparison of non-linear mathematical model with GA, all five problems are used with respect to the solution quality and execution time.

6.1 Data generation

The part-machine matrix (30x18) is obtained from a jewelry manufacturing company. Processing times are generated from uniform distribution with (15, 25) minutes. Each cell is independent, i.e. part undergoes all the operations in only one cell and comes out as a final product. Therefore, intercell movement of parts is not allowed. Independent cells are used in certain systems where intercell movement of parts is either not possible (pharmaceutical manufacturing) or may cause serious problems due to product mix up. One piece flow principle is assumed for the entire cellular manufacturing system. Set-up times for the parts are assumed as zero.

Demand Category	**Annual Uniform Demand Distribution**	
	Lower Bound	Upper Bound
1	250	750
2	751	1250
3	1251	1750
4	1751	2250
5	2251	2750

Table 12. Annual Uniform Demand Data Generation

Cell capacity is assumed as 2000 hours (50 weeks x 40 hrs per week) per year. Remaining two weeks is allocated to compensate for unexpected system breakdowns and plant shutdowns. Demand for each product is assumed to follow normal distribution. The mean demand is generated from uniform distribution from five categories. Each product is assigned randomly to a category (see Table 12). The standard deviation of demand is generated via multiplying the mean demand with a factor. The factor is obtained randomly via uniform distribution (0.1, 0.5).

6.2 Comparison of Süer et al.'s (2010) model with the proposed stochastic approach

In this section, the results of Süer et al.'s (2010) deterministic approach (see Figure 6) and the proposed stochastic approach (see Figure 7) are provided. The deterministic model grouped the products into 10 families/cells based on the hybrid similarity matrix. Simulation experiment resulted in 100% utilization in cell number 5 and second highest utilization is observed in cell number 2. Since deterministic model only considers mean capacity requirements of products, some of the cells are utilized over 95% when deterministic mathematical model is used. However, these high utilization rates resulted in the same or lower utilization with simulation experimentation. The overall trend of utilization obtained from simulation is observed as similar to the result of mathematical model.

According to the results of the proposed stochastic non-linear approach (Figure 7), products are formed as 13 families/cells. The proposed approach increased the number of cells by 3 and the number of machines by 18%. A similar trend between the result of mathematical model and simulation is also observed with the stochastic approach. In contrast to very high utilization observed on 2 cells in the deterministic approach, the highest utilization is obtained as 82% with the stochastic approach (from simulation). Simulation model resulted in 1% to 7% less utilization of cells than mathematical model's results. There is no overutilization observed in any cell with simulation since the proposed approach considers variance of capacity requirements in addition to similarities during the cell formation.

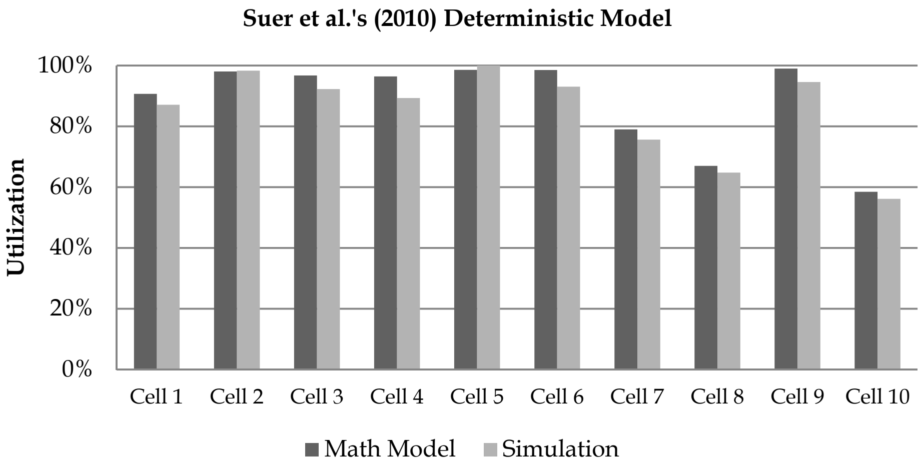


Fig. 6. Utilization Results of Deterministic Model by Süer et al. (2010)

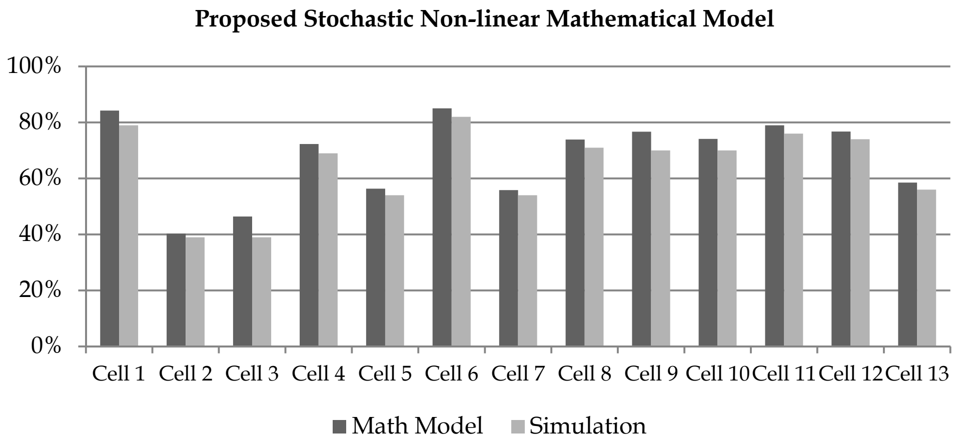


Fig. 7. Utilization Results of the Proposed Stochastic Model

Utilization-based comparison is important to observe the model validation with simulation experiments. However, the behavior of model is also important in model validation. Therefore, system performance is also included in comparison. The performance measures considered are 1) cell utilization, 2) the number of machines, 3) work-in-process (WIP) inventory, 4) average waiting time and 5) average number waiting. The results obtained from both approaches are shown in Table 13.

According to the system performance comparison (Table 13), WIP, average waiting time and average number of waiting decreased significantly since there is no overutilization occurred with stochastic approach. However, the number of machines increased from 94 to 111.

Deterministic Approach (Süer et al., 2010)					Proposed Stochastic Approach				
Cell	# M/C	WIP	Av. Waiting Time	Av. Number Waiting	Cell	# M/C	WIP	Av. Waiting Time	Av. Number Waiting
1	10	6.06	18.8	1.22	1	17	3.53	11.73	0.44
2	13	7.84	12.17	1.05	2	10	2.27	2.09	0.06
3	10	6.06	19.11	1.09	3	9	1.64	2.72	0.11
4	12	6.62	17.57	0.97	4	8	1.72	2.50	0.14
5	13	298.46	6256.56	294.30	5	7	5.00	5.42	0.18
6	12	6.83	18.45	1.03	6	11	6.02	9.90	0.43
7	4	2.52	1.04	0.03	7	11	5.53	4.73	0.22
8	8	3.9	0.39	0.01	8	6	4.37	3.98	0.30
9	8	5.72	21.37	1	9	8	2.95	5.24	0.26
10	4	1.94	0.07	0	10	9	7.60	7.90	0.42
Total	94	345.97	6365.53	300.71	11	4	1.90	1.02	0.03
					12	7	3.86	7.52	0.30
					13	4	1.95	0.07	0.00
					Total	111	48.34	64.82	2.90

Table 13. System Performance Based Comparison

6.3 Comparison of the proposed stochastic non-linear mathematical model with GA

The problem studied is NP-Hard. Therefore, it is important to provide alternative solution approaches that are providing near optimal solution in faster times when dealing with larger problems. In this section, proposed mathematical model and genetic algorithms are compared. The comparison is made in two ways: 1) system performance and 2) solution quality and execution time. Five datasets are used consisting of 10, 20, 30, 45 and 60 parts. Mathematical model could only provide global optimal solutions for the first 4 datasets. To be able to analyze system performance, both solutions for dataset with 30 parts obtained from mathematical model and GA are simulated. The simulation results are shown in Table 14. According to the results, both approaches formed 30 products as 13 families and opened 13 dedicated cells. The utilizations of cells vary from 39% to 82%. In terms of total number of machines, GA resulted in 10 less machines. Even though, there is parallelism between WIP and average number waiting results, in terms of average waiting times, GAs solution provided more than 50% lower average waiting time. However, both mathematical model and GA results showed better performance than Süer et al.'s (2010) deterministic model.

Secondly, both approaches are compared based on the solution quality and execution time. The optimal solution of mathematical model, and best and average solution of GA, execution times and the average % distance (gap) GA's solution from mathematical model and parameter set used in GA are shown in Table 15. Mathematical model provided global optimal solution for the first 4 datasets. On the other hand, GA provided the optimal solution with 100% frequency with 10 parts dataset. With the datasets: 20, 30 and 45 parts; GA provided near optimal solutions with great improvements in execution times. Since the solutions of 30 parts dataset provided by GA and mathematical models are compared in detail in Table 14, it can be concluded that near optimal solutions with respect to similarity coefficient can still provide good or even better system performance. However, system performance should always be assessed with validation methods e.g. simulation.

	Stochastic Non-linear Mathematical Model				Genetic Algorithms			
Cell	# M/C	WIP	Av. Waiting Time	Av. Number Waiting	# M/C	WIP	Av. Waiting Time	Av. Number Waiting
1	17	3.53	11.73	0.44	8	0.94	2.93	0.11
2	10	2.27	2.09	0.06	6	2.39	2.03	0.15
3	9	1.64	2.72	0.11	5	2.46	1.85	0.11
4	8	1.72	2.50	0.14	5	0.13	0.04	0.00
5	7	5.00	5.42	0.18	4	1.41	0.51	0.02
6	11	6.02	9.90	0.43	8	0.28	0.03	0.00
7	11	5.53	4.73	0.22	10	1.92	3.00	0.12
8	6	4.37	3.98	0.30	14	6.15	5.43	0.34
9	8	2.95	5.24	0.26	8	4.57	1.48	0.11
10	9	7.60	7.90	0.42	6	5.48	0.46	0.02
11	4	1.90	1.02	0.03	7	1.87	0.20	0.01
12	7	3.86	7.52	0.30	12	14.54	3.58	0.15
13	4	1.95	0.07	0.00	8	5.73	4.25	0.20
Total	**111**	**48.34**	**64.82**	**2.90**	**101**	**47.86**	**25.79**	**1.34**

Table 14. System Performance Comparison of Mathematical Model and GA

	Stochastic Non-linear Mathematical Model		Genetic Algorithms				
Problem Size	Optimal Solution	Sol. Time (Sec)	Best Solution	Avr. Solution	Avr. % Gap	Avr. Exe. Time (Sec)	Parameter Set (IP,NOG,CP,MP,MR)
10	4.606	145	4.606	4.606	0%	9.70	(100, 200, 0.7-0.9, 0.01, 10-20%)
20	10.87	1522	10.57	10.11	7%	16.90	(500, 100, 0.7-0.9, 0.01, 10-20%)
30	12.29	40836	11.80	11.53	6%	32.93	(1000, 1000, 0.7-0.9, 0.01, 10-20%)
45	20.064	198662	17.024	16.315	19%	78.13	(1000, 2000, 0.7-0.9, 0.01, 10-20%)
60	N/A	N/A	26.855	25.164	-	114.31	(1000, 2000, 0.7-0.9, 0.01, 10-20%)

IP: Initial population, NOG: Number of generations, CP: Crossover probability, MP: Mutation probability, MR: Migration rate

Table 15. Comparison Based on Solution Quality and Execution Time

7. Conclusion

In this study, the impact of probabilistic demand and processing times on cell formation is addressed. The uncertainty in demand and processing times is one of the most common problems of manufacturing world. Manufacturing system design is also directly influenced by such factors which may result in million dollars of wrong investment on machines and equipments. Majority of literature on cell formation and manufacturing system design either neglects the uncertainty in demand and processing times or assumes the impact as limited. Süer et al. (2010) proposed two approaches; deterministic and stochastic. Deterministic model used in their study formed the cells based on the expected (mean) capacity requirements. On the other hand, stochastic approach proposed a layered manufacturing system design to deal with uncertainty in demand and processing times. They allowed a family to have more than one cell to deal with uncertainty. In this study, each family is restricted to one cell and the impact of variance is included in the proposed approach. A hierarchical methodology is used to solve the problem. First of all a new similarity coefficient is introduced which incorporates the route and demand similarity in one similarity coefficient, “hybrid similarity coefficient”. Hybrid similarity matrix is built based on the data obtained from a jewelry manufacturing company where the operations are labor intensive. To deal with the cell and family formation under the impact of uncertain demand and processing times, a stochastic non-linear mathematical model is developed.

The proposed model and Suer et al.'s (2010) deterministic model are experimented with 30x18 (part x machine) dataset. During the modeling and experimentation phases, independent cell and family formation is considered. Therefore, intercell movement of parts is not allowed. Each formed product family is assumed to be produced by a dedicated cell. In addition to solving the cell and family formation problem with mathematical optimization, model validation is also considered. To validate the designed cellular manufacturing system, simulation models are developed for both proposed stochastic and Süer et al's (2010) deterministic approaches. According to the simulation results, similar cell utilization patterns are obtained between the mathematical model and simulation results for both proposed and Suer et al.'s (2010) deterministic approaches. In the further analysis, both approaches are compared based on the system performance. Five performance indicators are considered in comparison, namely: cell utilization, the number of machines, WIP, average waiting time, average number waiting. According to the results, the proposed approach resulted in more number of cells and more machines, but it performed better with respect to all other performance measures.

Even though, the proposed non-linear stochastic model with respect to hybrid similarity coefficient guarantees the optimal solution, solution time increases exponentially as the number of parts increases due to the NP-hardness of problem studied. Metaheuristics are usually employed to deal with NP-hard problems. Therefore, a Genetic Algorithm (GA) model is also developed to deal with the medium and large scaled problems. The probabilistic parameters are also reflected in GA. Experimentation is performed with five datasets, namely 10, 20, 30, 45 and 60 parts. All other datasets are generated as a portion or random replication of 30 parts datasets. First of all, the solution obtained from GA for 30 parts is simulated. Both stochastic non-linear mathematical model and GA's simulation results are compared. According to the simulation results, GA performed better in terms of all performance measures. In addition, both GA and mathematical model resulted in better performance than the deterministic model. Secondly, remaining datasets are also experimented with both proposed mathematical model and GA. According to the results, GA found the optimal solution with 100% frequency for the first dataset (10 parts). For the larger datasets 20, 30 and 45 parts, GA provided 6% to 19% distant solution than the optimal solution. The largest (60 parts) dataset can only be solved by GA. Besides, GA significantly outperformed the mathematical model in all datasets in terms of execution time. Another important conclusion is that the optimal solution with respect to similarity does not guarantee that system performance will be better. The proposed CMS is required to be validated with simulation.

Even though GA provided good and faster solutions and to the best knowledge of authors there has not been any GA approach including stochastic components proposed for the problem studied in literature yet, the proposed GA is planned to include alternative genetic operators to be able to increase the solution quality. In addition, the problem studied can be extended to include other features of CMS design such as allowance of intercell movement of parts, system implementation costs, setup times etc. The impact of variance on system design is another potential important side of the problem which may affect the solution significantly. Moreover, identification of the bottleneck machine in systems where bottleneck machine shifts is also another important potential future work. Finally, in addition to genetic algorithms, other meta-heuristics such as Simulated Annealing, Tabu Search can be considered to cope with the larger problems.

8. References

Akturk M.S., & Turkcan A. (2000). Cellular manufacturing system design using a holonistic approach. *International Journal of Production Research*, *38*(10), 21. Taylor and Francis Ltd, ISSN: 0020754.

Albadawi, Z., Bashir, H. A., & Chen, Mingyuan. (2005). A mathematical approach for the formation of manufacturing cells. *Computers and Industrial Engineering*, *48*(1), 3 - 21, ISSN: 03608352.

Asokan, P., Prabhakaran, G., & Satheesh Kumar, G. (2001). Machine-Cell Grouping in Cellular Manufacturing Systems Using Non-traditional Optimisation Techniques - A Comparative Study. *The International Journal of Advanced Manufacturing Technology*, *18*(2), 140-147. Springer London, ISSN: 0268-3768.

Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review*, *12*(3), 183-210, ISSN: 08837066

Cao, D., & Chen, M. (2004). Using penalty function and Tabu search to solve cell formation problems with fixed cell cost. *Computers & Operations Research*, *31*(1), 21-37, ISSN: 03050548.

Chen, Mingyuan. (1998). *A mathematical programming model for system reconfiguration in a dynamic cellular manufacturing environment*. *Annals of Operations Research* (Vol. 77, pp. 109-128-128). Springer Netherlands, DOI: 10.1023/A:1018917109580.

Harhalakis, G., Nagi, R., & Proth, J. M. (1998). An Efficient Heuristic in Manufacturing Cell Formation for Group Technology Applications. *International Journal of Production Research*, *28*(1), 185-198, DOI: 10.1080/00207549008942692.

Heragu, S., & Chen, J. (1998). Optimal solution of cellular manufacturing system design: Benders' decomposition approach. *European Journal of Operational Research*, *107*(1), 175-192, ISSN: 03772217.

Jayaswal, S., & Adil, G. K. (2004). Efficient algorithm for cell formation with sequence data, machine replications and alternative process routings. *International journal of production research*, *42*(12), 2419-2433. Taylor & Francis, ISSN: 0020-7543.

Kamrani, A. K. A. K., Parsaei, H. R., & Leep, H. R. (1995). *Planning, Design, and Analysis of Cellular Manufacturing Systems*. *Manufacturing Research and Technology* (Vol. 24, pp. 351-381). Elsevier, ISBN10: 0-444-81815-4

Kusiak, A. (1987a). The generalized group technology concept. *International Journal of Production Research*, *25*(4), 561-569, ISSN: 0020-7543.

Kusiak, A. (1987b). The generalized group technology concept. *International Journal of Production Research*, *25*(4), 561-569m, ISSN: 01663615.

McAuley, J. (1972). Machine grouping for efficient production. *Production Engineer*, *51*(2), 53-57, ISSN: 0032-9851 .

Moon, C., Gen, M., & Suer, G. A. (1999). A genetic algorithm-based approach for design of independent manufacturing cells. *International Journal of Production Economics*, *60-61*(2), 421-426, ISSN: 09255273.

Purcheck, G. F. K. (1974). A mathematical classification as a basis for the design of group-technology production cells. *Production Engineer*, *54*(1), 35-48, ISSN: 0032-9851.

Rajamani, D., Singh, N., & Aneja, Y. P. (1990). Integrated design of cellular manufacturing systems in the presence of alternative process plans. *International journal of production research*, *28*(8), 1541-1554. Taylor & Francis, ISSN: 0020-7543.

Saad, S. (2003). The reconfiguration issues in manufacturing systems. *Journal of Materials Processing Technology*, *138*(1-3), 277-283, ISSN: 09240136.

Seifoddini, H. (1990). A probabilistic model for machine cell formation. *Journal of Manufacturing Systems*, *9*(1), 69-75, ISSN: 02786125.

Shafer, S., & Rogers, D. (1991). A goal programming approach to the cell formation problem. *Journal of Operations Management*, *10*(1), 28-43, ISSN: 02726963.

Shtubt, A. (1989). Modelling group technology cell formation as a generalized assignment problem. *International Journal of Production Research*, *27*(5), 775 - 782. Taylor & Francis, ISSN: 00207543.

Silver, E. A., & Peterson, R. (1985). *Decision Systems for Inventory Management and Production Planning (Wiley Series in Production/Operations Management)* (p. 736). Wiley, ISBN: 0471867829.

Sofianopoulou, S. (1999). Manufacturing cells design with alternative process plans and/or replicate machines. *International Journal of Production Research*, *37*(3), 707-720. Taylor & Francis, ISSN: 0020-7543.

Solimanpur, M., Vrat, P., & Shankar, R. (2004). A multi-objective genetic algorithm approach to the design of cellular manufacturing systems. *International Journal of Production Research*, *42*, 1419-1441, ISBN: 978-1-4244-3334-6.

Suer, G. A., Huang, J., & Sripathi, M. (2010). Design of dedicated, shared and remainder cells in a probabilistic demand environment. *International journal of production research*, *48*(19-20), 5613-5646. Taylor & Francis, ISSN: 0020-7543

Suer, G. A., Pena, Y., & Vazques, R. (2003). Cellular design to minimise investment costs by using evolutionary programming. *International Journal of Manufacturing Technology and Management*, *5*(1), 118-132, DOI: 10.1504/IJMTM.2003.002536

Süer, G. A., Mese, E. M., & Eḡilmez, G. (2011). Cell Loading, Family and Job Scheduling to minimize TT. *International Conference on Intelligent Manufacturing Logistics Systems, February 27-March 1, 2011, Chung-Li, Taiwan.*

Wang, J. (1998). A linear assignment algorithm for formation of machine cells and part families in cellular manufacturing. *Computers & Industrial Engineering*, *35*(1-2), 81-84, ISSN: 03608352

Wei, J. C., & Gaither, N. (1990). An Optimal Model for Cell Formation Decisions. *Decision Sciences*, *21*(1), 416-433.

Wemmerlov U., & Johnson D. J. (1997). Cellular manufacturing at 46 user plants: implementation experiences and performance improvements. *International Journal of Production Research*, *35*(1), 29-49. Taylor and Francis Ltd, DOI: 10.1080 /002075497195966

Wicks, E. M., & Reasor, R. J. (1999). *Designing cellular manufacturing systems with dynamic part populations*. *IIE Transactions* (Vol. 31, pp. 11-20-20). Springer Netherlands, DOI: 10.1023/A:1007516532572.

20

Facility Layout

Xiaohong Suo
China Institute of Industrial Relations, Beijing, P.R.China

1. Introduction

The facility layout problem is concerned with finding the most efficient non-overlapping arrangement of n indivisible departments with unequal area requirements within a facility. Generally, about 20%-50% of the total operating expenses in manufacturing are attributed to material handling costs. Effective facility layout could reduce these costs by 10%-30% annually. Moreover, good facility planning could also improve the material handling efficiency, reduce the throughput time, decrease the space utilization area of manufacturing system, etc. So, the facility layout affects the total performance of manufacturing system, such as, material flow, information flow, productivity, etc.

Facility layout, being a significant contributor to manufacturing performance, has been studied many times over the past decades. Raman et al. showed that facility layout has a direct impact on operational performance, as measured by manufacturing lead time, throughput rate, and work-in-process (WIP).

2. Classification of facility layout problem

It is well known, facility layout problem is concerned with the allocation of activities to space such that a set of criteria are met and/or some objectives are optimized. There are numerous derivations for the facility layout problems in manufacturing systems, which have been investigated in Table 1. These derivations can be classified into six categories (product, process, equipment, production, manufacturing system and company). Any changes in items of these six categories can lead to facility layout problem. Once one item changes, other items will change correspondingly, e.g., the introduction of new products results in changes in process and equipment. Generally, combinations of items of these six categories are the derivations for the facility layout problem.

When the flows of materials between the departments are fixed during the planning horizon, facility layout problem is known as the static (single period) facility layout problem (SFLP). Researchers had paid more attentions to SFLP, and now SFLP has two new trends. With more fierce competitive in the global market, facility layout must react on the changes in designs, processes, quantities, scheduling, organizations, and management idea rapidly. Once these items change frequently, manufacturing systems must be reconfigurable and their structure must be modified as well. However, SFLP can hardly meet this demand. Company need to design a flexible layout which is able to modify and expand easily the original layout. Flexibility can be reached by modular devices, general-purpose devices and material handling devices. The trends of SFLP are shown in Fig.1. Under a volatile environment, SFLP need to add flexibility to meet the production requirement. Approaches to get flexibility for SFLP include to modify the SFLP and to increase the robustness of the SFLP. Gradually, SFLP develops these two approaches to the dynamic facility layout problem (DFLP) and robust layout, respectively.

Up to now, there are existing three basic types of layout problem, including SFLP, DFLP and robust layout problem. Research on the relationships among SFLP, DFLP and robust layout is important due to the impact of the types of layout problem on productivity, quality, flexibility, cost, etc. How to select the suitable type of layout problem is an urgent task. The classification procedure of facility layout problem is shown in Fig.2, where researchers choose the appropriate type of layout problem based on the judgment conditions. The judgment conditions include whether the material handling flows change over a long time or not, and whether it is easy for rearrangement or not when the production requirement changes drastically. If the material handling flows change over a long time, choose DFLP or robust layout; if not, choose SFLP. If rearrangement is easy when the production requirement change drastically, choose DFLP; if not, choose robust layout.

No. Classify	1	2	3	4	5
Product	Increase or decrease in the demand for a product	Addition or deletion of a product	Changes in the design of a product	Introduc-tion of new products	
Process	Changes in the design of process	Replacement of characteristics of process	Installation of new processes		
Equipment	Installation of new equipment	Replacement of one or more pieces of equipment			
Production	Failure to meet schedules	High ratio of material handling time to production time	Excessive temporary storage space	Bottlenecks in production	Crowded conditions
Manufacturing systems	Conflict between productivity and flexibility in general manufacturing systems	Flexibility does not meet the demands of changes in product mixes of FMS.			
Company	Adoption of a new safety standard	Organizational changes within the company	A decision to build a new plant		

Table 1. Derivations for facility layout in manufacturing systems

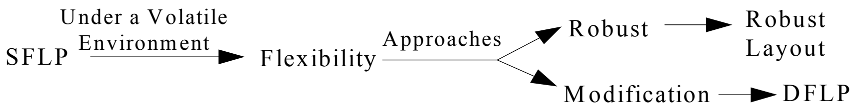


Fig. 1. Trends of SFLP

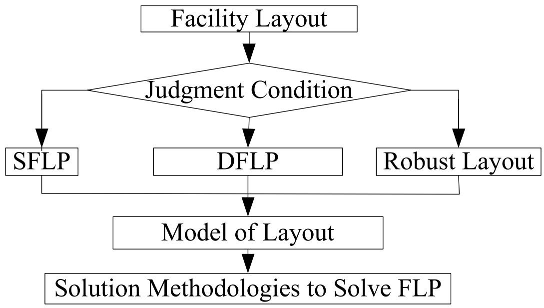


Fig. 2. Classification procedure of facility layout problems

2.1 Relationships among SFLP, DFLP and robust layout

Due to the impact of the locations of facility on material handling costs, throughput, and productivity of the facility, facility layout is an important module of manufacturing systems design. The FLP is the arrangement of departments within a facility with respect to some objective. The most common objective considered is the minimization of material handling cost. Material handling costs are determined based on the amounts of materials that flow between the departments and the distances between the locations of the departments. SFLP is appears when the flows of materials between departments are fixed during the planning horizon, which can be formulated as a quadratic assignment problem (QAP). So, SFLP is used under the static environment. When the flows of material between departments change during the planning horizon, this problem is known as the dynamic (multiple-period) facility layout problem (DFLP) [1]. Therefore, DFLP is widely used when the condition is changeable and the future demand of product can be forecasted. A robust layout is one that is good for a wide variety of demand scenarios even though it may not be optimal under any specific ones [2]. A robust layout procedure considers minimizing the total expected material handling costs over a specific planning horizon. Robust layout is selected when the demand is stochastic and the re-layout is prohibited.

2.1.1 SFLP vs. DFLP

SFLP Converting to DFLP.

Fiercer competition of the world makes SFLP covert to DFLP. Under today's changeable market situation, demand is changed irregularly from one production period to another. Generally, 40% of a company's sales come from new products, *i.e.* products that have only recently been introduced [3]. When these changes frequently occur and the location of an existing facility is a decision variable, SFLP convert to DFLP. The procedure of SFLP converting to DFLP is given in Fig. 3. The changes in product, process, equipment, etc. can bring on the facility layout problem. If the material flows are consumed to be constant, SFLP is sufficient. However, this assume are contradiction with the practice production. In order to correct the deficiency, SFLP will be converted to DFLP.

During the process of SFLP converting to DFLP, rearrangement costs arise. The changes in locations of facility can reduce the material flows between department pairs during a planning horizon. Meanwhile, rearranging the locations of facility will result in some shifting (rearrangement) costs depending on the departments involved in this shift. The procedure for rearrangement costs is illustrated in Fig. 4. Generally, when the products change often and the facility location is static, the material flows are increased drastically. In order to reduce the material flows, the facilities are shifted to different location which will result in the rearrangement (shifting) costs. The DFLP is based on the anticipated changes in flow that will occur in the future. Moreover, the future can be divided into any number of time periods, and a period may be defined in months, quarters, or years. In addition, different periods can be of different lengths [4].

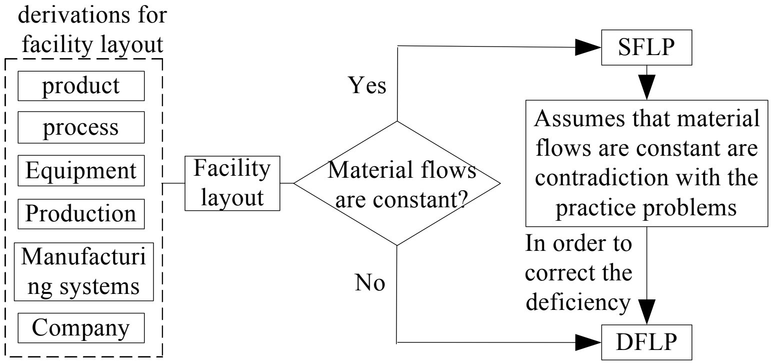


Fig. 3. Procedure of SFLP converting to DFLP

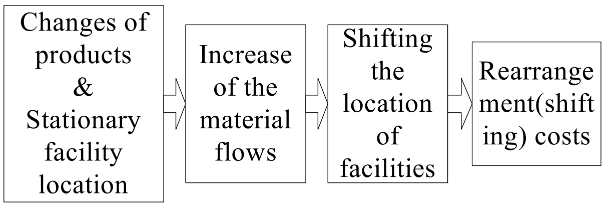


Fig. 4. Procedure for rearrangement costs in DFLP

SFLP for each period and rearrangement are the two parts of DFLP. For DFLP, it is assumed that the flow data during each period remains constant, respectively. Therefore, the facility layout during one period in the planning horizon can be obtained by solving the SFLP for each period. However, the flow data in whole planning horizon are changeable in DFLP. There exist rearrangement costs between the layouts for each pair of adjacent periods. That is to say, DFLP is composed of a series of SFLP if the rearrangement costs can be neglected. Therefore, DFLP involves selecting a SFLP for the first period and then deciding whether to change to a different SFLP in the next period. The dynamic layout shows flow dominance.

For each period, some departments have higher material handling inflows than the others. During the adjacent periods, if the higher material handling inflows do not change, the same SFLP will be used in these two periods. If these flow dominant departments change during the adjacent periods, changing to a different SFLP will occur in the following period.

For DFLP, the cycle of rearrangement depends on the rearrangement costs. If the rearrangement costs are relatively low, the layout configuration would tend to change more often to retain material handling efficiency. The reverse is also true for high rearrangement costs. The structural diagram of DFLP is given in Fig.5. Table 2 gives the comparison of main characteristics for SFLP and DFLP.

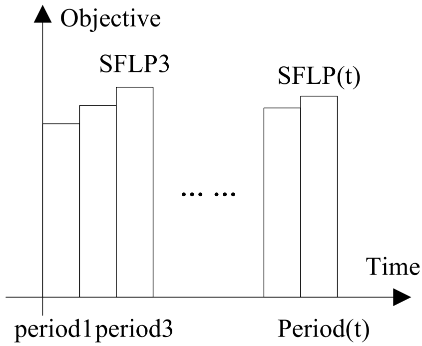


Fig. 5. Structural diagram of DFLP

	Period number	Rearrangement	Optimize objectives	Method for generating static layout of each period
SFLP	One period	No rearrangement	Minimize the material handling costs	The best static layout
DFLP	Multiple periods	Rearrangement costs	Minimize the material handling costs and rearrangement costs	The best static layout, or the random layout, or mixed layout*

*Mixed layout: combination of random layouts and the best layouts, each contributing half of the layout selected.

Table 2. Comparison between SFLP and DFLP

Objective Function of DFLP.

The objective function of a DFLP is generally defined as the minimization of the total costs, material handling costs for a series of SFLP plus rearrangement costs between periods. In each period, material handling costs among departments are calculated by the product of the probability of a flow matrix occurring in that period, the associated flows, and the distances. The formulation of the DFLP is given below.

$$\text{Min DFLP} = \sum_{t=1}^{T} \text{SFLP}(t) + \sum_{t=1}^{T-1} A_{t(t+1)} \tag{1}$$

where t is the number of periods in DFLP; $A_{t(t+1)}$ is the rearrangement costs between each pair of adjacent periods; and T is the total number of periods in the whole period horizon. Rearrangement costs are incurred when moving machines (or departments) from one location to another in order to minimize material handling costs in consecutive periods. Rearranging the layout will result in some shifting costs depending on the departments of the layouts. Rearranging includes the changes in location and orientation [5]. Each of the two aspects for facility will impact the total rearrangement costs. When the location and orientation of a facility in the new layout are the same as in the existing layout, no rearrangement cost is incurred for that facility. Otherwise, if the location or orientation of a facility has changed, then the specific facility rearrangement costs will be added to the objective function.

There are many approaches to calculate rearrangement costs for DFLP. Furthermore, the rearrangement of departments may lead to production loss, and it may also require specialized labor and equipment. Therefore, rearrangement costs consist of labor cost, equipment cost, out-of-pocket moving expenses and the cost of operational disruptions. Generally, the rearrangement (shifting) costs may be viewed as fixed costs, or a linear function of the distance between the various locations, or the linear function of square-feet being rearranged, or variable costs associated with moving a particular facility in a given period, or the accumulation of fixed costs due to changing in facility configuration, interrupting or disrupting production, using personnel and equipment to move the facility, or any combination of the above [5,6]. One special application of the DFLP is the rearrangement of existing facilities. In a rearrangement problem, the first period is the current SFLP and subsequent periods are the revised layouts.

Computation Complexity of DFLP.

As the SFLP, the DFLP is also the computationally intractable problem. In other words, the number of possible solutions or layout plans is $N!$ for a SFLP instance with N departments, while $(N!)^T$ for a DFLP instance with N departments and T periods. For this reason, with the same computer's configuration, only small problems can be optimally solved in reasonable computation time for DFLP while large and media problems can be solved for SFLP. The numbers of possible solutions and their methodologies for SFLP and DFLP are listed in Table 3.

For instance, even for a six-department, five-period problem, $(N!)^T$ is 1.93×1014 combinations. Thus for large problems obtaining optimal solutions is not nearly possible. So in practices, for small problem, $n=N!$; for large problem, $n<<N!$, where n is the number of static layout during each period in DFLP. n in each period depends on the capability of the software and hardware used to solve the DFLP. The more power these are, the larger n can be selected. Logically, larger n should lead to better solutions.

	Number of solutions	Size of problems	Methodologies
SFLP	$N!$	Small	Exact algorithm
		Media and large	Heuristic or meta-heuristic algorithm
DFLP	$(N!)^T$	Small and media	heuristic or meta-heuristic algorithm

Table 3. Numbers of solutions and methodologies for SFLP and DFLP

DFLP Degenerates into SFLP.

The SFLP has one period and no rearrangements, so it is just a special case of the DFLP. Under some conditions, DFLP may degenerate into SFLP. In those environments where material handling flows do not frequently change over a long time, SFLP analysis would be sufficient. When the rearrangement costs are negligible, dynamic layout analysis is not necessary. In the other cases, if the rearrangement costs are prohibitive, such as in the case of very heavy machinery, the same layout is used for the total planning horizon. In this situation, DFLP is also not necessary.

2.1.2 DFLP vs. robust layout

Robustness is defined as the frequency that a layout falls within a pre-specified percentage of the optimal solution for different sets of production scenarios [7]. For a robust layout, it is good for a wide variety of demand scenarios even though it may not be optimal under any specific demand scenarios. The objective of a robust layout is trying to minimize the total material handling costs over the specific planning horizon. Fig.6 illustrates the robust layout design framework [6].

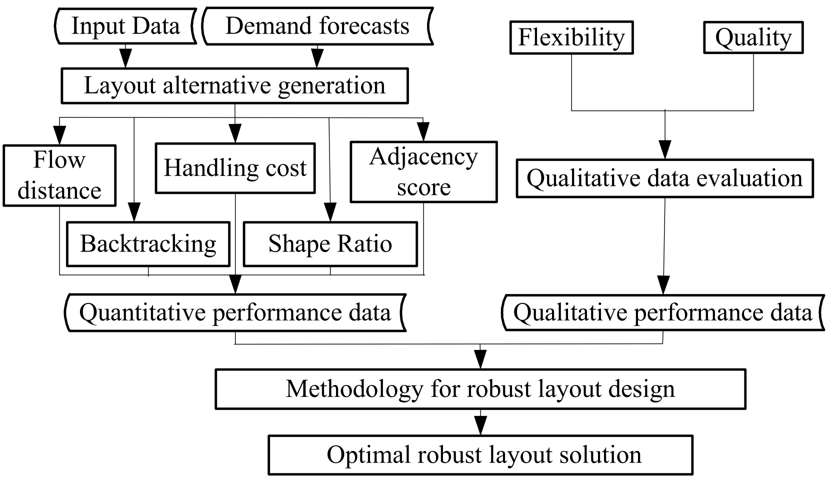


Fig. 6. Robust layout design framework

To some extent, robust layout is similar to SFLP. However, there is a difference between them: robust layout is researched under dynamic environment, and SFLP is assumed that the environment is static. Under the volatile environment, if the total planning horizon is divided into a lot of periods, DFLP is considered to be composed of a series of SFLP. Under some condition, if the number of periods is small over the planning horizon, *i.e.* each period has long time, DFLP is considered to be composed of a series of robust layout based on the definition of robust layout.

For DFLP, a set of planning time are referred to as a consecutive period with a layout rearrangement occurring only at the beginning of each period. The number and length of these periods are determined based on the trade-off between material handling costs and facility rearrangement costs. If rearrangement costs are larger than material handling costs, the number of these periods will be small and length will be long. The reverse is also true. Thus, the aim of DFLP is to modify the layout at the beginning of each period, but not to change the layout within these periods. In Contrast to DFLP, when rearrangement costs are extremely high, a pure robust layout will be selected, and it has the period equal to a total planning horizon. DFLP and robust layout are compared in Table 4. Robust layout will have different demand levels in the total planning horizon and will choose the demand level to minimize the material handling costs, adjacency scores and backtracking costs.

	Selective conditions	Period number	Function for material handling costs
DFLP	if production requirements change drastically and the rearrangement is easy	Multiple periods	Between the lower bound and upper bound on the expected material handling costs
Robust layout	if machine rearrangement costs are high	One period equal to the total planning horizon	Provide an upper bound on the expected material handling costs

Table 4. Comparison between DFLP and robust layout

As described in the last column of Table 4, robust layout provides an upper bound on the expected material handling costs while creating a new layout for each period provides a lower bound [8]. But in practical problems, creating a new layout for each period is unrealized. Therefore DFLP, which material handling costs are between the lower bound and upper bound, is necessary.

2.2 Discussions

The comparison among SFLP, DFLP and robust layout is presented in Table 5.

The fifth column of Table 5 refers the converting conditions of SFLP, DFLP and robust layout with each other. When the flows between department pairs are changeable, the SFLP will convert to DFLP or robust layout. When rearrangement costs are negligible, DFLP will convert to SFLP. However, when rearrangement costs are extremely high, DFLP will convert to robust layout. Under stochastic demand and when forecasting the future demand of products is difficult, robust layout will convert to DFLP. Since studies show that material handling costs make up 20-50% of the total operating costs and 15-70% of the total costs of manufacturing a product [9], the most common objective considered are the minimization of material handling costs. For DFLP, its objective function involves material handling costs and rearrangement costs. However, the objective of robust layout includes the total material handling costs for all product mix in total planning horizon.

The connection and difference between each pair of the types of layouts are described in Table 6.

Since assuming that the material handling flows are constant for SFLP, uncertainly of future production requirements are relatively low. DFLP is suitable for dynamic environment, so its uncertainty of future production requirements is high. Although SFLP and robust layout both have one period and no rearrangement, the application scopes are different-one for static environment and the other for dynamic environment. DFLP and robust layout are both used under the dynamic and changeable environment, but DFLP considers the rearrangement and robust layout does not consider for only using one facility layout solution during the whole planning horizon.

	Application scope	Advantages	Disadvantages	Converting condition	Objective function
SFLP	Flows between pairs of departments do not change over a long time	Widely used; low computational complexity; modeling easily	Flows between departments are assumed to be constant over time	Flows between department pairs are changeable	$\min \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} f_{ij} d_{ij}$ $i, j = 1, 2, ..., n$
DFLP	The relative material flows between departments change over time	Moving expenses and operational disruption	High computational complexity	Rearrange-ment costs are either negligible or extremely high	$\min(\sum_{t=1}^{P} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} f_{tij} d_{tij} X_{t}$ $+ \sum_{i=1}^{t-1} A_{t(t+1)})$
Robust Layout	Under stochastic demand and the re-layout is prohibited	Offer low mean and variance in distance traveled; lower material handling costs over large changes in product demand	It is not the best optimal solution for a special product.	Forecasting the future demand of products is difficult	$\min \sum_{m=1}^{M} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{mij} f_{mij} d_{m}$ $i, j = 1, 2, ..., n;\ m$ product mix

Table 5. Comparison among SFLP, DFLP and robust layout

<table>
<tr><th></th><th>SFLP</th><th>DFLP</th><th>SFLP</th><th>Robust layout</th><th>DFLP</th><th>Robust layout</th></tr>
<tr><td>Connection</td><td colspan="2">SFLP is the base and the special case of DFLP</td><td colspan="2">One period and no rearrangement</td><td colspan="2">In the dynamic and changeable environment</td></tr>
<tr><td rowspan="2">Difference</td><td colspan="2">Uncertainty of future production requirements</td><td rowspan="2">Static environ-ment</td><td rowspan="2">Dynamic environ-ment</td><td rowspan="2">Multiple periods and rearrangement</td><td rowspan="2">One period and no rearrangement</td></tr>
<tr><td>Low</td><td>High</td></tr>
</table>

Table 6. Connection and difference between each pair of the types of layouts

2.3 Summaries

In this part, the relationship among SFLP, DFLP and robust layout are researched. The characteristics of SFLP, DFLP are analyzed first. Then DFLP and robust layout are compared. The research results are given as followings:

1. The application scope of DFLP is different from SFLP and robust layout.

2. DFLP can convert to SFLP or robust layout in some conditions.
3. Among SFLP, DFLP and robust layout, the majority of practical problems will be classified to DFLP.
4. SFLP is the base of DFLP and robust layout.
5. SFLP and robust layout both are the special case of DFLP.

3. Objectives and constraints of FLP

The facility layout problem, block layout, considers the assignment of facilities to locations so that the quantitative (qualitative) objective of the problem is minimized (maximized) under various constrains.

3.1 Objectives of FLP

Traditionally, there are two basic types of objectives for FLP [10]. The first one is the quantitative (distance-based) objective aiming at minimizing the total material handling cost between departments based on a distance function. The distance-based objective, considers all distance pairs, but due to department areas, inter-department distances may be misleading. To help relieve this concern, distances have been measured in a variety of ways: from department centroid-to-centroid, expected distance, distance from department boundaries, distance along the material handling network, etc. Even so, since the choice between general-purpose and special-purpose material handling devices may depend on whether or not departments are adjacent, the same inter-department distance may not have the same material handling cost.

In general, the actual cost to move a unit load of material between two departments will be the sum of a fixed cost and a variable cost. The fixed cost is dependent on the waiting time to obtain the appropriate material handling method, the time to pickup and deposit the unit load, and possibly some charge for the initial purchase cost of the material handling method. The variable material handling cost is dependent on the distance the unit load travels. In single-floor facilities this has a near-linear relation, while in multi-floor facilities it is a non-linear relationship.

The second one is the qualitative (adjacency-based) goal, aimed at maximizing the closeness relationship scores between departments based on the placement of departments that utilize common materials, personnel, or utilities adjacent to one another, while separating departments for reasons of safety, noise, or cleanliness. The adjacency-based objective, if interpreted from a material handling cost perspective, is based on the assumption that the material handling costs between two departments are reduced significantly when the two departments are adjacent. The adjacency-based objective appears to assume that fixed costs dominate the total costs (to the extent we can ignore the variable material handling costs) and that more efficient (and less costly) material handling methods may be used when departments are adjacent. On the other hand, the distance-based objective models the variable material handling costs and ignores the fixed costs.

Over the years, extensive research has been conducted on FLP. Yet, most of the research conducted in this field has concerned a single objective, either qualitative or quantitative goodness of the layout. In general, minimization of the total material handling costs is often used as the optimization criterion in FLP. However, closeness, hazardous movement or safety, and similar criteria are also important in FLP. Inherently, real-life layout problems are multi-objective by nature and the FLP must consider both quantitative and qualitative objectives simultaneously. Consequently, FLP falls into the category of multi-objective optimization problem.

3.2 Constraints of FLP

Facility layout plays a crucial role in determining the throughout time of a manufacturing process. The objective of the facility layout problem in manufacturing environment is the arrangement of facilities on a floor shop [11], subject to the following constraints:

1. to reduce the flows among all facilities;
2. to have a regular flow of the parts and products not permitting bottleneck in the production;
3. to rationalize the space occupied by the facilities;
4. to permit flexibility considering that with the technological progress and the new demands in the market, facilities could be added or changed.
5. to locate in a specified location.

Facilities are including machines, departments, storage equipments, factory, material-handling systems, commerce and warehouse. In the manufacturing system it may be distinguished machines, material handling systems and storage equipments.

4. Mathematic formulation of facility layout problem

For the past decades researchers have been working on facility layout problem while considering various aspects which vary with the nature of production demand, shape of the facilities, number of floors, and nature of material flow. Despite these variations, the process of obtaining optimal solutions involves two steps: modeling the facility layout problem, and developing a solution approach. Modeling helps clearly define the problem and consider the factors that are imperative in developing layouts.

The facility layout problem is one of the best-studied problems in the field of combinatorial optimization. A number of formulations have been developed for this problem. Models are categorized depending on their nature, assumptions and objectives. More particularly the FLP has been modeled as quadratic assignment problem (QAP), quadratic set covering problem (QSP), linear integer programming problem (LIP), mixed integer programming problem (MIP), and graph theoretic problem.

4.1 Quadratic assignment problem (QAP)

Koopmans and Beckman were the first to model the problem of locating plants with material flow between them as a quadratic assignment problem (QAP) in 1957 [12]. The name was so given because the objective function is a second-degree function of the variables and the constraints are linear functions of the variables. More specifically, it is an NP-hard problem and one of frequently used formulation to solve FLP.

Consider the FLP of allocating a set of facilities to a set of locations, with the objective to minimize the cost associated not only with the distance between locations but with the flow also. Each location can be assigned to only one facility, and each facility can be assigned to only one location. There is material flow between the different departments and cost (material handling) associated with the unit flow per unit distance. Thus, different layouts have different total material handling costs depending on the relative location of the facilities. F_{ik} is the flow between facilities i and k, and D_{jl} is the distance between location j and l. The FLP has been formulated as follows:

$$\sum_{\substack{i=1 \\ i\neq k}}^{n}\sum_{\substack{j=1 \\ j\neq l}}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} F_{ik} * D_{jl} * X_{ij} * X_{kl} \tag{2}$$

$$s.t. \ \sum_{i=1}^{n} X_{ij} = 1 \qquad \forall \ j = 1,...,n \tag{3}$$

$$\sum_{j=1}^{n} X_{ij} = 1 \qquad \forall \ i = 1,...,n \tag{4}$$

$$X_{ij} \in \{0,1\} \qquad \forall \ i, j = 1,...,n \tag{5}$$

$X_{ij} = 1$ if facility i is assigned to location j and $X_{ij} = 0$ if facility is not assigned to location j, where n is the number of facilities. Equation (2) seeks to minimize the sum of flow multiplied by the distance for all pairs of facilities in a given layout. Equation (3) ensures that each location contains only one facility while equation (4) ensures that each facility is assigned to only one location.

4.2 Quadratic set covering problem (QSP)

Bazaraa formulated facility layout problem as a quadratic set covering model in 1975 [13]. In this formulation, the total area occupied by all facilities is divided into a number of blocks where each facility is assigned to exactly one location and each block is occupied by at most one facility. The distance between the locations is taken to be from centriods of the locations and the flow between facilities is minimized. The disadvantage of this approach is that the problem size increases as the total area occupied by all the facilities is divided into smaller blocks.

4.3 Linear integer programming problems (LIP)

Several integer programming formulations have been proposed for the facilities layout problem. Lawler was the first one to formulate the FLP as a linear integer programming model [14]. He proved that his model is equivalent to QAP. QAP has n^2 k_{ij} variables and $2n$ constraints while integer programming problem has n^4+2n+1 constrains and n^4 Y_{ijkl} while n is the number of locations, X_{ij} is the integer variable of facility i at location j, and Y_{ijkl} is the integer variable of facility i at location j in arrangement k of location l.

Assumption $y_{ijkl} = x_{ij} \cdot x_{kl}$

$$\text{Min}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} b_{ijkl} y_{ijkl} \tag{6}$$

$$s.t. \quad \sum_{j=1}^{n} x_{ij} = 1 \,, \; i = 1,2,...,n; \tag{7}$$

$$\sum_{i=1}^{n} x_{ij} = 1 \qquad j = 1,2,...,n; \tag{8}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} y_{ijkl} = n^2 \tag{9}$$

$$x_{ij} + x_{kl} - 2y_{ijkl} \geq 0 \; i,j,k,l = 1,2,...,n \tag{10}$$

$$x_{ij} \in \{0,1\} \qquad i,j = 1,2,...,n. \tag{11}$$

$$y_{ijkl} \in \{0,1\} \; i,j,k,l = 1,2,...,n \tag{12}$$

$$b_{ijkl} = \begin{cases} f_{ik} c_{jl} + a_{ij} & i = k, j = l \\ f_{ik} c_{jl} & i \neq k \;\; or \;\; j \neq l \end{cases} \tag{13}$$

4.4 Mixed integer programming problems (MIP)

Kaufman and Broeckx developed a linear mixed integer programming model in 1979 [15], which has the smallest number of variables and constraints among all integer programming formulations of the QAP. The equivalence between QAP and the mixed integer programming has been proposed through this model.

Assumption $w_{ij} = x_{ij}\sum_{k=1}^{n}\sum_{l=1}^{n} b_{ijkl} x_{kl}$, $e_{ij} = \sum_{k=1}^{n}\sum_{l=1}^{n} b_{ijkl}$

$$\text{Min}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} b_{ijkl} x_{ij} x_{kl} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}\left(\sum_{k=1}^{n}\sum_{l=1}^{n} b_{ijkl} x_{kl}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \tag{14}$$

$$s.t \; \sum_{j=1}^{n} x_{ij} = 1 \;\; i = 1,2,...,n; , \tag{15}$$

$$\sum_{i=1}^{n} x_{ij} = 1 \qquad j = 1,2,...,n; , \tag{16}$$

$$e_{ij} x_{ij} + \sum_{k=1}^{n}\sum_{l=1}^{n} b_{ijkl} x_{kl} - w_{ij} \leq e_{ij} \tag{17}$$

$$w_{ij} \geq 0 \quad i, j = 1, 2, ..., n. \tag{18}$$

$$x_{ij} \in \{0,1\} \quad i, j = 1, 2, ..., n. \tag{19}$$

4.5 Graph theoretic formulations

In graph theoretic formulations it is assumed that the desirability of locating each pair of facilities adjacent to each other is known [16]. In this model a closeness rating indicating desirability of locating facility *i* adjacent to facility *j* is assumed. The model seeks to maximize the closeness rating of the facilities.

5. Solution methodologies for facility layout problem

Several researches have been done in the facility layout problem. The solution methodologies for FLP can be divided into exact algorithms, heuristics and meta-heuristic algorithms [17]. The exact methods such as the branch-and-bound and cutting plane algorithm have been successfully applied to FLP when the number of facilities is less than 16. However, when the number of facilities is larger than 16, FLP cannot be solved optimally in reasonable time. In order to obtain good (near optimal) solution in a reasonable computational time, heuristics were developed. Recently, meta-heuristic approaches such as simulated annealing (SA), genetic algorithms (GA), tabu search (TS), and colony optimization have been successfully applied to solve large FLP.

5.1 Exact algorithms

Exact algorithms are clever version of exhaustive search approach. Branch-and-bound and cutting plane algorithms are used to solve the FLP modeled as QAP optimally. These exact algorithms are complete in the sense that the existence of a feasible solution and then the optimal solution can be determined with certainty once such exact algorithm is successfully terminated. The main disadvantage of these exact algorithms is that they entail heavy computational requirements when applied even to small size problems.

5.1.1 Branch and bound algorithms

Branch and bound methods are used to find an optimum solution of quadratic assignment formulated FLP because QAP involves only binary variables. In branch and bound algorithms, the solution procedure proceed on the basis of stage by stage or parallel search of single assignment or pairs of assignments of facilities to locations. At each stage back tracking occurs, certain assignments are excluded and the forward search process is resumed.

Only optimal solutions up to a problem size of 16 are reported in literature. Beyond n=16 it becomes intractable for a computer to solve it and, consequently, even a powerful computer can not handle a large instance of the problem.

5.1.2 Cutting plane algorithms

Cutting plane methods are exact algorithms for integer programming problems. They have proven to be very useful computationally in the last few years, especially when combined with a branch and bound algorithm in a branch and cut framework. Cutting plane algorithms work by solving a sequence of linear programming relaxations of the integer programming problem. The relaxations are gradually improved to give better approximations to the integer programming problem, at least in the neighborhood of the optimal solution. For hard instances that can not be solved optimality, cutting plane algorithms can produce approximations to the optimal solution in moderate computation times, with guarantees on the distance to optimality.

5.2 Heuristic algorithms

In order to obtain good (near optimal) solution in a reasonable computational time, heuristic algorithms were developed [18]. A heuristic algorithm can be defined as a well-defined set of steps for quickly identifying good quality solutions. The quality of a solution is defined by an evaluation criterion, e.g., minimize material handling cost, and the solution must satisfy the problem constraints. Basically, heuristic algorithms for FLP can be classified into four classes: construction algorithms, improvement algorithms, hybrid algorithms and graph theoretic algorithms.

5.2.1 Construction algorithms

Construction algorithms are considered to be the simplest and oldest heuristic approaches to solve the QAP, from a conceptual and an implementation point of view. A construction algorithm consists of successive selection and placement of facilities until a complete layout is achieved. These methods are probably the oldest ones, dating back to the early 60s. The simplicity of construction algorithm is often associated with poor quality of the resulting solutions.

But these construction algorithms can be used to provide initial solutions for improvement algorithms. Improvement methods start with a feasible solution and try to improve it by interchanges of single assignments.

5.2.2 Improvement algorithms

An improvement algorithm starts with an initial solution (existing layout). This existing layout is improved by exchanging the locations of a pair of facilities. The exchange, which produces the best solution, is retained and the procedure continues until the solution cannot be improved any further or until a stopping criterion is reached. Hence, the solution quality of improvement algorithms greatly depends on the initial layout provided, and the systematic procedure of the location exchange.

The greedy nature of pair-wise exchange makes it susceptible to converge to a local optimum. Therefore, the shortcomings of improvement algorithms originate not only from the initial solution provided but also from the greedy nature of the systematic exchange procedure. The greedy nature of the procedure is exposed because only the location exchanges, which result in the greatest cost reduction, are accepted. Hence, the nature of the exchange procedure often impedes the algorithm from finding the global optimum and causes the algorithm to converge to a local optimum.

Improvement methods can easily be combined with construction methods.

Improvement algorithms can be meta-heuristic such as SA and TS, which require one feasible solution as starting solution for the execution of these algorithms.

5.2.3 Hybrid algorithms

In hybrid algorithms the solution of QAP is determined by using a combination of two optimal or sub-optimal algorithms. Such combination of algorithms is essential in some cases to improve solution quality. This classification is extended to include certain algorithms, which use the principal of construction algorithms and improvement algorithms. FLAC and DISCON are examples of such hybrid algorithms.

5.2.4 Graph theoretic algorithms

Graph theoretic algorithms identify maximal planar subgroups of a weighted graph that show the relationships between the facilities [19]. The dual of a maximal planar sub graph determines the layout of the facilities. Seppanen and Moore proposed graph theoretic solutions procedure in which a heuristic algorithm, which uses this strategy, was also presented. The algorithm determines the maximum spanning tree based on the weighted graph. With the help of one edge adding process, the maximum spanning tree is the used to obtain a maximal planar sub graph. The dual of the maximal planar sub graph determines a layout of the facilities.

5.3 Meta-heuristic algorithms

The development of meta-heuristic algorithms has greatly influenced the performance of improvement algorithm and uses a general strategy like pair-wise exchange heuristic. There are three classes widely used of meta-heuristic algorithms in layout problem i.e. Simulated annealing (SA), tabu search (TS), and genetic algorithms (GA).

5.3.1 Simulated annealing algorithms (SA)

Simulated annealing (SA) is a general probabilistic local search algorithm, proposed by Kirkpatrick et al in 1983, to solve difficult optimization problems. Many large instances of difficult real life problems were successfully solved by simulated annealing algorithms. Its ease of implementation, convergence properties and its use of hill-climbing moves to escape local optima has made it a popular technique over two decades. SA is based on the analogy between the annealing of solids and the solving of combinatorial optimization problems [20]. SA is a step-by-step method which could be considered as an improvement of the local optimization algorithm. This process accepts not only better solutions but also worse solutions with a certain probability which is called the probability of accepting. The probability of accepting is determined by the temperature. The probability of accepting a worse solution is large at a higher temperature. As the temperature decreases, the probability of accepting a worse solution also decreases as well.

SA has advantages and disadvantages compared to other global optimization techniques, such as genetic algorithms, tabu search algorithms, and neural networks algorithms. Among its advantages are the relative ease of implementation and the ability to provide reasonably good solutions for many combinatorial problems. Though a robust technique, its drawbacks include the need for a great deal of computer time for many runs and carefully chosen tunable parameters.

5.3.2 Genetic algorithms (GA)

Genetic algorithms (GA) is a heuristic search that mimics the process of natural evolution, which encode a potential solution to a specific problem on a simple chromosome-like data structure and apply operators like mutation, recombination to create new data strings and to preserve critical information [21,22].

GA gained more attention during the last decade than any other evolutionary computation algorithms; it utilizes a binary coding of individuals as fixed-length strings over the alphabet {0, 1}.

Evolution, or more specifically biological evolution, is the change over time in one or more inherited traits of individuals. Natural selection, genetic drift, mutation, gene flow are the four corresponding common mechanisms of evolution. After a long enough time, only the adaptive individuals survive as a consequence of natural selection. To put it concisely, whether the individual should survive or not is decided by two factors, the gene in the individual and the fitness of the gene in the whole population. Mimicking the mechanism, genetic algorithm applies as a searching tool finding out the fittest individuals among a population. More often, the algorithm is viewed as a function optimizer, implementing first by defining two attributes of the individuals: the gene (a data string specifies the individual's character) and the fitness (a function evaluates the individual's vitality). Thus, the two main components of most genetic algorithms that are problem dependent are: the problem encoding (the gene of the individual) and the evaluation function (the fitness). Subsequently, certain operators like mutation and recombination are applied to select the fittest offspring after several generations as the finial individuals, the corresponding optimal answer to the problem when decoded.

GA iteratively search the global optimum, without exhausting the solution space, in a parallel process starting from a small set of feasible solutions and generating the new solutions in some random fashion.

5.3.3 Tabu search algorithms (TS)

Tabu search (TS) was proposed by Glover and has quickly become one of the most effective methods using local search techniques to find near-optimal solution to combinatorial optimization problems. It uses a deterministic local search technique which is able to escape local optima by using a list of prohibited neighbor solutions known as the tabu list. In addition to escaping local optima, using the tabu list can also prevent cycling by forbidding or penalizing moves which take the solution, in the next iteration, to points in the solution space previously visited, and thus save computational time.

A drawback of tabu search is that if it reaches a previously visited solution, it will cycle following the same path unless a tabu neighbor exists. In other words, if the search moves to a previously visited solution that has not been tabu for the last two iterations, then a loop is encountered.

5.3.4 Ant colony algorithms (ACO)

Recently, a few papers have appeared where an ant colony algorithm (ACO) has been attempted to solve large FLP. The first ACO system was introduced by Marco Dorigo in his Ph.D. thesis in 1992, and was called ant system.

ACO is a heuristic search technique to seek for an optimal path in a graph, inspired by the ability of ants to find food sources by using a substance called pheromone. Ant belongs to a colony leave the nest and randomly search for a food source. When an ant finds a food source, it returns to the nest to let others know about the source. On the way back to the nest, the ant places pheromone, which ants are sensitive to, to mark the path from the food source to the nest. Ants select their path to food sources according to the pheromone concentration on different paths. Pheromone evaporates over time, and this causes less frequently visited food sources to lose their address hence to be less visited by others ants. This mechanism has been the cornerstone to devise meta-heuristic algorithms for finding good solutions for difficult optimization problems.

5.4 Other approaches

The major drawbacks of the aforementioned approaches lie in the fact that the search for the best layout is not very efficient and the multi-objective nature are not considered in the problem. As a matter of fact, facility layout problem can be considered one of the truly difficult ill-structured, multi-criteria and combinatorial optimization problems. Many researchers still finding out for new and recent developments rather than conventional approaches to overcome the aforementioned drawbacks. Intelligent techniques such as expert systems, fuzzy logic and neural networks have been used as new advancements for the tackled problem.

6. Computer simulations

The typical absence of some encompassing, closed-form, and analytical fitness functions renders computer simulations a useful alternative. Such an approach would provide detailed analysis, modeling, and evaluation of complex layout design problems. However, simulation models are not easily amenable to optimization and make procurement of a superior layout alternative difficult to achieve. Recently, some efforts have been made to optimize layout design simulation models using genetic algorithms in various facility layout design contexts in order to expedite the process and procure a diverse set of superior in layout alternatives. Nevertheless, computer simulations are usually very time consuming and could become prohibitive in the facility layout design process.

7. Facility layout based on manufacturing costs

Facility layout is composed of product, its process routing, machine and some space. Different combinations of these entities and their activities affect the type of facility layout. Considering the criteria of material handling route, the types of facility layout are classified into three types: single-row layout, multi-row layout and loop layout, as shown in Fig. 6. The application scope, advantages and disadvantages are illustrated in Table 7. The single-row layout includes three shapes such as linear, U-shape and semi-circular. In the linear layout, there may exist bypassing and backtracking, as shown in Fig. 1(a). Backtracking is the movement of some parts from a machine to another machine that precedes it in the sequence of placed machines in a flow line arrangement. Bypassing occurs when a part skips some machines while it is moving towards the end of a flow line arrangement. Table 8 gives the comparison of main characteristics for backtracking and bypassing.

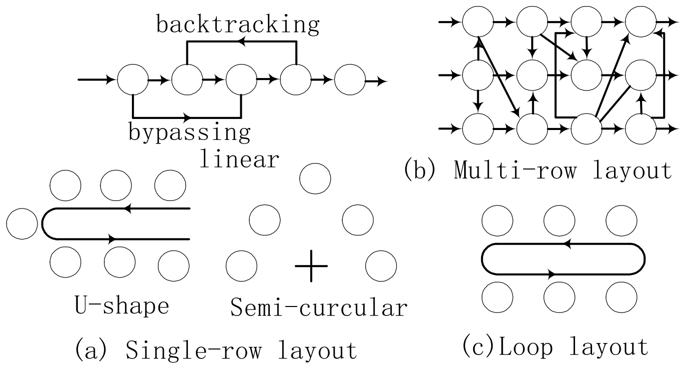


Fig. 6. Types of facility layout based on the criteria of material handling route

Type of layout		Application scope	Advantages	Disadvantages
Single-row layout	Linear U-shape Semi-circular	Within GT cells, in facilities that implement JIT, and sometimes with FMS	Material flow are moving along the sequence of operations of all the parts; small material handling cost and time; less delays; better control of operations; the ability to use conveyors.	When several parts having different sequence of operations are processed, the benefits of a flow line arrangement are reduced since the movement of parts may not always be unidirectional.
Multi-row layout		Suitable for FMS	Adjacent lines share common equipments; low investment; small space area; high machine utilization rate;	Complicated process management; coordinate multi-task difficulty.
Loop layout		Used in FMS	High flexibility in material handling system	

Table 7. Comparison of three types of layout

	Direction	Derivation	Disadvantages	Objective	Scope
Backtracking	Adverse sequence of operations in the flow line	The difference in the sequences of operations of the parts	Impacts the movement cost and productivity of facility	Should be minimize.	In traditional facilities
Bypassing	Same sequence of operations in the flow line	The same with above.	Unnecessary travel time and cost	The same with above	The same with above.

Table 8. Comparison of backtracking and bypassing

7.1 Problem statement

Selling price of products is the concerned problem for customers. Therefore how to decrease the selling price by effective layout planning is an important issue.

7.1.1 Manufacturing cost

Skinner provides the breakdown of costs for a manufacturing product [17], shown in Fig. 7. About 40% of the selling price of a product is manufacturing cost. Material and parts make up the largest percentage of total manufacturing cost, at round 50%. Direct labor is responsible for operating the facilities and is a relatively small proportion of total manufacturing cost: 12%. It is only about 5% of selling price. Machinery, plant and energy etc. are about 26% of manufacturing cost. Therefore, decreasing the manufacturing cost is the key to lower the selling price of products.

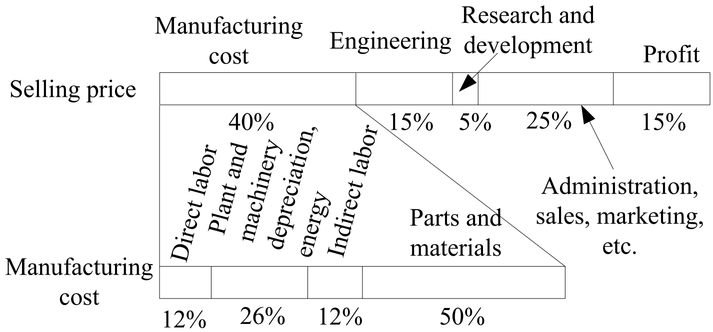


Fig. 7. Breakdown of costs for a manufactured product

Manufacturing cost has two classify methods. The first one classifies the manufacturing cost into fixed costs and variable costs. The second separates manufacturing cost into: (1) direct labor, (2) material, and (3) overhead. In this chapter, the second one is selected. The classification of manufacturing cost is shown in Fig. 8. The direct labor cost is the sum of the wages and benefits paid to the direct labor. The smaller the number of direct labor, the lower the manufacturing cost. The material cost is the cost of all raw material used to manufacture the parts or products. Overhead costs are all of the other expenses associated with running the manufacturing firm. Overhead divides into two categories: (1) factory overhead and (2) corporate overhead. Detail expenses of overhead costs are listed in Table 9.

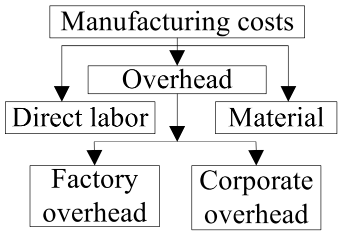


Fig. 8. Classification of manufacturing cost

Factory overhead costs		Corporate overhead costs	
Plant supervision	Applicable taxes	Corporate executives	Applicable taxes
Line foreman	Insurance	Sales and marketing	Cost of space
Maintenance crew	Heat and air conditioning	Accounting department	Security personnel
Custodial services	Light	Finance department	Heat and air conditioning
Security personnel	Power for machinery	Legal counsel	Light
Tool crib attendant	Factory depreciation	Engineering	Insurance
Material handling	Equipment depreciation	Research and development	Fringe benefits
Shipping and receiving	Fringe benefits	Other support personnel	Other office costs

Table 9. Typical overhead costs

Factory overhead consists of the costs of operating the factory other than direct labor and materials. Corporate overhead is the cost of running the company other than its manufacturing activities. As shown in Table 9, material handling cost in factory overhead and cost of space in corporate overhead are the two parts which relate to the facility layout. When material handling cost and cost of office space increase, the manufacturing cost increase accordingly.

7.1.2 Objectives of facility layout based on manufacturing cost

Groover observes that materials spend more time waiting or handling than in process [23]. His observation is illustrated in Fig. 9. Only 5% of the time is spent on the machine. About 95% of a part's time is spent either moving or waiting. This figure shows that the material handling and storage are significance in a typical factory. Furthermore, studies show that material handling cost makes up 20%-50% of the total operating cost and 15%-70% of the total manufacturing cost. Therefore, the most common objective of facility layout is the minimization of material handling cost.

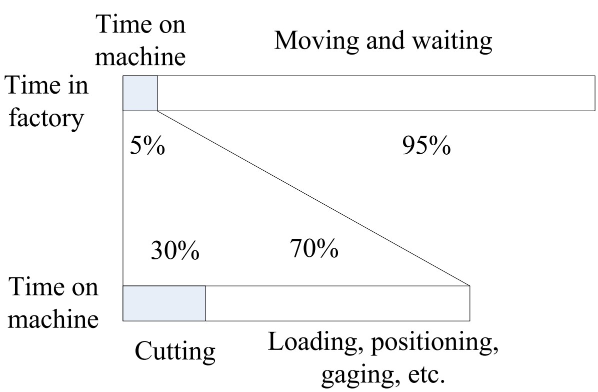


Fig. 9. How time is spent by a typical part in batch production machine shop

When the cost of space increases, the manufacturing cost increase accordingly. With the given purchase price of unit space area, the higher the area utilization rate is, the lower the manufacturing cost is. So, one objective of facility layout is to maximize the area utilization rate.

Combination of the Fig. 7 and 8 show that other than materials and parts, direct labor and machinery are main parts of manufacturing cost. The higher the utilization rate of direct labor and machinery are, the lower the manufacturing cost is. Increased utilization of existing machinery could lead to smaller machine inventories since less machinery would be sitting unused, and the direct labor are the same. Hence the maximization of the utilization rate of direct labor and machinery are also the objective of facility layout.

To sum up, the objectives of facility layout based on manufacturing cost include: (1) minimizing the material handling cost, (2) maximizing the area utilization rate, and (3) maximizing the utilization rate of direct labor and machinery.

7.2 Model formulation

7.2.1 Material handling cost

For manufacturing facilities, material handling cost is the most significant measure for determining the efficiency of a layout and is most often considered. It is determined based on the flows of materials between departments and the distances between the locations of the departments. The material handling cost model has the following form:

$$F = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} f_{ij} d_{ij} \qquad i, j = 1, 2, ..., n \tag{20}$$

Note that c_{ij} is the unit cost (the cost to move one unit load one distance from department i to j), f_{ij} is the material flow between the department i and j, d_{ij} is the distance between the centers of department i and j.

7.2.2 Area utilization rate

Area utilization rate of whole layout is a ratio of total areas required of all facilities to the smallest possible rectangle, which can envelop all the facilities [24]. Hence, the area utilization area rate of whole layout is shown as follow:

$$R_s = \frac{\sum_{i=1}^{n} A_i}{\sum_{i=1}^{n} A_i + \sum B_j} \times 100\% \tag{21}$$

Note that R_s is the Area utilization rate, A_i is the area of department i where equipment i is sitting, B_j is the blank area of layout.

7.2.3 Equipment utilization

Up to now, there exist three views about equipment utilization. Østbye defines that equipment utilization is measured as of the ratio of the number of units reported in use by the surveyor relative to the total number of units recorded as present by the surveyor [25]. This measure was calculated daily, for both the intervention and control wards, during the pre- and post-intervention control periods as well as the intervention period. Optimal efficiency of utilization would have a ratio of 1, indicating that every unit present was in use. Increased utilization of existing equipment could lead to smaller equipment inventories since less equipment would be sitting unused. Michael Vineyard thinks that equipment utilization measures the percentage of time the machines are in use and considers factors beyond just maintenance downtime [26]. Steege puts forward that equipment utilization has three affecting factor: rate of quality, availability and performance efficiency [27]. Rate of quality measures the percentage of defect-free product that is manufactured by a piece of equipment. It determines the effect of the equipment on product yield. It is equally significant from a productivity point of view whether or not the equipment is running at full capacity. Equipment availability measures the percentage of time that equipment is ready to perform its manufacturing function. Performance efficiency is the percentage of available time that equipment is producing sellable product.

In this chapter, performance efficiency is selected to evaluate the equipment utilization. The equipment utilization affected by performance efficiency can be formally stated as follow:

$$R_{EU} = \frac{T_O}{T_A} \times 100\% \tag{22}$$

Note that R_{EU} is equipment utilization, T_O is the operation time of equipment, including processing time, unload time, and setup time, T_A is the available time of equipment.

7.2.4 Labor utilization

Labor utilization measures as average hours worked through overtime work (and possibly short-time work) [28]. The optimal labor utilization, i.e. hours worked per employee is explained by average wage rates which are functions of the hours worked [29]. Thus, Average labor utilization of layout can be written as:

$$R_L = \frac{T_W}{T_T} \times 100\% \tag{23}$$

Note that R_L is average labor utilization, T_W is the work time of labor, including processing, loading, unloading, loaded and empty travel time, T_T is the total time of labor in a factory.

7.3 Simulation and results

The simulation was carried out in Deneb/QUEST platform in order to investigate the performance of three types of facility layout. QUEST is a discrete event simulation software package, used to model and simulate the operation of complex automated manufacturing systems. Using 3D CAD geometry, QUEST analyzes the performance of existing or proposed manufacturing facilities by simulating the process behavior over a specified time. QUEST combines a graphical user interface with material flow logic grouped in modules for: labor, conveyors, automated guided vehicles (AGVs), kinematics, power and free conveyors, and automated storage and retrieval systems (AS/RS). A Value-Added Costing module assists in implementation of Activity Based Costing during the simulation analysis, Statistical results can be viewed with graphical and numerical analysis capabilities.

The piston production line-117 is chosen as an example to simulate. The simulation procedure is illustrated in Fig. 10. Virtual facility model can be gain from the equipment database of QUEST. The process parameters of piston are inputted to QUEST through process DB. Logic and algorithm DB provide rules and procedures that govern the behavior of the element and algorithm for the simulation system. At the final of this procedure, QUEST supplies simulation data to user in order to analysis the three types of facility layout.

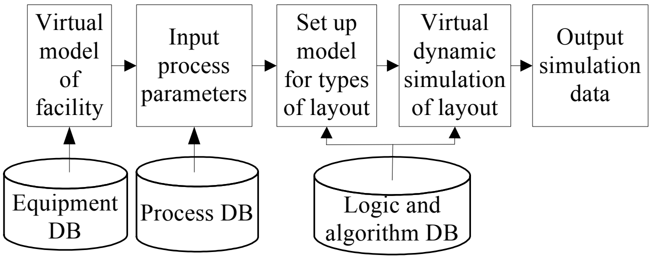


Fig. 10. Simulation procedure of facility layout

7.3.1 Assumptions and constrains

Two major assumptions made in the proposed models are as follows: (1) Machines are rectangular with the same dimensions and the distance between the machines is calculated with respect to their centers. (2) The clearance between each pair of machines is fixed.

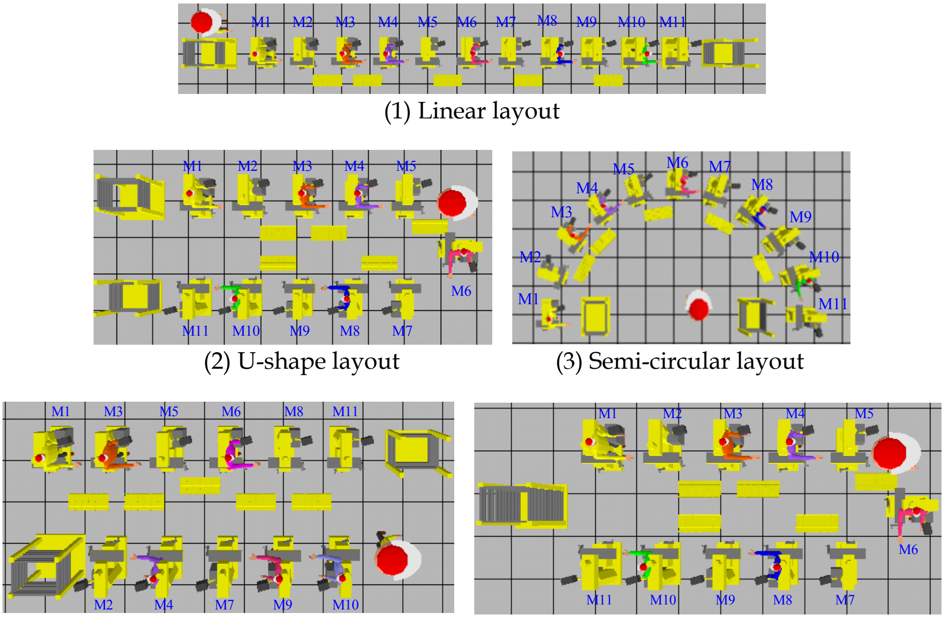
For three types of facility layout, two set of constrains are considered: (1) one machine is assigned to each location and each machine is assign to only one location; (2) the clearance between each pair of macines has the minimal value to avoid intervening and overlapping with each other.

Based on assumptions and constrains, the simulation models of the facility layout described above are given in Fig. 11.

7.3.2 Simulation results and analysis

The man-hour arrangement of piston production line is listed in Table 10. Quantity represents the number of machines. Setting the simulation time is one work day, i.e. 6.5 hours. The loaded travel time is measured by minute. Setting the moving velocity of labor is 304mm/sec and c_{ij} is 1. The simulation results are compared in Table 11~14.

The simulation results listed in Table 11 show that the material handling cost of loop layout is the lowest, and multi-line layout is the highest due to the distances between the centers of departments in which the materials are handled. As for area utilization rate shown in Table 12, loop layout is higher than the others, and semi-circular layout is the lowest of the three types of facility layout as a result of the blank areas of it is the largest. The results in Fig. 7 show that the equipment utilization of U-shape layout is the higher than others in the same moment. At aspect of labor utilization presented in Fig. 8, the U-shape layout is better than the others, and linear layout is the worst owing to the sum of its hanling materials is minimum.



(1) Linear layout

(2) U-shape layout (3) Semi-circular layout

(4) Multi-line layout (5) Loop layout

Fig. 11. Simulation model of three types of facility layout

No	Machine	Quantity	Labor	Man-hour(s）	No	Machine	Quantity	Labor	Man-hour(s）
1	M1	1	L1	43.28 (dual-workstation)	6	M6	1	L4	31.3
2	M2	1	L1	74.3	7	M7	1	L4	50.26
3	M3	1	L2	74.3	8.9	M8,9	2	L5	136.2
4	M4	1	L3	36	10	M10	1	L6	70.77
5	M5	1	L3	36	11	M11	1	L6	14.7

Table 10. Man-hour arrangement of piston production line

Type of facility layout	Linear layout	U-shape layout	Semi-circular layout	Multi-line layout	Loop layout
Material handling cost	688.607	723.2155	681.5551	880.5325	662.1082

Table 11. Material handling cost of piston production line (/yuan)

Type of facility layout	Linear layout	U-shape layout	Semi-circular layout	Multi-line layout	Loop layout
Material handling cost	30	48.26	19.096	48.26	66.47

Table 12. Area utilization rate of piston production line (%)

Type of facility layout	Linear layout	U-shape layout	Semi-circular layout	Multi-line layout	Loop layout
Equipment utilization rate	47.19764	51.43864	47.22245	46.719	45.50791

Table 13. Equipment utilization rate of piston production line (%)

Type of facility layout	Linear layout	U-shape layout	Semi-circular layout	Multi-line layout	Loop layout
Labor utilization rate	24.30267	32.38983	25.89	29.1795	27.006

Table 14. Labor utilization rate of piston production line (%)

7.4 Conclusions

In this example, single-row layout, multi-row layout and loop layout are compared with each other first. The backtracking and bypassing in linear layout are introduced. Due to customers' demand for as possible as low selling price of a product, the concept of manufacturing cost is presented secondly. Based on analyzing the components of manufacturing cost, four models are established lately. These models describe the function of material handling cost, area utilization rate, equipment utilization, and labor utilization, respectively. Through modeling and simulation on QUEST platform, three types of facility layout (linear, U-shape, semi-circular, multi-line, and loop) are compared finally. From above discussion, some conclusions can be achieved as follows: (1) material handling cost of loop layout is lowest among the three types of facility layout; (2) area utilization rate of loop layout is higher than the others; (3) equipment utilization of U-shape layout is the higher than others in the same moment; (4) labor utilization of U-shape layout is highest, and linear layout is the worst.

8. References

[1] A.R. McKendall, J. Shang. Hybrid ant systems for the dynamic facility layout problem. Computers & Operations Research, 2006, 33(3) , pp.790

[2] P. Kouvelis, A.A. Kurawarwala and G. Gutierrez. Algorithms for robust single and multiple period layout planning for manufacturing systems. Eur. J. Oper. Res., 1992, 63(2) , pp.287

[3] A.L. Page. PDMA new product development survey: Performance and best practices. Paper presented at PDMA Conference, 1991, 13

[4] B. Jaydeep, H.C. Chun and G.C. Daniel. An improved pair-wise exchange heuristic for the dynamic plant layout problem. INT. J. PROD. RES.,2000, 38(13) , pp.3067

[5] T. Yang, B.A. Peters. Flexible machine layout design for dynamic and uncertain production environments. Eur.J.Oper.Res., 1998,108(1) , pp.49

[6] T. Ertay, D. Ruan and U.R. Tuzkaya. Integrating data envelopment analysis and analytic hierarchy for the facility layout design in manufacturing systems. Information Sciences, 2006,176(3) , pp.237

[7] M.J. Rosenblatt, H.L. Lee. A robustness approach to facilities design. INT. J. PROD.RES., 1987, 25(4) , pp.479

[8] T.L. Urban. Computational performance and efficiency of lower-bound procedures for the dynamic facility layout problem. Eur.J.Oper.Res., 1992,57(2) , pp.271

[9] J.A. Tompkins, J. A. White and Y. A. Bozer. Facilities planning. NewYork: Wiley; 1996, pp.137

[10] R.D. Meller, K.Y. Gau. The facility layout problem: recent and emerging trends and perspectives. J. Manuf. Systems, 1996, 15(5),pp.351

[11] R.L. Francis, J.A. White. Facility layout and location. New Jersey: Prentice Hall, Englewood Cliffs, 1974

[12] Koopmans T. C., Beckman M. Assignment problems and the location of economic activities. Econometrica, 1957, 25(5),pp.53

[13] Bazaraa M. S.. Computerized layout design: a branch and bound approach. AIIE Transactions, 1975, 7(4) ,pp.432

[14] Lawler E. L.. The quadratic assignment problem[J]. Management Science, 1963, 9(4), pp.586

[15] Kaufman L., Broeckx F.. An algorithm for the quadratic assignment problem using Benders'decomposition. European Journal of Operational Research, 1978, 2(3) ,pp.207

[16] Foulds L. R., Robinson D. F.. Graph theoretic heuristics for the plant layout problem. Operations Research Quarterly, 1976, 27(4) ,pp.845

[17] W. Skinner. The focused factory. Harvard Business Review, May-June 1974,pp.113

[18] Heragu S.S.. Recent models and techniques for solving the layout problem. European Journal of Operational Research, 1992, 57(2) , pp.136

[19] Seppannen J.J., Moore J.M.. Facilities Planning with Graph Theory. Management Science, 1970, 17(4):242-253.49. Seppannen J.J., Moore J.M.. String Processing Algorithms for Plant Layout Problems. Int. J. Prod. Res., 1975, 13(3) , pp.239

[20] Elmohamed S., Fox G., and Coddington P.. A comparison of annealing techniques for academic course scheduling. Proceedings of the 2nd international conference on the practice and theory of automated timetabling, 1998, New York, pp.146

[21] Azadivar F., G. Tompkins. Simulation Optimization with Quantitative Variables and Structural Model Changes: A Genetic Algorithm Approach. Euro. J. of Op. Res., 1999, 113, pp.169

[22] Azadivar F., J. Wang. Facility layout Optimization using Simulation and Genetic Algorithms. International Journal of Production Research, 2000, 38(17) , pp.4369

[23] M.P. Groover. Fundamentals of Modem Manufacturing: Materials, Processes, and Systems. New York, John Wiley & Sons, Inc.,1996

[24] M.J. Wang. A solution to the unequal area facilities layout problem by genetic algorithm. Computers in Industry, Vol.56(2005),pp.207

[25] T. Østbye, D.F. Lobach, D. Cheesborough, A.M.M. Lee, K.M. Krause, V. Hasselblad, and D. Bright, "Evaluation of an infrared/radiofrequency equipment-tracking system in a tertiary care hospital" , Journal of Medical Systems, 2003,27(4),pp.367

[26] M. Vineyard, A.G. Kwasi, J.R. Meredith, "An evaluation of maintenance policies for flexible manufacturing systems: A case study", Int. J. Oper. & Prod. Mgmt., 2000,20(4),pp.409

[27] P. Steege, "Overall equipment effectiveness in resist processing equipment", Advanced Semiconductor Manufacturing Conference. IEEEISEMI, 1996

[28] H. Entorf, H. König, W. Pohlmeier, “Labor utilization and nonwage labor costs in a disequilibrium macro framework”, Scand. J. of Economics, 1992, 94(1),pp.71
[29] T. Airaksinen. On Optimal Utilization of Labor and Capital Stocks in the Neoclassical Theory of the Firm. Scand. J. of Economics, 1977, pp.28

21

Real-Time Petri Net Based Control System Design for Distributed Autonomous Robotic Manufacturing Systems

Gen'ichi Yasuda
Nagasaki Institute of Applied Science, Japan

1. Introduction

Generally speaking, flexible manufacturing systems are made up of some flexible production machines with some local storage facilities for tools and parts, some handling devices such as robots and a versatile transportation system. It is expected that more and more robots will be introduced into manufacturing systems to automate various operations in the near future. However, it is quite obvious that a single robot cannot perform effective tasks in an industrial environment, unless it is provided with some additional equipment that allows the machine to grasp, handle and dispose correctly workpieces or mechanical parts onto which technological operations are to be performed. Therefore, in order to avoid the need of loading and unloading of parts to the robot manually, it is usually required to integrate the robot into the production line that also includes machine tools, conveyors, and other special purpose machines. Mainly to provide flexibility to robots, a lot of researches have been done to develop an effective programming method for robots. But not much research has been done to integrate a system which includes various machines (robots and other devices) that cooperate in the same task (Holding & Sagoo, 1992). A common programming language for tasks that involve more than one robot or machine should be provided (Holt & Rodd, 1994).

Robot programs often must interact with people or machines, such as feeders, belt conveyors, machine tools, and other robots. These external processes are executing in parallel and asynchronously; therefore, it is not possible to predict exactly when events of interest to the robot program may occur. The programmable logic controllers (PLC) are widely used to the programming and control of flexible manufacturing systems. Implementation languages can be based on ladder diagrams or more recently state machines. However, when the local control is of greater complexity, the above kinds of languages may not be well adapted. It is important to have a formal tool powerful enough to develop validation procedures before implementation. Conventional specification languages such as ladder diagrams do not allow an analytical validation. Presently, the implementation of such control systems makes a large use of microcomputers. Real-time executives are available with complete sets of synchronization and communication primitives (Yasuda, 2000). However, coding the specifications is a hazardous work and debugging the implementation is particularly difficult when the concurrency is important.

The Petri net and its graphical representation is one of the effective means to describe control specifications for manufacturing systems. From the plant control perspective, the role and the presence of nets were considered in the scheduling, the coordination and the local control level (Silva, 1990). However, in the field of flexible manufacturing cells, the network model becomes complicated and it lacks the readability and comprehensibility. Therefore, the flexibility and expandability are not satisfactory in order to deal with the specification change of the control system. Despite the advantages offered by Petri nets, the synthesis, correction, updating, etc. of the system model and programming of the controllers are not simple tasks. The merging of Petri nets and knowledge based techniques seems to be very promising to deal with large complex discrete event dynamic systems such as flexible manufacturing systems (Gentina & Corbeel, 1987; Maletz, 1983; Wang & Sarides, 1990).

The aim of this chapter is to introduce manufacturing engineering specialists to the basic system level issues brought up by the development of computer-controlled robotic manufacturing systems and how Petri nets are applied to resolve the above mentioned problems of control system design. After some terminology concerning basic Petri nets, the extensions of Petri nets for manufacturing system control are briefly reviewed. Based on the hierarchical and distributed structure of the manufacturing system, the net model of the system is decomposed into a set of interacting local nets and a system coordinator net to perform distributed autonomous multitasking control based on Petri nets.

2. Modeling of discrete event manufacturing systems with Petri nets

The Petri net is one of the effective means to represent discrete event manufacturing systems. Considering not only the modeling of the systems but also the well-defined control, the guarantee of safeness and the capabilities to represent input and output functions are required. Therefore the Petri net has been modified and extended.

2.1 Modification of basic Petri nets

A Petri net is a directed graph whose nodes are places shown by circles and transitions shown by bars. Directed arcs connect places to transitions and transitions to places. Formally, a Petri net is a bipartite graph represented by the 4-tuple $G = \{P, T, I, O\}$ (Murata, 1989) such that:

$P = \{p_1, p_2, ..., p_n\}$ is a finite, not empty, set of places;

$T = \{t_1, t_2, ..., t_m\}$ is a finite, not empty, set of transitions;

$P \cap T = \phi$, i.e. the sets P and T are disjointed;

$I : T \to P^\infty$ is the input function, a mapping from transitions to bags of places;

$O : T \to P^\infty$ is the output function, a mapping from transitions to bags of places.

The input function I maps from a transition t_j to a collection of places $I(t_j)$, known as input places of a transition. The output function O maps from a transition t_j to a collection of places $O(t_j)$, known as output places of a transition.

Each place contains integer (positive or zero) marks or tokens. The number of tokens in each place is defined by the marked vector or marking $M = (m_1, m_2, ..., m_n)^T$. The number of tokens in one place p_i is simply indicated by $M(p_i)$. The marking is shown by dots in the places. The marking at a certain moment defines the state of the net, or the state of the system described by the net. The evolution of the state therefore corresponds to an evolution of the marking, caused by the firing of transitions. The firing of an enabled transition will change the token distribution (marking) in a net according to the transition rule. In a basic Petri net, a transition t_j is enabled if $\forall p_i \in I(t_j)$, $M_k(p_i) \geq w(p_i, t_j)$, where the current marking is M_k and $w(p_i, t_j)$ is the weight of the arc from p_i to t_j.

Because discrete event manufacturing systems are characterized by the occurrence of events and changing conditions, the Petri net type considered is the condition-event net, in which conditions can be modeled by places whilst events can be modeled by transitions. Events are actions occurring in a system. The occurrence of these events is controlled by system states. Because the condition-event system is essentially asynchronous, events always occur when their conditions are satisfied. Consequently, bumping occurs when despite the holding of a condition, the preceding event occurs. This can result in the multiple holding of that condition. From the viewpoint of discrete event process control, bumping phenomena should be excluded. So, the firing rule of the basic Petri net should be modified so that the system is free of this phenomenon. Thus the axioms of the modified Petri net are as follows:

1. A transition t_j is enabled if for each place $p_k \in I(t_j)$, $m_k = 1$ and for each place $p_l \in O(t_j)$, $m_l = 0$;
2. When an enabled transition t_j is fired, the marking M is changed to M', where for each place $p_k \in I(t_j)$, $m'_k = 0$ and for each place $p_l \in O(t_j)$, $m'_l = 1$;
3. In any initial marking, there must not exist more than one token in each place.

The number of arcs terminated at or started from a place or a transition is unlimited, but at most one arc is allowed between a transition and a place. According to these axioms, the number of tokens in each place never exceeds one, thus, the modified Petri net is said to be a safe graph. The modified Petri net is a subclass of the Petri net, and it is transformed into the equivalent Petri net as shown in Fig. 1.

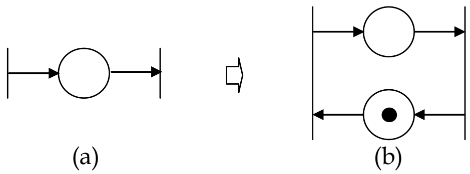


Fig. 1. (a) A place in the modified Petri net and (b) its equivalent Petri net

2.2 Extensions for real-time control

The extended Petri net adopts the following elements as input and output interfaces which connect the net to its environment: gate arc and output signal arc. A gate arc connects a transition with a signal source, and depending on the signal, it either permits or inhibits the occurrence of the event which corresponds to the connected transition. Gate arcs are classified as permissive or inhibitive, and internal or external. When the signal is 1 (true), a permissive arc permits the occurrence of the event. On the other hand, an inhibitive arc inhibits the occurrence of the event when the signal is 1. An internal arc deduces the signal from a place, and the signal is 1 when a token exists in the place, otherwise 0 (false). An external arc deduces the signal from an external machine. An output signal arc sends the signal from a place to an external machine. In addition to the axiom 1, a transition is enabled if it does not have any internal permissive arc signaling 0 nor any internal inhibitive arc signaling 1. An enabled transition is fired if it does not have any external permissive arc signaling 0 nor any external inhibitive arc signaling 1. Thus the enabling condition and the external gate condition are formally expressed as follows.

$$t_j = \bigcap_{m=1}^{M} p_{j,m}^{I} \wedge \bigcap_{n=1}^{N} \overline{p_{j,n}^{O}} \wedge \bigcap_{q=1}^{Q} g_{j,q}^{IP} \wedge \bigcap_{r=1}^{R} \overline{g_{j,r}^{II}} \tag{1}$$

$$g_j^{E} = \bigcap_{u=1}^{U} g_{j,u}^{EP} \wedge \bigcap_{v=1}^{V} \overline{g_{j,v}^{EI}} \tag{2}$$

where

M : set of input places of transition j

$p_{j,m}^{I}$: state of input place m of transition j

N : set of output places of transition j

$p_{j,n}^{O}$: state of output place n of transition j

Q : set of internal permissive gate signals of transition j

$g_{j,q}^{IP}$: internal permissive gate signal variable q of transition j

R : set of internal inhibitive gate signals of transition j

$g_{j,r}^{II}$: internal inhibitive gate signal variable r of transition j

U : set of external permissive gate signals of transition j

$g_{j,u}^{EP}$: external permissive gate signal variable u of transition j

V : set of external inhibitive gate signals of transition j

$g_{j,v}^{EI}$: external inhibitive gate signal variable v of transition j

All the variables are logical binary variables, and $\wedge$, $\vee$ denote the logical product and the logical sum, respectively, and $\bigcap_{i=1}^{m} a_i = a_1 \wedge a_2 \wedge \ldots \wedge a_m$. The state (marking) change, that is, the addition or removal of a token of a place, is described as follows:

$$p_{j,m}^{I} = p_{j,m}^{I} \wedge \overline{(t_j \wedge g_j^{E})} = RST(t_j \wedge g_j^{E}) \tag{3}$$

$$p_{j,n}^{O} = p_{j,n}^{O} \vee (t_j \wedge g_j^{E}) = SET(t_j \wedge g_j^{E}) \tag{4}$$

where $SET()$ and $RST()$ denote the set and the reset function, respectively.

Fig. 2 shows an example of extended Petri net model of robotic task control by transition firing with permissive and inhibitive gate arcs. The robot starts the loading operation based on signals from the switches, sends the commands through output signal arcs, and receive the status signals from the sensors through permissive gate arcs. Fig. 3 shows an example detailed net model of the lowest level local control of a machining center.

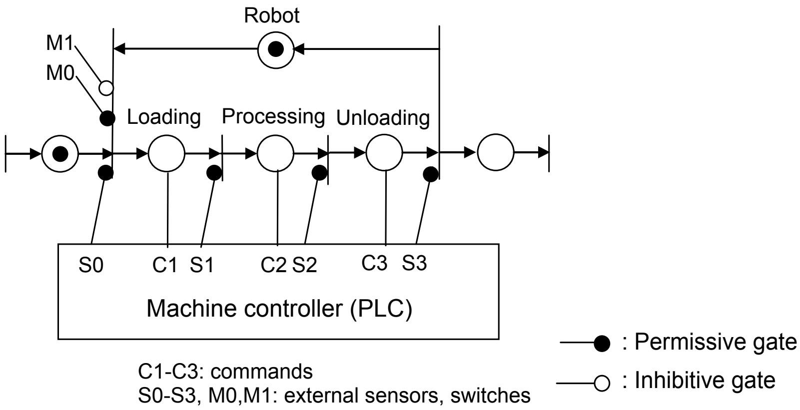


Fig. 2. Extended Petri net representation of robotic task with output signal arcs and gate arcs.

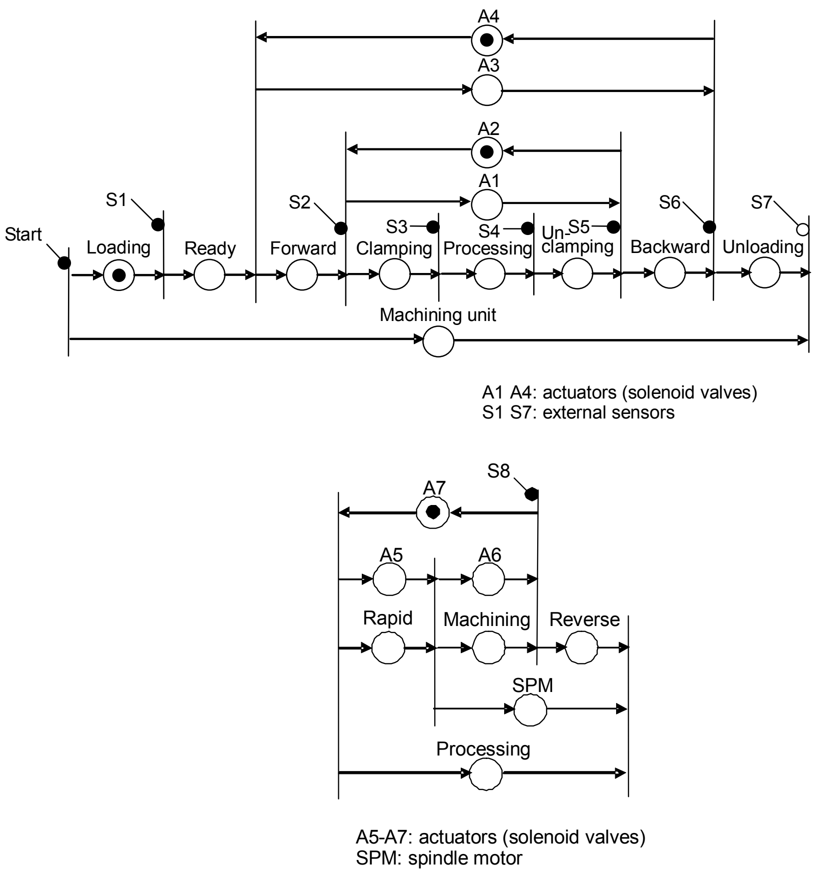


Fig. 3. Detailed net model of real-time control of manufacturing tool

3. Edition and simulation of net models

When programming a specific task, the task is broken down into subtasks. These subtasks are represented by a place. The internal states of machines are also represented by a place. The relations between these places are explicitly represented by interconnections of the transitions, arcs and gates. The whole task is edited with a net edition and simulation system. In parallel a graphic robot motion simulator system is used to edit a subtask program for a robot. The basic edition and simulation procedure is shown in Fig. 4.

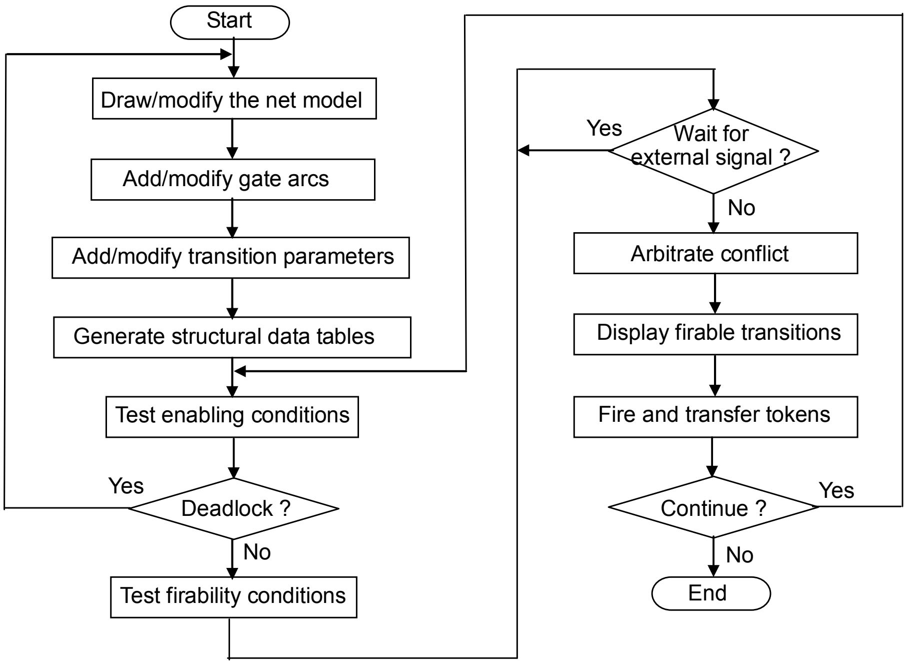


Fig. 4. Flow chart of net edition and simulation procedure

The net simulator is a tool for the study of condition-event systems and used to model condition-event systems through its graphical representation. When the net modeling is finished, the net is transformed into a tabular form and several data tables corresponding to the connection structure of the net are automatically generated (Yasuda, 2008). These tables are the following ones:

1. The table of the labels of the input and output places for each transition;
2. The table of the transitions which are likely to be arbitrated for each conflict place;
3. The table of the gate arcs which are internal or external, permissive or inhibitive, for each transition.

Although a variety of software implementations of Petri nets is possible using multitask processing (Taubner, 1988), a simple implementation method is adopted, where just one process is provided for the management of all places and tokens. Through the simulation steps, the transition vector table is efficiently used to extract enabled or fired transitions.

The table of marking indicates the current marking for each place. Using these data tables, the flow of the net simulation consists in the following steps:

1. Search enabled transitions using the axiom 1 or (1);
2. Test the enabled transitions considering gate conditions (2);
3. Arbitrate enabled transitions in conflict using some arbitration rule;
4. Execute transition firing and output corresponding signals to external machines;
5. Change the marking to the new marking using the axiom 2 or (3), (4) and update the system state.

The flow chart of the enabling condition test is shown in Fig. 5. The simulation algorithm is based on the execution rules of the net. The simulator tests each transition as to whether its input and output places and its internal gate arcs satisfy the enabling condition. If there is no enabled transition, it means that the net is in a deadlock condition. The simulator warns and requires the operator to change the initial marking or structure of the net. If there are some enabled transitions, it tests each of them as to whether its external gate arcs satisfy the firability condition, as shown in Fig. 6. If there is no firable transition, the simulator stops and shows which transitions are waiting for the gate signals.

For an example net as shown in Fig, 7, the enabling condition and the firability condition are written as (5), (6), respectively. The simulator tests each transition in the specified order of (5), (6). Fired transitions are memorized, and through their output places the output transitions of each place are searched. The enabling condition test is performed only for these transitions in order to shorten computation time. In Fig. 7, the enabling condition of only the transition t1 is evaluated, since the transition t5 is fired previously.

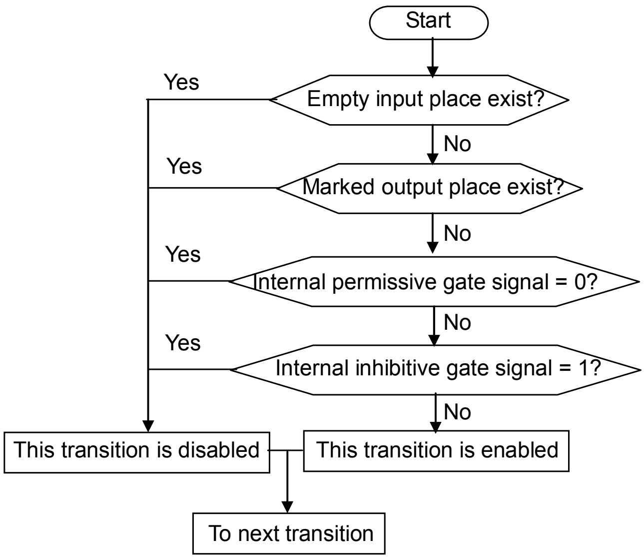


Fig. 5. Flow chart of enabling condition test

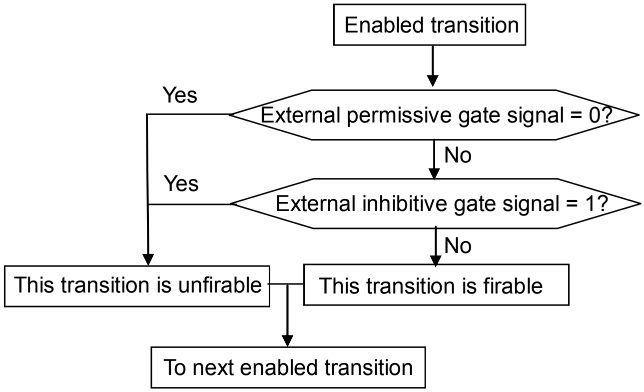


Fig. 6. Flow chart of firability condition test

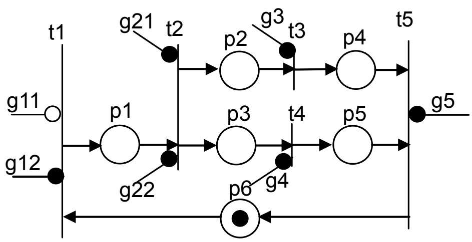


Fig. 7. Example of net representation with parallel activities.

$$
\begin{aligned}
t_1 &= p_6 \cdot \overline{p_1} \\
t_2 &= p_1 \cdot \overline{p_2} \cdot \overline{p_3} \\
t_3 &= p_2 \cdot \overline{p_4}
\end{aligned}
\qquad
\begin{aligned}
t_4 &= p_3 \cdot \overline{p_5} \\
t_5 &= p_4 \cdot p_5 \cdot \overline{p_6}
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
p_6 &= RST(t_1 \cdot \overline{g_{11}} \cdot g_{12}) \\
p_1 &= SET(t_1 \cdot \overline{g_{11}} \cdot g_{12}) \\
p_1 &= RST(t_2 \cdot g_{21} \cdot g_{22}) \\
p_2 &= SET(t_2 \cdot g_{21} \cdot g_{22}) \\
p_3 &= SET(t_2 \cdot g_{21} \cdot g_{22}) \\
p_2 &= RST(t_3 \cdot g_3)
\end{aligned}
\qquad
\begin{aligned}
p_4 &= SET(t_3 \cdot g_3) \\
p_3 &= RST(t_4 \cdot g_4) \\
p_5 &= SET(t_4 \cdot g_4) \\
p_4 &= RST(t_5 \cdot g_5) \\
p_5 &= RST(t_5 \cdot g_5) \\
p_6 &= SET(t_5 \cdot g_5)
\end{aligned}
\tag{6}
$$

If the transitions connected to a conflict place may happen to be in conflict, according to the rules of the net, only one of them is chosen to fire arbitrarily and the others become unfirable. The arbiter assigns the right of the order of firing among the transitions connected to a conflict place. But the right vanishes when the specified transition is not firable. The arbiter has a pointer to memorize the transition to be assigned the right next. The procedure of the arbitration is shown in Fig. 8. After the arbitration, all the firable transitions are displayed and fired. The simulator moves the tokens; it remove tokens in all the input places of the fired transitions and put a token in each output place of the transitions. If some error is found or the simulation result does not satisfy the specification, it can be easily amended by reediting the net and by simulating it again. The edition and simulation are performed in an interactive form on a graphic display. The software written in Visual C# under OS Windows XP allows net models be modified on-line and simulation immediately restarted.

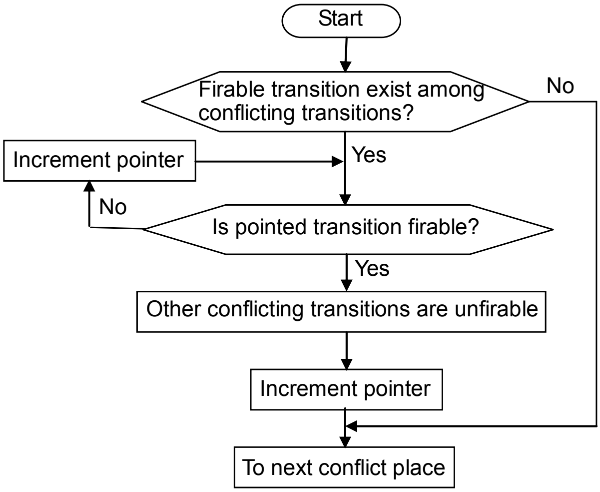


Fig. 8. Flow chart of arbitration procedure

In the basic Petri net, the firing of a transition is indivisible; the firing of a transition has duration of zero. The real-time performance of systems can be studied by adding time to the basic Petri net. An approach known as the timed Petri net associates a time parameter T with a transition, such that once the transition is enabled, it will fire after the period T. If the enabling condition is not satisfied before the schedule time comes, then the transition can not be fired and the passage of time is cancelled. Time values may be associated with places in order to maintain the instantaneous firing rule for transitions. A place with capacitance C_N, such as buffers in manufacturing systems, can be represented as a cascade connection of ordinary places with capacitance 1. The internal gate signal from the place is 1 when the number of tokens in the place is C_N, and 0 when the number is 0. These extensions are illustrated in Fig. 9(a) and (b).

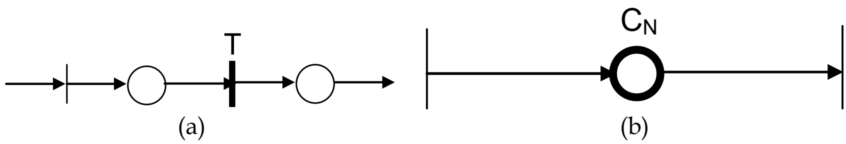


Fig. 9. Example of representation of (a) timed transition (b) place with capacitance N

4. Net models of multitasking control

Manufacturing tasks are a combination of several processes. These processes represent subtasks that are composed of task units. Tasks that include cooperative subtasks of different machines are typical examples of concurrent processes. A system with one process is the degenerate case of a system of concurrent processes, which is obtained by combining nets representing several processes. Every sequential program can be represented by a flow chart. A flow chart is composed of nodes and arcs between them. It represents the flow of control in a program and can be represented by a Petri net, by replacing the nodes with places and the arcs with transitions as shown in Fig. 10. Each arc of the flow chart is represented by exactly one transition in the corresponding net. Petri net models of sequential constructs are shown in Fig. 11. A token residing in a place means that the program counter is positioned ready to execute the next instruction. Places for motion and computational actions have a unique output transition. Decision actions introduce conflict into the net. The choice can either be made nondeterministically or may be controlled by some external signal.

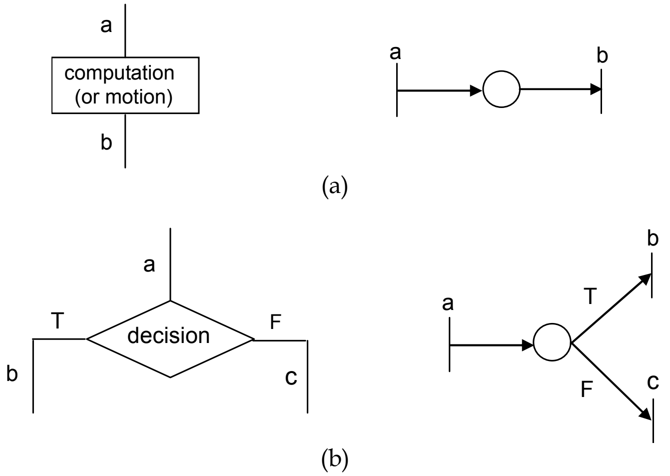


Fig. 10. Translation from nodes in a flow chart to places in a Petri net: (a) computation or motion, (b) decision

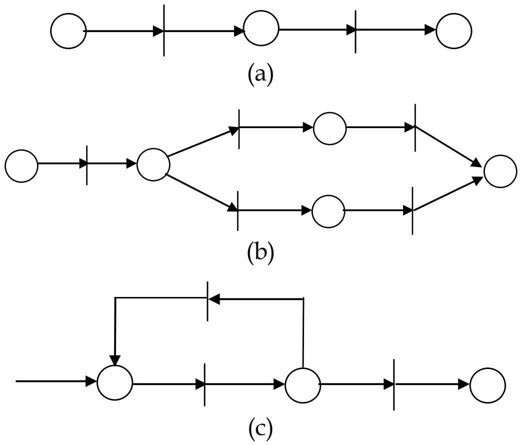


Fig. 11. Net representations of sequential constructs; (a) sequence, (b) decision, (c) iteration

In the case of two concurrent processes, where each process can be represented by a net model of a sequential process, the composite net which is simply the union of such nets can represent the concurrent execution of two processes. Parallelism is usefully introduced into a system only if the component processes can cooperate in the system. Such cooperation requires the sharing of information and resources between the processes. This sharing must be controlled to ensure correct operation of the overall system. One of the most popular synchronization mechanisms has been the P and V operations on semaphores. The WAIT and SIGNAL statements are used in a program written in a high level robot language and provides a variation of the P and V operations as a basic inter-process communication mechanism. Fig. 12 shows the net representation of an example of synchronization mechanism.

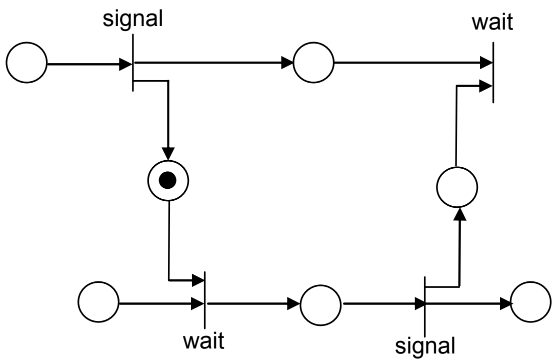


Fig. 12. Net representation of synchronization mechanism using asynchronous communication

Fig. 13 shows the net representation of cooperative operation using synchronization mechanism, where shared transitions require mutual synchronization between two robots. In contrast to decentralized implementation, synchronization can be also implemented by centralized coordination (Yasuda, 2010).

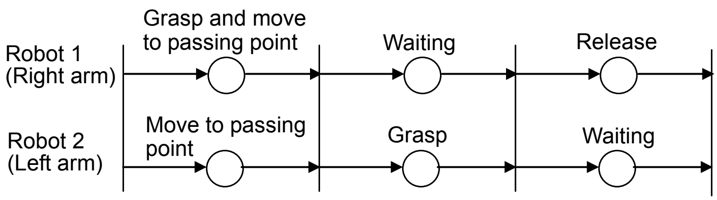


Fig. 13. Net representation of operation that passes a part from right arm to left arm

The main flow of execution control of robotic action using output signal arc and permissive gate arc is described as the following steps:

1. When a token is placed in a place which represents an action, the net based controller initiates the execution of the action (subtask) attached to the fired transition by sending the "start" signal through the output signal arc to the machine controller.

2. Then the machine controller interprets the request and runs the execution routine by sending the commands through serial interface to the robot or other external machine.
3. When the action is completed, the machine controller informs the system controller to proceed with the next activations by sending the "end" signal through external permissive gate arc.

When a token is placed in a place which is "ready" state in the net model, the controller sends the "ready" signal. If the machine receives the signal, it runs the processing routine which performs the initializations and other preliminary processing for the next execution routines. When the processing routine is completed, it sends the "ack" (acknowledgement) signal to the system controller. The "end" and "ack" signals work as gate signals for the system controller.

5. Implementation of real-time control system for robotic cells

To implement the Petri net based modeling and control method, the net based task editor and simulator, and the real-time controller based on tasks represented as net models were developed (Yasuda, 2008). The subtasks and sets of point data needed to execute the whole task are initially identified. Then they are edited and tested with the net based edition and simulation system. Initially, the proposed method is used to execute a simple example of pick-and-place task by a single robot. The experimental set up includes the following equipment: a small industrial robot with an arm (Mitsubishi Electric, Movemaster II RM501), two belt conveyors with their sequence control circuits, a NC machine tool and a general PC. All the software is written in Microsoft Visual C# on Windows XP. The task specification is represented as the flow of a workpiece and written as the following steps:

1. A workpiece arrives at point E1.
2. Conveyor CV1 carries the workpiece to point E2.
3. Robot R1 transfers the workpiece to point E3.
4. Machining operation M1 is done.
5. Robot R1 transfers the workpiece to point E4.
6. Conveyor CV2 carries the workpiece to point E5.

Synchronous cooperation is required to perform the loading and unloading operations between the robot and the conveyor or machining center. The cooperation can be implemented by a system coordinator which coordinates the machine controllers such that associated transitions of the local net models fire simultaneously. For high efficiency, it is desirable that the system accepts as many workpieces as possible, but it must not be in a deadlock condition. Generally, if there are some paths between two transitions, the largest number of tokens in each path is the smallest number of places of the paths. The task specification is shown as follows. Using the place of capacity control, the net representation of the task program written under these requirements is shown in Fig. 14.

Another example is a cooperative task by two arm robots which must synchronize their actions with each other. The task specification is summarized as the following steps:

1. A workpiece arrives at point E1.
2. Robot R1 transfers the workpiece to the exchange area, and at the same time Robot R2 moves to the exchange area.

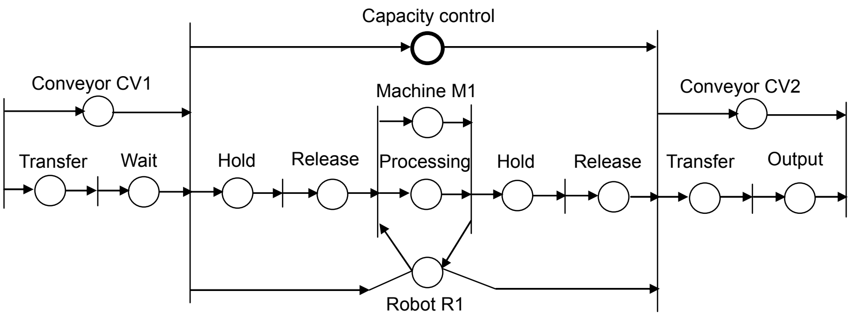


Fig. 14. Net representation of pick-and-place operation with a single robot

3. The workpiece is exchanged from robot R1 to robot R2.
4. Robot R2 changes the workpiece orientation.
5. Robot R2 transfers the workpiece to the exchange area. Robot R1 moves to the exchange area.
6. The workpiece is exchanged from robot R2 to robot R1.
7. Robot R1 transfers the workpiece to point E2.

Following the same procedure of the former example, the subtasks and sets of point data needed to execute the whole task are initially identified. Then they are edited and tested with the net based edition and simulation system. The net representation is written using shared transitions for system coordination as shown in Fig. 15. An experimental view of the cooperative task, passing and exchanging a workpiece, by two robots is shown in Fig. 16.

The detailed procedure of the implemented real-time control based on tasks represented as net models is described as follows. If there is a token in a place corresponding to subtasks, the net based controller sends a message to the respective hardware controllers such as arm, hand, sensor, etc. to execute the defined subtask with certain point data. These parameters (hardware controller code, subtask file code, point data file code) are defined during the net edition procedure. The net based controller was developed with all functions of the edition and simulation to permit correction or modification of the net model on-line. This characteristic is important to facilitate the debugging work. By executing the net model, the

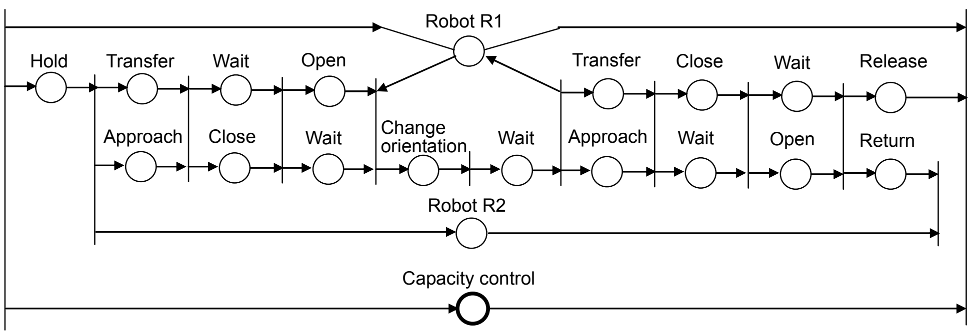


Fig. 15. Detailed net model of cooperative task by two arm robots

Fig. 16. Experiments of cooperative task by two arm robots

developed control system activates the arm, hand, and sensor, etc. and coordinates each individual controller. In making these experiments, it was verified that the implemented system can be used as an effective tool for introducing robots into the manufacturing system. The system can be used to verify and correct control algorithms including robot movements and to evaluate the effectiveness of a robot and other machines in the planning stage.

A multi-computer control architecture composed a system computer and several control computers has been adopted as shown in Fig. 17. The computer control architecture was

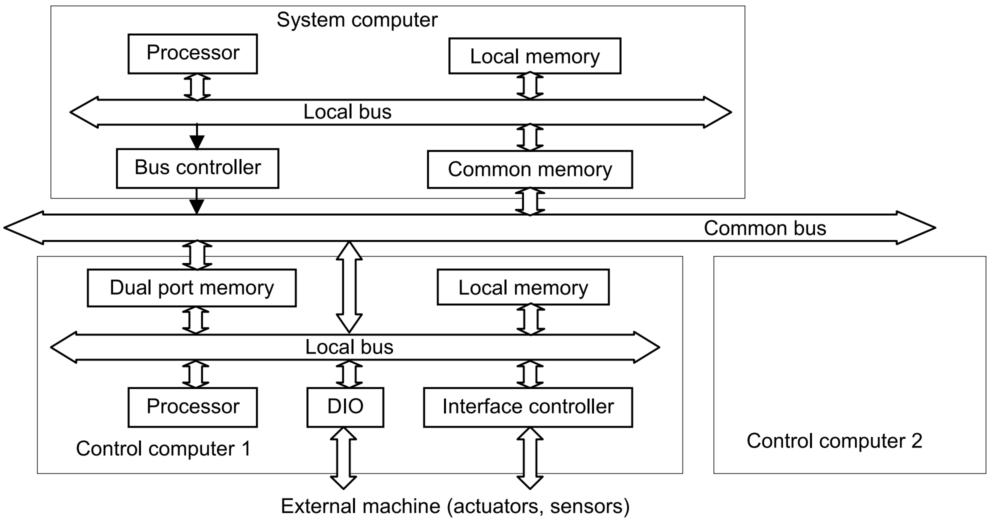


Fig. 17. Multi-computer control architecture composed a system computer and several control computers with dual port memory

developed for the use of distributed autonomous control of independent actuators or machines in compact factory automation systems (Yasuda & Tachibana, 1987). The system controller controls communication between the system controller and control computers through the bus controller based on the master-slave mode. The system computer installs the conceptual net model for system coordination and installs local net models in the control controllers through the common bus. The control computers are equipped with interface circuits to actuators and external sensors for direct machine control and monitoring. Then, in the real-time control, the system computer communicates with each control computer through dual port memory with respect to firing of shared transitions and gate arc signals (Yasuda, 2011). The presented control flow of the net model is successfully executed using output signal arc and permissive gate arc. The net model in the system controller is conceptual for system coordination and not so large. The computation speed of 50 MHz of the general microprocessor is satisfactorily high in comparison with those of controlled devices such as robotic arms, conveyors, machine tools and external sensors.

6. Conclusions

A Petri net based specification and real-time control method for large complex robotic manufacturing systems was introduced as an effective prototyping tool to realize distributed autonomous control systems corresponding to the hardware structure of robotic manufacturing systems. From the design point of view, the use of nets has many advantages in modeling, qualitative analysis, performance evaluation and code generation. The Petri net appears as a key formalism to describe, analyze and implement the distributed autonomous control system for manufacturing systems in future.

7. References

Gentina, J. C. & Corbeel, D. (1987). Coloured Adaptive Structured Petri-Net: A Tool for the Automatic Synthesis of Hierarchical Control of Flexible Manufacturing Systems, *Proceedings of 1987 IEEE International Conference on Robotics and Automation,* pp. 1166-1172

Holt, J. D. & Rodd, M. G. (1994). An Architecture for Real-Time Distributed AI-Based Control Systems. In: *IFAC Distributed Computer Control Systems 1994*, 47-52

Holding, D. J., & Sagoo, J. S. (1992). A Formal Approach to the Software Control of High-Speed Machinery. In: *Transputers in Real-Time Control,* G. W. Irwin & P. J. Fleming (Eds.), 239-282, Research Studies Press, Taunton, Somerset, U.K.

Maletz, M. C. (1983). An Introduction to Multi-robot Control Using Production Systems, *Proceedings of IEEE Workshop on Languages for Automation*, pp. 22-27

Murata, T. (1989). Petri Nets: Properties, Analysis and Applications. *Proceedings of the IEEE*, Vol. 77, No. 4, pp. 541-580

Silva, M. (1990). Petri Nets and Flexible Manufacturing. In: *Advances in Petri Nets* G. Rozenberg (Ed.), LNCS 424, 374-417, Springer-Verlag, Berlin, Germany

Taubner, D. (1988). On the Implementation of Petri Nets. In: *Advances in Petri Nets* G. Rozenberg (Ed.), LNCS 340, 419-439, Springer-Verlag, Berlin, Germany

Wang, F. & Saridis, G. N. (1990). A Coordination Theory for Intelligent Machines, *Proceedings of the 11th IFAC World Congress*, pp. 235-240

Yasuda, G. (2000). A Multiagent Control Architecture for Multiple, Cooperating Robot Systems, *Proceedings of the International Conference on Production Research Special ICPR-2000,* Paper ID 224, August, 2000

Yasuda, G. (2008). Implementation of Distributed Control Architecture for Industrial Robot Systems Using Petri Nets, *Proceedings of the 39th International Symposium of Robotics*, pp.533-538

Yasuda, G. (2010). Petri Net Based Implementation of Hierarchical and Distributed Control for Discrete Event Robotic Manufacturing Systems, *Proceedings of the 2010 IEEE International Conference on Control Applications, Part of 2010 IEEE Multi-Conference on Systems and Control*, pp.251-256

Yasuda, G. (2011). Design and Implementation of Distributed Control Architecture for Flexible Manufacturing Cells Based on Petri nets, *Proceedings of 12th Asia-Pacific Industrial Engineering & Management Systems Conference*, pp. 852-864

Yasuda, G. & Tachibana, K. (1987). A Multimicroprocessor-Based Distributed Processing System for Advanced Robot Control, *Proceedings of the IXth International Conference on Production Research*, pp.1926-1933