

Materials Forming, Machining and Tribology

J. Paulo Davim *Editor*

# Modern Mechanical Engineering

Research, Development and Education

 Springer

# **Materials Forming, Machining and Tribology**

*Series editor*

J. Paulo Davim, Aveiro, Portugal

For further volumes:

<http://www.springer.com/series/11181>

J. Paulo Davim  
Editor

# Modern Mechanical Engineering

Research, Development and Education

 Springer

*Editor*  
J. Paulo Davim  
Department of Mechanical Engineering  
University of Aveiro  
Aveiro  
Portugal

ISSN 2195-0911                      ISSN 2195-092X (electronic)  
ISBN 978-3-642-45175-1            ISBN 978-3-642-45176-8 (eBook)  
DOI 10.1007/978-3-642-45176-8  
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013955555

© Springer-Verlag Berlin Heidelberg 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

Nowadays, Mechanical Engineering is currently defined as a discipline “*which involves the application of principles of physics, design, manufacturing and maintenance of mechanical systems.*” Among others, fundamental subjects of classical mechanical engineering include: kinematics and dynamics, materials and manufacturing processes, thermodynamics and heat transfer, fluid mechanics, automation and control, applied mechanics and design, etc. With a special emphasis, this book also covers some modern subjects of mechanical engineering such as nanomechanics and nanotechnology, mechatronics and robotics, computational mechanics, alternative energies, sustainability, as well as aspects related with mechanical engineering education. It look to cover original contributions that advance understanding of both the fundamentals of mechanical engineering and its application to the solution of problems in modern industry.

The main aim of this book is to present a collection of examples illustrating research, development, and education in modern mechanical engineering. [Chapter 1](#) of the book provides sustainability in mechanical engineering discipline. [Chapter 2](#) is dedicated to application of solar distillation systems with phase change material storage. [Chapter 3](#) describes magneto-rheological fluid technology. [Chapter 4](#) contains information on tribological behaviour of rare-earth lubricating oils. [Chapter 5](#) is dedicated to structural dynamics and viscoelastic passive damping treatments. [Chapter 6](#) describes thermo mechanical modeling of multiphase steels (classical and modern engineering analysis). [Chapter 7](#) provides optimal real-time management automated production lines. [Chapter 8](#) deals with modeling optimization of mechanical systems and processes. [Chapter 9](#) contains information on implementing STEP-NC (exploring possibilities for the future advanced manufacturing) and [Chap. 10](#) is dedicated of optimum CNC free-form surface machining through design of experiments in CAM software. [Chapter 11](#) describes modeling of micromachining and [Chap. 12](#) is dedicated of micromilling. [Chapter 13](#) provides digital image processing in machining. [Chapter 14](#) describes formability and simulative tests in modern sheet metal forming education. Finally, [Chap. 15](#) is dedicated to multimedia resources in engineering education.

With a great usefulness, this book can be used as a research book for final undergraduate engineering course or as a topic on mechanical engineering at the postgraduate level. It also can serve book can serve as a useful reference for academics, researchers, mechanical, manufacturing, industrial and materials

engineers, professionals in mechanical engineering and related industries. The interest of scientific in this book is evident for many important research centers, laboratories, and universities as well as industry. Therefore, it is hoped this book will inspire and fill with enthusiasm others to undertake research in modern mechanical engineering.

The Editor acknowledges Springer for this opportunity and for its enthusiastic and professional support. Finally, I would like to thank to all the chapter authors for their availability for this work.

Aveiro, Portugal, October 2013

J. Paulo Davim

# Contents

<b>1</b>	<b>Sustainability in Mechanical Engineering Discipline</b> . . . . .	<b>1</b>
	S. Vinodh, K. Jayakrishna and K. E. K. Vimal	
<b>2</b>	<b>Application of Solar Distillation Systems with Phase Change Material Storage</b> . . . . .	<b>15</b>
	S. K. Shukla	
<b>3</b>	<b>Magneto-Rheological Fluid Technology</b> . . . . .	<b>43</b>
	Ali A. Alghamdi, Ruben Lostado and Abdul-Ghani Olabi	
<b>4</b>	<b>Tribological Behaviour of Rare-Earth Lubricating Oils</b> . . . . .	<b>63</b>
	Tianxia Liu, Xianguo Hu, Enzhu Hu and Yufu Xu	
<b>5</b>	<b>Structural Dynamics and Viscoelastic Passive Damping Treatments</b> . . . . .	<b>89</b>
	R. A. S. Moreira	
<b>6</b>	<b>Thermomechanical Modeling of Multiphase Steels: Classical and Modern Engineering Analyses</b> . . . . .	<b>109</b>
	A. Andrade-Campos, P. Vasconcelos, J. F. Caseiro and J. A. Oliveira	
<b>7</b>	<b>Optimal Real-Time Management for Automated Production Lines</b> . . . . .	<b>147</b>
	Joao Figueiredo	
<b>8</b>	<b>Modeling and Optimization of Mechanical Systems and Processes</b> . . . . .	<b>169</b>
	Ramón Quiza, Gerardo Beruvides and J. Paulo Davim	
<b>9</b>	<b>Implementing STEP-NC: Exploring Possibilities for the Future of Advanced Manufacturing</b> . . . . .	<b>199</b>
	Kelvin Hamilton, Jean-Yves Hascoet and Matthieu Rauch	

<b>10 Optimum CNC Free-form Surface Machining Through Design of Experiments in CAM Software . . . . .</b>	<b>241</b>
N. A. Fountas, N. M. Vaxevanidis, C. I. Stergiou and R. Benhadj-Djilali	
<b>11 Modeling of Micromachining . . . . .</b>	<b>285</b>
Angelos P. Markopoulos and Dimitrios E. Manolakos	
<b>12 Micro Milling . . . . .</b>	<b>325</b>
Emel Kuram and Babur Ozcelik	
<b>13 Digital Image Processing in Machining . . . . .</b>	<b>367</b>
Samik Dutta, Surjya K. Pal and Ranjan Sen	
<b>14 Formability and Simulative Tests in Modern Sheet Metal Forming Education . . . . .</b>	<b>411</b>
P. A. F. Martins, L. Montanari, V. A. Cristino and M. B. Silva	
<b>15 Multimedia Resources in Engineering Education . . . . .</b>	<b>449</b>
B. F. Yousif, Marita Basson and Carola Hobohm	
<b>Index . . . . .</b>	<b>463</b>



# Chapter 1

## Sustainability in Mechanical Engineering Discipline

S. Vinodh, K. Jayakrishna and K. E. K. Vimal

**Abstract** Sustainable development is very much essential in the contemporary scenario due to increasing environmental concerns, demand for energy, global warming and end of life safe disposal of products. Engineering Professionals possess a vital role to enable sustainable development. This situation necessitated the engineers to understand the challenges and opportunities associated with sustainable development. This chapter highlights the need for including sustainability concepts in modern mechanical engineering curriculum. This chapter provides the details about sustainability, three pillars and orientations of sustainability, fundamental concepts and tools/techniques of sustainability. The insights on challenges and recent advancements in sustainability are also presented with the curriculum contents of sustainability in mechanical engineering education.

### 1.1 Introduction

In order to suit the needs of modern industries, the engineering education curriculum need to be renewed. Sustainable development is the contemporary need and mechanical engineering professionals are expected to understand the challenges and provide sustainable solutions. This section presents the fundamentals of sustainability which includes the definition, pillars and orientations of sustainability along with the emphasis for sustainability in Mechanical Engineering discipline.

---

S. Vinodh (✉) · K. Jayakrishna · K. E. K. Vimal  
Department of Production Engineering, National Institute of Technology,  
Tiruchirappalli, Tamil Nadu, India  
e-mail: vinodh\_sekar82@yahoo.com

### ***1.1.1 Definition of Sustainability***

Sustainable development is the development that meets the needs of the present without compromising the abilities of future generations. Sustainable development considers economic crisis, climatic changes, pressure from all levels of stake holders and scarcity of materials [1]. The present world is facing problems like rapid depletion of natural resources and undesired environmental changes on a global scale. The manufacturing organizations are realizing the importance of mitigating the present crisis and are adopting sustainable manufacturing principles. Sustainability in general deals with reduction of consumption of resources including material, energy and water resources and reducing emission of hazardous and non-hazardous wastes into environment.

### ***1.1.2 Three Pillars of Sustainability***

Sustainability addresses Triple Bottom Line (TBL). Three pillars of sustainability include environment, economy and society. They are alternatively referred to as profitability, people and planet. These are the general core pillars of sustainability. The improvement efforts need to be concentrated on these three pillars to cumulatively improve sustainability. Recent studies on sustainability focus on economic and social issues on par with environmental aspects.

### ***1.1.3 Three Orientations of Sustainability***

Three orientations of sustainability include material, product design and manufacturing processes. It uses both technological and non-technological solutions, in selection of materials, product design and production processes [2]. Material selection for a product affects the amount of natural resources consumed and recovery of the material at the end. Decision made during product design is a major factor because it directly affects the resources consumption and waste generation pattern throughout the product life cycle. Developing technologies to transform materials with reduced emissions of greenhouse gases, reduced use of non-renewable or toxic materials and reduced generation of waste. Sustainability aspects can be incorporated at product and process levels. At the process level, there exists a need to achieve optimized technological improvements and process planning for reduced energy and resource consumptions, toxic wastes, occupational hazards, etc. At the product level, sustainability can be achieved by reducing the energy of the product and recovering the waste generated.

### ***1.1.4 Need for Sustainability in Mechanical Engineering***

Sustainability in mechanical engineering is necessary to design the processes and systems in such a way that they use energy and resources sustainably. Sustainability is very much important in mechanical engineering curriculum because it provides a platform for sharing the latest thinking from research and practice viewpoint. Mechanical Engineers have a critical role to enable countries to achieve sustainable development. They must understand the challenges and opportunities that arise from the needs to achieve sustainable development. It is necessary to train students on sustainability for enabling them to be equipped with appropriate knowledge and skills to meet client needs.

## **1.2 Fundamental Concepts of Sustainability**

The term sustainable manufacturing was often referred to as eco-efficiency, remanufacturing, green technology, cleaner production etc. To achieve sustainability in manufacturing, materials used, product design, processes used for product manufacturing need to be sustainable. In other means throughout the supply chain starting from product design to End of life (EoL) activities needs to be environmentally friendlier. In this section, various concept models and frameworks to ensure sustainability, Sustainable materials, Sustainable product, Sustainable processes, Sustainable supply chain, importance of sustainability assessment will be discussed.

### ***1.2.1 Concept Models of Sustainability***

There have been numerous ways of representing sustainable development in a model that captures this extremely complex concept and a new way of thinking. In general sustainability models can be classified into five types namely Pictorial Visualization Models, Quantitative Models, Physical Models, Conceptual Models and Standardizing Models [3]. Among these conceptual models, gains much importance as it can be directly used for practical applications like concept selection, assessment and improvement proposal identification. Labuschagne et al. [4] proposed a comprehensive framework of sustainability criteria that can be used to assess the sustainability of projects, technologies, as well as the overall organizational sustainability. In this process, they reviewed existing frameworks namely Global Reporting Initiative, United Nations Commission on Sustainable Development Framework, Sustainability Metrics of the Institution of Chemical Engineers and Wuppertal Sustainability Indicators. Based on the review they proposed a modified framework for sustainability assessment which has criteria for

economic sustainability, environmental sustainability and social sustainability. They validated the appropriateness of the proposed model by measuring the sustainability in South African process industry. Vinodh [5] proposed a conceptual model for sustainability assessment. This model addresses three perspectives namely environment, economy and society. This model has 12 indicators and 37 attributes. This model was used for assessing the sustainability level of the organization using Multi grade fuzzy approach.

### ***1.2.2 Sustainable Materials***

In the past, energy use and content is one of the few quantifiable aspects of environmental performance and could therefore be used in materials selection exercises. However, with increased environmental awareness in all sectors of industry, environmental data on the effect of material production and processing is becoming more readily available. Although quantifying the amount of a single pollutant, for example CO<sub>2</sub> emitted through production of a material is a complex task. It is relatively simple when compared to the problems which can arise when trying to assess the overall environmental effect. The environmental performance of materials needs to be evaluated during their entire life cycle, encompassing extraction and processing resources, distribution, use, recycling, and final disposal [6]. Environmental impact of products is directly influenced by many factors but environmental properties of the materials make significant contribution such as energy costs, emissions involved in production and manufacturing phases, and recyclability. The choice of materials therefore assumes strategic importance, and requires an extension of the characterization of materials, integrating conventional characterization aimed at defining physical cum mechanical properties, with a complete characterization of environmental performance. For the designer to make an optimal choice of materials, harmonizing performance characteristics and properties of eco-compatibility, the selection process must take into account a wide range of factors: constraints of shape and dimension, required performance, technological and economic constraints associated with manufacturability of materials, environmental impact of all phases of the life-cycle.

The optimal choice of materials, also in relation to environmental demands, requires a complete environmental characterization, with particular regard to the following aspects [7]:

- Environmental impacts associated with production processes (energy costs and overall impact);
- Environmental impacts associated with end of life phases (recycling or disposal);
- Suitability for recycling (expressed by the recyclable fraction).

### ***1.2.3 Sustainable Products***

A sustainable product is a product, which will give as little impact on the environment as possible during its life cycle. The life cycle in this simple definition includes extraction of raw materials, production, use and final recycling (or deposition). The material in the product as well as the material (or element) used for producing energy is also included here. This definition is in fact not totally defined according to the amount of impact on the environment or the nature. Hence the impact cannot be zero, it must be reasonably minimized. This opens up the fact that there will be similar products, which are more or less sustainable when they are compared concurrently. The environmental impact can be reduced by minimizing negative environmental impact during various stages of product development. The environmental impact during usage and recycling can be measured in terms of Carbon foot print, Water eutrophication, Air acidification and Total energy consumption [8].

- Carbon foot print is defined as the total sets of greenhouse gas emissions caused by an organization or product.
- Water Eutrophication is the ecosystem response to the addition of artificial or natural substances, such as nitrates and phosphates, through fertilizers or sewage, to an aquatic system.
- Air acidification includes acid gases that are released into air or resulting from the reaction of non-acid components of the emissions taken up by atmospheric precipitations.
- Total energy consumption is the energy consumed in the process of converting raw materials into finished products.

The above mentioned parameters need to be controlled to ensure product sustainability.

### ***1.2.4 Sustainable Processes***

The prime orientations of sustainability are product, process and material. In this, process sustainability gains more importance as it directly contributes the energy consumption and negative environmental impact. The main factors contributing to environmental impacts include material and energy consumption during machining operations. An integrated process covers various parameters such as Process parameters, Machine type, Machine components (e.g. drives) and surrounding machine equipment (e.g. filter concept, exhaust system). Manufacturing process constitute major activity that contributes to the growth of the global economy. Intensive research work in manufacturing process ensures the improvement in manufacturing performance. But due to global warming issues, the focus has been shifted towards the development of environmentally benign process. The

environmental and health benign process and advanced technologies were developed for achieving cleaner, healthier, safer and also economic process. During the manufacturing process electrical energy, compressed air, cutting fluids, and raw material have to be considered as main inputs. Steps need to be taken to reduce or eliminate the consumptions of input which are contributing to the environmental negative impact.

### ***1.2.5 Sustainable Supply Chains***

A focus on supply chains is a step towards broader adoption and development of sustainability, since the supply chain considers the product from initial processing of raw materials to delivery to the customer. However, sustainability also must integrate issues and flows that extend beyond the core of supply chain management: product design, manufacturing by-products, by-products produced during product use, product life extension, product end-of-life, and recovery processes at end-of-life.

### ***1.2.6 Sustainability Assessment***

The concept of sustainability, or sustainable development, is clearly the basis of sustainability assessment. Sustainable development was first described by Brundtland Commission in 1987. Most of the sustainable assessment models were developed on the basis of Triple Bottom Line (TBL). The theory of sustainability assessment as currently expressed in the literature has largely evolved from work undertaken by practitioners of environmental impact assessment (EIA), and more recently strategic environmental assessment (SEA), which in turn has been influenced by policy analysis techniques [9]. The literature reflects a widely held belief that environmental assessment processes such as EIA and SEA can, and do, make valuable contributions towards sustainability.

It is important to note at this point that sustainability is a difficult concept to define in a way that is meaningful and sufficiently practical to allow it to be operationalized. It has been suggested that the difficulty arises because sustainability is a concept which remain ‘fuzzy’ until applied in a specific context. This situation is not aided by the fact that many alternative theoretical formulations and applications of sustainability have been developed [10].

At first, proper performance indicators are selected covering different aspects of sustainability. Azapagic [11] emphasized that indicators should be quantitative whenever possible; however, for some aspects of sustainability, qualitative descriptions may be more appropriate (e.g. societal aspects). Each indicator selected should be tracked periodically and equipped with symbol and unit of measurement. The selected indicators are grouped according to main aspects of

sustainability (economy, environment, and society). Economic group of indicators concerns the impacts of the organization on the well-being of its stakeholders and on economic systems at the local, national and global levels. Environmental group of indicators cover impacts of the company on living and nonliving natural systems, including ecosystems, land, air and water. Societal group of indicators reflects the attitude of the company to the treatment of its own employees, suppliers, contractors and customers, and also its impact on society at large [12]. As sustainability can be classified into three orientations namely material, process and product, sustainability assessment needs to be done with respect to these three orientations. Few researchers have done sustainability in case organizations with respect to above mentioned three orientations. As the nature of assessment remains fuzzy in nature, fuzzy based techniques namely Multi Grade Fuzzy, weighted fuzzy approach, Fuzzy inference system and Fuzzy logic based Kano model were already tested for sustainability assessment [5, 13–15].

### **1.3 Tools/Techniques of Sustainability**

Sustainability is a strategy in the design, development and manufacture of new products that is extensively acknowledged in principle, while not yet widely adopted [16]. This subsection discusses tools/techniques majorly available in design, development and manufacture of new products, based on a paradigm for sustainable manufacturing.

#### ***1.3.1 Design for Environment***

The conventional product development approach does not include any environmental aspects in the development process. Environmental requirements are be introduced to any one of the development stages by applying various tools and methodologies so as to bridge the gap between product development and environmental management. The design stage of product development is often renowned as a significant stage for determining costs and profitability, also the potential for infusing environmental characteristics into the product is more. During the early design stage, the vital factors like materials, process, and energy source needs to be clearly defined as the environmental impacts of the product for its entire life cycle is based on these factors. Assimilation of environmental factors into products design is known as design-for-the-environment (DfE), eco-design, life-cycle design, sustainable product design (SPD) or design for sustainability [17]. Design for environment is also defined as a system approach to design that incorporates environmental, health and safety concerns over the complete product life cycle. Design for environment principles and guidelines helps design engineers to create greener products during the early stages of design when life cycle

analysis is not completely feasible [18]. The eco designer (one designs by considering the environmental impact) must also be cognizant of changes throughout the manufacturing sector, especially with respect to the environment, and at the international level.

Following are the future needs with which the eco designer should design a product [19],

- Products should be designed for reuse.
- Better reprocessing technologies must be developed.
- Integration of financial and environmental systems.
- Reuse/life prediction modeling.
- Accounting system for the value in processing/design selection.

Ehrenfeld et al. [20] claims that DfE provides competitive advantages through

- Reduction of manufacturing costs;
- Reduction of wastes;
- Satisfaction of consumers' environmental demands;
- Lightening of regulatory burden;
- New sources of revenues and profits;
- Creating a culture that encourages change.

### ***1.3.2 Environmentally Conscious Quality Function Deployment***

Sustainable system tries to maximize resources efficiency for the production of sustained components [21]. Also in the contemporary manufacturing scenario, the environmental requirements must be treated with equal importance to traditional product requirements [22]. The reduction of environmental problems for ensuring clean and green atmosphere is the focus of contemporary manufacturing organizations [23]. Environmentally Conscious Quality Function Deployment (ECQFD) integrates environmental requirements into QFD methodology and extending it so as to evaluate improvement concepts [24]. ECQFD consists of four phases. In Phase I, Voice of Customers (VoC), Voices of the Environment (VoE) and quality characteristics (QC) for traditional and environmental qualities are correlated, while QC and components are also correlated in Phase II. Components can be regarded as function units or part characteristics. The outputs of Phase I and II are the identification of the QC and components that should be considered. Following this, the so-called ECQFD team will examine design improvement options for the product in a semi-quantitative manner in Phases III and IV. To do so, the team determines a redesign target and those changes are expressed by a combination of QC and component to be improved. Then, they evaluate the effects of the design changes on the VoC and VoE using semi-quantitative information represented in the two correlation matrices in Phases I and II [1]. The main advantage of ECQFD



is that several technical attributes and environmental concerns can be prioritized so that product design team can concentrate on their limited resources on critical issues to develop customer oriented environmentally friendly products.

### ***1.3.3 Life Cycle Assessment***

There are two LCA standards created by the International Organization for Standardization (ISO): ISO 14040 and ISO 14044. Life Cycle Assessment, as defined by ISO 14040 and ISO 14044, is the compiling and evaluation of the inputs and outputs and the potential environmental impacts of a product system during a product's lifetime.

LCAs are used by a variety of users for a range of purposes. According to ISO standards LCA, can assist in:

- Identifying opportunities to improve the environmental aspects of products at various points in their life cycle;
- Decision making in industry, governmental or non-governmental organizations (e.g. strategic planning, priority setting, product and process design or redesign);
- Selection of relevant indicators of environmental performance, including measurement techniques; and
- Marketing (e.g., an environmental claim, eco-labeling scheme or environmental product declarations).

Life Cycle Assessment is conducted according to ISO 14040 and 14044 standards. LCA consist of four steps, namely:

1. Goal and Scope Definition
2. Inventory Analysis
3. Impact Assessment
4. Interpretation

These four steps are described in detail in the following sections.

#### **1.3.3.1 Goal and Scope Definition**

According to ISO 14040 standards, the first phase of an LCA is the definition of the goal and scope. In this step, all general decisions for setting up LCA system are made. The goal and scope should be defined clearly and consistent with the intended application. An LCA is an iterative process and this allows redefining the goal and scope later in the study based on the interpretation of the results.

During the scope definition, the product or process system under study is characterized, all assumptions are detailed and the methodology used to set up the product system is defined.

### **1.3.3.2 Inventory Analysis**

Inventory Analysis is the LCA phase that involves the compilation and quantification of inputs and outputs for a given product system throughout its life cycle or for single processes. The inventory analysis includes data collection and compilation of data in an Life Cycle Inventory (LCI) table.

### **1.3.3.3 Life Cycle Impact Assessment**

Life Cycle Impact Assessment (LCIA) identifies and evaluates the amount and significance of the potential environmental impacts arising from the LCI. The inputs and outputs are first assigned to impact categories and their potential impacts are quantified according to characterization factors.

### **1.3.3.4 Interpretation**

In the interpretation phase, the results are checked and evaluated to see that they are consistent with the goal and scope definition and that the study is complete.

### **1.3.3.5 Use of LCA**

Large companies use LCA as a way of identifying environmental hot spots and to develop and advertise their environmental management strategies. LCA studies are often conducted by industry associations and environmental concepts and tools research. Governmental departments around the world are active promoters of LCA. Government use LCA for data collection and developing more effective environmental policies related to materials and products. There are many universities researching and developing LCA methodology and data. Software packages for conducting LCA studies are available namely GaBi5, Eco-Invent, Simapro etc.

## ***1.3.4 Environmental Impact Assessment***

The International Association for Impact Assessment (IAIA) defines an environmental impact assessment as “the process of identifying, predicting, evaluating and mitigating the biophysical, social, and other relevant effects of development proposals prior to major decisions being taken and commitments made” [25]. EIA is often regarded as a synonym for a local, point-source oriented evaluation of environmental impacts, which takes into account time-related aspects, specific geographic situation, and existing background pressure on the environment. Thus

EIA is a procedure that has to support decision making with regard to environmental aspects of a much broader range of activities. The effects of a project on the environment must be assessed in order to take account of concerns to protect human health, to contribute by means of a better environment to the quality of life, to ensure maintenance of the diversity of species and to maintain the reproductive capacity of the ecosystem as a basic resource for life.

## **1.4 Recent Advancements and Research Issues in Sustainability**

The advancement and research issues on sustainability in modern scenario are presented in the following subsections.

### ***1.4.1 Recent Advancements in Sustainability***

The performance indicators for sustainability include Financial health, Economic performance, Potential financial benefits, Trading opportunities, Air resources, Water resources, Land resources, Mineral and energy resources, Internal human resources, External population, Stake holder participation, Macro social performance, Market potential, End of life (EoL) disposal policy and Customer orientation. Several new indicators are also under development by the researchers. The assessment of sustainability was done using approaches namely multi grade fuzzy, fuzzy logic, neural networks, expert systems, grey relational approaches, fuzzy multi attribute decision making (MADM) approaches and several outranking and compromising multi criteria decision making (MCDM) methods. The assessment was done from environment, economy, society and business perspectives. The selection process in the context of sustainability is a MADM problem involving several attributes. Applications include concept selection, supplier selection, EoL alternative selection, design selection and alternate material selection. Sustainable material selection considers the environmental impacts associated with the materials and type of manufacturing process employed. Sustainable material selection enhances the product life and provides better EoL options.

### ***1.4.2 Challenges in Sustainability Research***

The contemporary manufacturing enterprises are facing the challenges towards developing eco friendly products. The development of performance indicators for sustainability is challenging as the developed indicators must reflect the exact

measurement of entities. Sustainability assessment is a complex yet a challenging task as the approach must address environment, economy, society and business aspects. Decision making in the context of sustainability has diversified applications starting from material selection, product development, process alternatives, supply chain practices, technology selection, EoL disposal policy selection etc.,. The effectiveness of the decision making process in the viewpoint of sustainability depends on the selection of appropriate criteria, analysis of interdependency between them and choice of appropriate MADM method. LCA of manufacturing processes gains much importance to assess the environmental impacts associated with them.

### ***1.4.3 Research Centres Involved in Sustainability Research***

As the research studies on sustainable development expands, number of research centres have grown up around the world. Environmental Sustainability Research Centre (ESRC) concentrates on innovative and interdisciplinary research concerning the environment, sustainability and social-ecological resilience. Sustainable Europe Research Institute (SERI) supports decision makers in the areas of sustainable development with information relevant to their goals within the framework of national, regional and European sustainable policies. International Research Institute in Sustainability (IRIS) is known for its state of art research and for participatory research methodologies which support change for sustainability. IRIS focuses on people's responses to sustainability. Science and Technology can assist in defining problems and identifying solutions but ultimately changes towards sustainability are dependent on people's decisions, choices and actions. Institute for Sustainability Solutions Research (ISSR) mainly focuses on solutions-based research and outreach in sustainability, involving students wherever possible, through their studies and most especially through involvement in opportunities in social and sustainable enterprise. Centre for Sustainable Manufacturing and Reuse/Recycling Technologies (SMART), Loughborough University, United Kingdom focuses on advanced technologies related to sustainable manufacture and to promote effective end-of-life product retirement strategies in the wake of European legislation and Corporate Social Responsibility. Centre for Sustainable Development in Cambridge University Engineering operates on a multi-disciplinary approach to focus on the context and complexity in which engineering products and services are delivered. Centre for Sustainable Technologies (CST), Indian Institute of Science (IISc), Bangalore, India focuses on inter-disciplinary research and technology development centre for providing sustainable solutions, primarily dealing with energy, buildings and environment. CST focuses on promoting sustainable technologies tailored to suit local conditions of resource availability and habitation. Innovation, Design Study and Sustainability Laboratory (IDeaS Lab) under Centre for Product Design and Manufacturing (CPDM) located at Indian Institute of Science (IISc), Bangalore, India has pioneered

research in the areas of design creativity, design methodology, sustainability, collaborative innovation, knowledge and life cycle management, and design research methodology.

#### ***1.4.4 Sustainability in Mechanical Engineering Education***

The meaning of sustainable development needs to be discussed well to understand its significance in contemporary scenario. The concept of TBL needs understanding from profitability, people and planet perspectives. Tools/Techniques of sustainability require in-depth understating among the students. Vital tools/techniques include Environmentally Conscious Quality Function Deployment (EC-QFD), Life Cycle Analysis (LCA), Design for Environment (DfE), Design for disassembly, R3 (Reuse, Recycle and Reduce), R6 (Reuse, Recycle, Recondition, Recovery, Reduce and Repair) cycles etc. Environment Impact Assessment (EIA) methods such as Institute of Environmental Sciences, University of Leiden (CML), Eco Indicator (EI) deserves proper attention from the viewpoint of theory and selection criteria. Various EoL disposal methods such as R6 require proper understanding so as to select the best alternative. Industrial applications of sustainability and Corporate Social Responsibility issues also need to be highlighted in Mechanical Engineering Curriculum.

### **1.5 Conclusions**

Efforts are required to improve curriculum to suit the engineering education for sustainable development. The modern mechanical engineers are required to focus on sustainable solutions to enable positive impact for the benefit of society. In this context, this chapter provides the insights on sustainable development from the perspective of sustainability orientations, Sustainability pillars, sustainable concepts, tools/techniques of sustainability. The challenges faced from Sustainability viewpoint, recent advancements, research centers involvement in sustainability research is also being highlighted. The principles being discussed in this chapter will provide a transition in mechanical engineering education to suit sustainable development.

### **References**

1. Vinodh S, Rathod G (2009) Integration of ECQFD and LCA for sustainable product design. *J Clean Prod Res* 37:1075–1091
2. Vinodh S, Girubha RJ (2012) Sustainable concept selection using ELECTRE. *Clean Technol Environ Policy*. doi [10.1007/s10098-011-0429-2](https://doi.org/10.1007/s10098-011-0429-2)

3. Todorov VI, Marinova D (2009) Models of sustainability. Paper presented at the 18th World IMACS/MODSIM Congress, Cairns, Australia 13–17 July 2009
4. Labuschagne C, Brent AC, Erck RPGV (2005) Assessing the sustainability performances of industries. *J Clean Prod* 13:373–385
5. Vinodh S (2009) Assessment of sustainability using multi-grade fuzzy approach. *Clean Technol Environ Policy* 13(3):509–515
6. Bovea MD, Vidal R (2004) Materials selection for sustainable product design: a case study of wood based furniture eco-design. *Mater Des* 25:111–116
7. Giudice F, Rosa GL, Risitano A (2005) Materials selection in the life-cycle design process: a method to integrate mechanical and environmental performances in optimal choice. *Mater Des* 26:9–20
8. Bolin CA, Smith ST (2011) Life cycle assessment of borate-treated lumber with comparison to galvanized steel framing. *J Clean Prod* 19:630–639
9. Smith S, Sheate W (2001) Sustainability appraisal of English regional plans: incorporating the requirements of the EU strategic environmental assessment directive. *Impact Assess Proj Apprais* 19:263–276
10. Gibson R (2001) Specification of sustainability-based environmental assessment decision criteria and implications for determining significance in environmental assessment. Available at <http://www.sustreport.org/downloads/SustainabilityEA.doc> (Accessed 12 August 2003)
11. Azapagic A (2003) Systems approach to corporate sustainability—a general management framework. *Trans. IChemE* 81(Part B): 303–316
12. IChemE (2002). Sustainable development progress metrics recommended for use in the process industries. Available at <http://www.icheme.org/sustainability/metrics.pdf> (2004)
13. Ghadimi P, Azadnia AH, Yusof NM, Saman MZM (2012) A weighted fuzzy approach for product sustainability assessment: a case study in automotive industry. *J Clean Prod* 33:10–21
14. Vinodh S, Jayakrishna K, Giruba RJ (2012) Sustainability assessment of an automotive organization using Fuzzy Kano's model. *Int J Sustain Eng* 6(1):1–9
15. Amindoust A, Ahmed S, Saghafinia A, Bahreininejad A (2012) Sustainable supplier selection: a ranking model based on fuzzy inference system. *Appl Soft Comput* 12(6):1668–1677
16. Sakao T (2007) A QFD-centred design methodology for environmentally conscious product design. *Int J Prod Res* 25:18–19
17. Kurk F, Eagan P (2008) The value of adding design-for-the-environment to pollution prevention assistance options. *J Clean Prod* 16:722–726
18. Fiksel (2009) *Design for environment: creating eco-efficient products and processes*. McGraw-Hill, USA
19. Jeswiet J, Hauschild M (2005) Ecodesign and future environmental impacts. *Mater Des* 26:629–634
20. John E, Michael L (1997) The development and implementation of DfE programmes. *J Sustain Prod Des* 1(1):17
21. Benedetto LD, Klemes J (2009). The environmental performance strategy, map: an integrated LCA approach to support the strategic decision-making process. *J Clean Prod* 17:900–906
22. Griese H (2005) Ecodesign and beyond-key requirements for a global sustainable development. Paper presented at the international conference on asian green electronics, Berlin, Germany
23. Kaebnick H, Antiyasari M, Kara S (2001) Technical and economic model for end-of-life options of industrial products. *Int J Environ Sustain Dev* 1(2):171–183
24. Masui K, Sakao T, Kobayashi M, Inaba A (2003) Applying quality function deployment to environmentally conscious design. *Int J Quality Reliab Manag* 20(1):90–106
25. Senecal P, Goldsmith B, Conover S, Sadler B, Brown K (1999) *Principle of environmental impact assessment best practice*. International association for impact assessment, USA

# Chapter 2

## Application of Solar Distillation Systems with Phase Change Material Storage

S. K. Shukla

**Abstract** This chapter presents the analysis of a solar distillation system with phase change material storage system. There is always a scarcity of clean and pure drinking water in many developing countries. Water from various sources is often brackish (i.e. contain dissolved salts) and/or contains harmful bacteria and therefore, cannot be used for drinking purpose. In addition, there are many coastal areas where seawater is abundant but potable water is not available in sufficient quantity. Apart from drinking purpose, pure water is also useful for health and industrial purposes such as hospitals, schools and batteries etc. Many parts of India, particularly, rural areas, coastal areas and many urban areas too, have a major drinking water problem. Sufficient drinking water, at accepted purity level, is just not available.

### 2.1 Introduction

Contaminated drinking water is one of the reasons of major health hazards responsible for almost 90 % of the health problems in rural areas. Women and children are mostly affected because they are quite vulnerable to water borne diseases. Generally, women are responsible for fulfilling the requirement of water in their household and thus, spend their large chunk of time for procuring the same. Out of 40–50 litres per capita per day (lpcd) of water requirement for domestic consumption, only 2 lpcd is the drinking water. A total amount of 5–10 lpcd water is needed for drinking and cooking purposes and thus, it is only this quantity of water that needs to meet the stringent quality standards of potability prescribed by W.H.O. or other similar agencies, whereas the remaining amount of water needed for washing and cleaning can be of intermediate quality. Intensive use of chemical fertilizers in agriculture and enhanced Industrial

---

S. K. Shukla (✉)  
Department of Mechanical Engineering, I.I.T (B.H.U.),  
Varanasi, U.P. 221005, India  
e-mail: skshukla.mech@itbhu.ac.in

activities cause the natural and inorganic pollutants to leach down to underground sub surface water and hence, the drawing of water through Hand-Pumps may not remain safe for drinking purpose. Keeping in view the poor paying capacity of people, water supply to remote areas through pipeline could be uneconomical and moreover, it also encourages wasteful use of high quality water in washing, cleaning and toiletries. Therefore, for economical and sustainable water management system, it is important to supply water at appropriate level of quality, which is suitable enough for the kind of use for which it is meant [29–33].

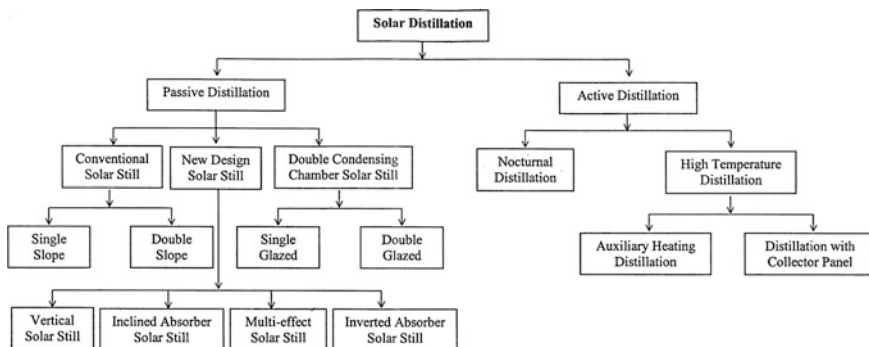
Water and energy are two inseparable items that govern our lives and promote civilization. Looking into the history of mankind one finds that water and civilization were also two inseparable entities. It is not a coincidence that all great civilizations were developed and flourished near large bodies of water. Rivers, seas, oases and oceans have attracted mankind to their coasts because water is the source of life. The transportation of drinking water from far-off regions is usually not economically feasible/desirable, desalination of available brackish water has been considered as an alternative approach. Conventional desalination processes based on distillation involve phase change. These are multistage flash distillation (MSF), multieffect distillation (MED) and vapor compression (VC), which could be thermal VC or mechanical VC. MSF and MED processes consist of a set of stages at successfully decreasing temperature and pressure. MSF process is based on the generation of vapors from seawater or brine due to a sudden pressure reduction when seawater enters to an evacuated chamber. The process is repeated stage by stage at successively decreasing pressure. This process requires an external steam supply, normally at temperature around 100 °C. The maximum temperature is limited by the salt concentration to avoid scaling and this maximum limits the performance of the process. On MED, vapors are generated due to the absorption of thermal energy by the seawater. The steam generated in one stage or effect is able to heat the salt solution in the next stage because next stage is at lower temperature and pressure. The performance of the process is proportional to stages or effects. MED plants normally use an external steam supply at temperature about 70 °C. On TVC and MVC, after initial vapor is generated from the saline solution, this vapor is thermally or mechanically compressed to generate additional production. On the other hand the membrane processes such as reverse osmosis (RO) and electro-dialysis (ED) do not involve phase change. The first one require electricity or shaft power to drive the pump that increase the pressure of saline solution to that required. This required pressure depends on the salt concentration of the resource of saline solution, which is normally around 70 bar for sea water desalination. Both of them RO and ED (membrane processes) are used for brackish water desalination whereas distillation is applicable to the entire range of salinities up to seawater. In fact it is the only process, which removes with certainty any organisms (bacteria, viruses and also pyrogens) contained in feed water. Theoretically distillation is capable of removing all non-volatile matter. In practice, however, some carryover of dissolved and colloidal matter into the distillate may take place. End product purity expressed in specific conductivity is between 4.0 and 0.066  $\mu\text{g}/\text{cm}$  at 25 °C, depending upon the technique used.



Since the cost of heat plays a decisive role in various distillation processes, viz., single stage, multistage distillation, flash distillation and vapor compression distillation; it seems advantageous to harness the heat of the sun for this purpose. Over a period of 100 years several types of solar stills have been designed and tested. There has been a significant progress in the field of solar distillation during the past four decades, perhaps, due to the general increase of interest in solar energy utilization. Because of cost free energy and low operating cost, as there are no moving parts involved in these systems, the solar distillation shows a comfortable economic advantage over other seawater distillation processes. Further, solar distillation requires simple technology and less maintenance so that it can be used at any rural place. Thus, in order to capture this very advantage of distillation process, cost effective Solar Stills have been designed and developed. And they have proved their performance and cost effectiveness.

When drying foods, the key is to remove moisture as quickly as possible at a temperature that does not seriously affect the flavor, texture and color of the food [1]. If the temperature is too low in the beginning, microorganisms may grow before the food is adequately dried. If the temperature is too high and the humidity is too low, the food may harden on the surface. This makes it more difficult for moisture to escape and the food does not dry properly. Although drying is a relatively simple method of food preservation, the procedure is not exact.

For the preservation of these agricultural products, the convective hot air drying using fossil fuels/grid-electricity is the most common technique employed in commercial dryers around the globe. However, due to unreliable or too expensive for the farmer to utilize them, more emphasis is being paid to solar energy as an alternative source for such applications [2]. In addition, it has tremendous potential especially in several regions of the world, where this source is abundantly available. In past four decades, various types of solar dryers have been designed, developed and tested with the aim of achieving faster drying of food product at a minimum cost. Ekechukwu and Norton [3] presented a comprehensive review on design, construction and operation of different types of solar dryers. However, all these dryers can be broadly grouped into three major types as direct, indirect and mixed mode, depending on arrangement of system components and mode of solar heat utilization [4]. The operation of these dryers is primarily based on the principle of natural or forced air circulation mode. In many rural regions of developing countries, the farmers have been preferably adopting natural convection over forced mode operated dryer, since it is inexpensive to construct and easy to operate without the need of grid connected electricity and supplies of other non-renewable sources of energy. In addition, natural convection cabinet dryer of direct type has been popular among farmers especially in India because of its ability for drying 10–15 kg fruits and vegetables at household level [5, 6]. Selection of solar dryer for a particular food product is primarily governed by quality requirements and economic factors. The common practice in predicting performance of solar energy system is to solve a set of several inter-related steady state heat balance equations representing various components [7–9].



**Fig. 2.1** Classifications of solar stills

### 2.1.1 Solar Distillation

Solar distillation uses the heat of the sun directly in a simple piece of equipment to purify water. The equipment, commonly called a solar still, consists primarily of a shallow basin with a transparent glass cover. The sun heats the water in the basin, causing evaporation. Moisture rises, condenses on the cover and runs down into a collection trough, leaving behind the salts, minerals, and most other impurities, including germs. Although it can be rather expensive to build a solar still that is both effective and long-lasting, it can produce purified water at a reasonable cost if it is built, operated, and maintained properly. The classification of solar distillation process is given in Fig. 2.1.

### 2.1.2 Basin Type Solar Still

A basin type solar still is a simple device for obtaining potable water from contaminated or saline water using solar energy. Such a solar still is shown in Fig. 2.2. It consists of a blackened tray containing saline water, covered with a sloping glass roof. The collection troughs are attached at the lower edges of the glass covers. The water condenses on the inner surface of the glass cover, flows into the collection troughs due to gravity. The pure water is removed as product.

The main energy flows in the basin type solar still have been shown in the flow chart Fig. 2.3. Solar energy, which is transmitted by the glass cover and absorbed by the water and the blackened tray, heats the water in the basin. This absorbed energy conducted through the haze; radiated, or convected to the glass cover, or is carried by evaporation of water to the glass cover. The cover is substantially transparent to solar radiation and opaque to infrared radiation and serves as a condenser for the saturated vapour within the still. The glass cover transfers heat to the surroundings; by radiation to the sky and by convection to the ambient air.

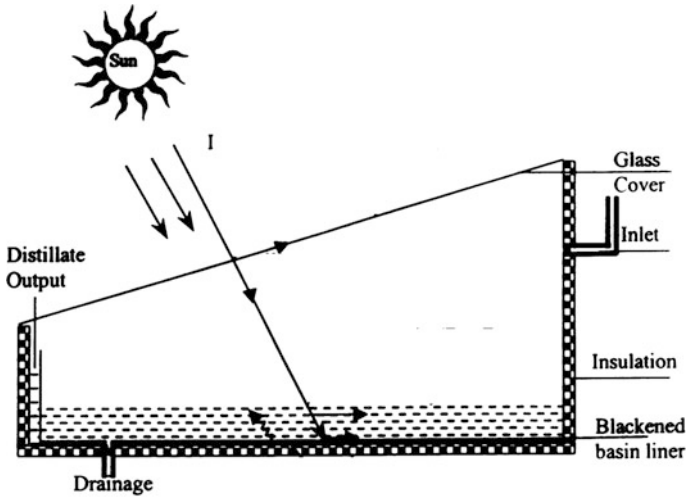


Fig. 2.2 Single basin solar still

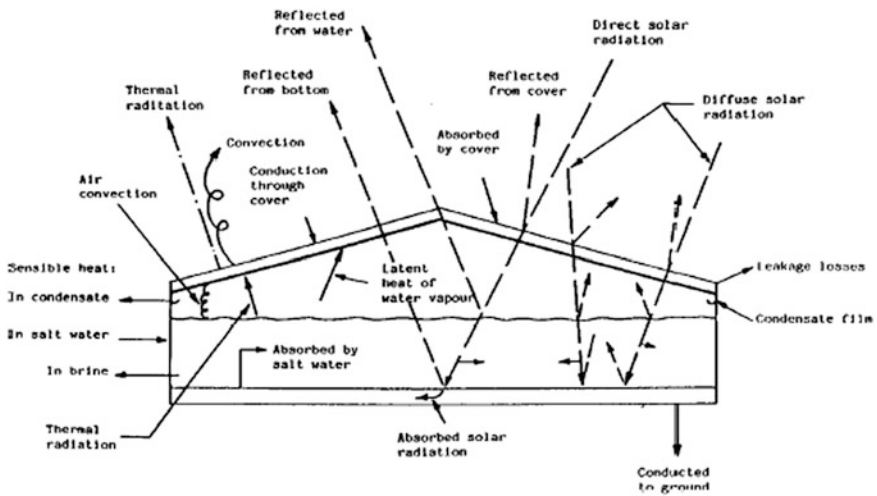


Fig. 2.3 Energy flow diagram

The basin type solar stills have been covered in the textbooks [9, 10]. The constructional and operational features of still have been described in the technical literature [9–20]. These features of the still are summarized by Dunkle [15]. A basin type solar still may be erected directly on dry ground, or well insulated at the bottom by a suitable insulating material. To cut down heat losses from sides and edges suitable insulation may be provided around the basin. However, it has been pointed out that the heat losses to the ground are small in case of larger basins even

when uninsulated [10, 14, 15]. Cooper [16] has explained that for large stills most of the heat conducted into the ground during the day is conducted back during the night. Malik et al. [10] have carried out a periodic heat transfer analysis and concluded that ground and edge losses are only 2 % of the incident energy.

### 2.1.3 Efficiency of the Still

If  $Q_i$  (in Joules/m<sup>2</sup> day) is the amount of solar energy incident on the glass cover of a still and  $Q_e$  (in Joules/m<sup>2</sup> day) is the energy utilised in vaporizing water in the still, then the daily output of distilled water  $M_e$  (in kg/m<sup>2</sup> day) is given by

$$M_e = \frac{Q_e}{l} \quad (2.1)$$

where  $l$  (in Joules/kg) is the latent heat of vaporization of water. The efficiency of the still is given by  $\frac{Q_e}{Q_i}$ . An ideal still is considered to be a still with zero conductive heat losses and with zero heat capacities of water, glass and insulation. Cooper [16] has shown that efficiency of such a still as high as 60 % for high values of solar insolation.

### 2.1.4 The Basic Heat Transfer Modes in a Solar Still

The physics of solar stills is mainly a superposition of heat and mass transfer by molecular diffusion and buoyant convection and of radiative heat exchange between the surrounding surfaces. The latter is practically independent from the foregoing two, but these are very complex processes, whose components must not be investigated separately.

The operation of a solar still is governed by various heat transfer modes, convection and radiation being the prominent ones. Free convection accompanied by evaporative mass transfer and radiation are modes of heat transfer between the water surface and the glass cover inside the still. The coefficient of heat transfer is incorporated in Nusselt number which is related to the Grashof and Prandtl numbers as,

$$\begin{aligned} N_u &= f(G_r \cdot P_r) \\ N_u &= \frac{h_{cw} \cdot x}{k_f} \\ G_r &= x^3 \cdot \rho_f^2 \cdot g \cdot \beta' \cdot \Delta T / \mu_f^2 \\ P_r &= \frac{C_p \cdot \mu_f}{k_f} \end{aligned} \quad (2.2)$$

In the case of a still  $x$ , the characteristic dimension of the system is the distance between the surface of water and the glass cover.

The steady state performance of a basin type solar still can be predicted following the solar energy textbooks (Malik et al. [10] Duffie and Beckman, [9]). The textbooks have recommended Dunkle's [15] heat transfer model which has been in the technical literature for some time. This heat transfer model is described below. The basic heat transfers in a still can be grouped into two categories:

- (a) Inner heat transfer (between the water surface and the glass cover):

The basic heat transfer modes between the water surface and the glass cover are convection, radiation and evaporation. The radiative, convective, and evaporative heat transfer coefficient between water surface and glass cover (Dunkle [15]) are given below assuming water surface and glass cover as two infinite parallel plates.

- (i) The radiative heat transfer coefficient,

$$h_{rw} = \frac{\sigma \left[ (T_w + 273)^4 - (T_g + 273)^4 \right]}{\left( \frac{1}{\epsilon_g} + \frac{1}{\epsilon_w} - 1 \right) (T_w - T_g)} \quad (2.3)$$

and rate of heat transfer by radiation is expressed as

$$Q_{rw} = h_{rw}(T_w - T_g)$$

- (ii) The heat transfer between the water surface and glass cover is by free convection. The convective heat transfer coefficient,

$$h_{cw} = 0.884 \left[ (T_w - T_g) + \frac{(P_w - P_g)(T_w + 273)}{(268.9 \times 10^3 - p_w)} \right]^{1/3} \quad (2.4)$$

and rate of convective heat transfer is expressed as

$$Q_{cw} = h_{cw}(T_w - T_g)$$

Dunkle [15] made use of the relation between Nusselt number and Grashof number for turbulent convection in a horizontal enclosed air space. The buoyancy term in Grashof number is modified by the density effect due to composition (due to evaporation) as well as temperature based in part on experiments with the evaporation of water into quiescent air [19]. Malik et al. [10] have explained the mathematical development. The correlation assumes that the air is saturated with water vapour at the temperature of water in the basin and at the temperature of glass cover.

(iii) The evaporative heat transfer coefficient,

$$h_{ew} = \frac{9.15 \times 10^{-7} h_{cw} (P_w - P_g) h_{fg}}{(T_w - T_g)} \quad (2.5)$$

is based on the relation between mass transfer and heat transfer, assuming a Lewis number 1.0.

(b) Outer heat transfer (between the glass cover and the environment):

(i) The radiative heat transfer coefficient between glass cover and the “sky” (atmosphere) is given by,

$$h_{rg} = \sigma \varepsilon_g (T_g^2 + T_s^2) (T_g + T_s) \quad (2.6)$$

where  $T_s$  is the apparent (effective) temperature of the atmosphere which acts as the sink for radiative heat transfer.

(ii) Bottom and Edge Heat Transfer

The heat is also transferred or loss from water in ambient through the insulation and subsequently convection/radiation of the bottom or side surface of the box. Hence the bottom loss can be written as

$$U_b = \left[ \frac{1}{K_i/L_i} + \frac{1}{h_{cb} + h_{rb}} + \frac{1}{h_w} \right] \quad (2.7)$$

where  $L_i$ ,  $K_i$ ,  $h_{rb}$ ,  $h_{cb}$  and  $h_w$  are thickness, thermal conductivity of the insulation, radiative and convective heat transfer coefficients from bottom to ambient and the convective heat transfer coefficient between basin liner of the still and the water. However last two terms in the above expression do not contribute significantly and the heat loss can be considered only through conduction from bottom,

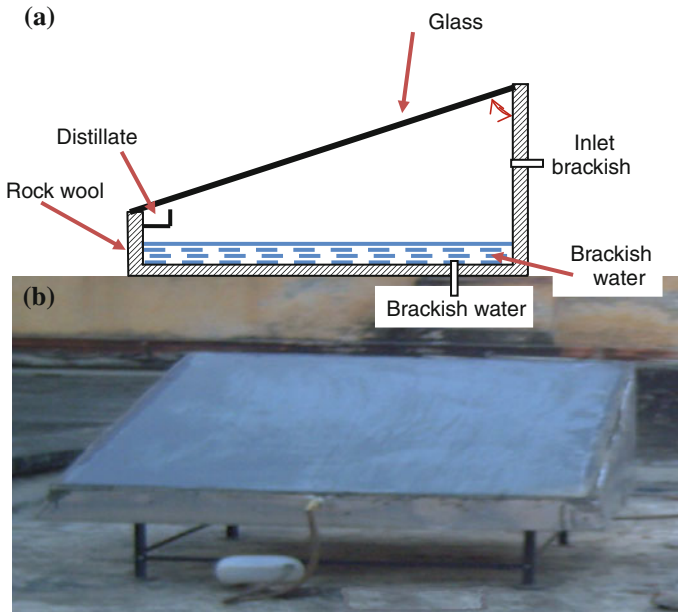
$$U_b = K_i/L_i \quad (2.8)$$

Similarly edge loss is also mainly by conduction through the insulation and can be written as

$$U_e = \frac{U_b \cdot A_e}{A_b} \quad (2.9)$$

where  $A_b$ ,  $A_e$  are bottom and edge areas of the still.

(iii) The convective heat transfer coefficient due to wind,



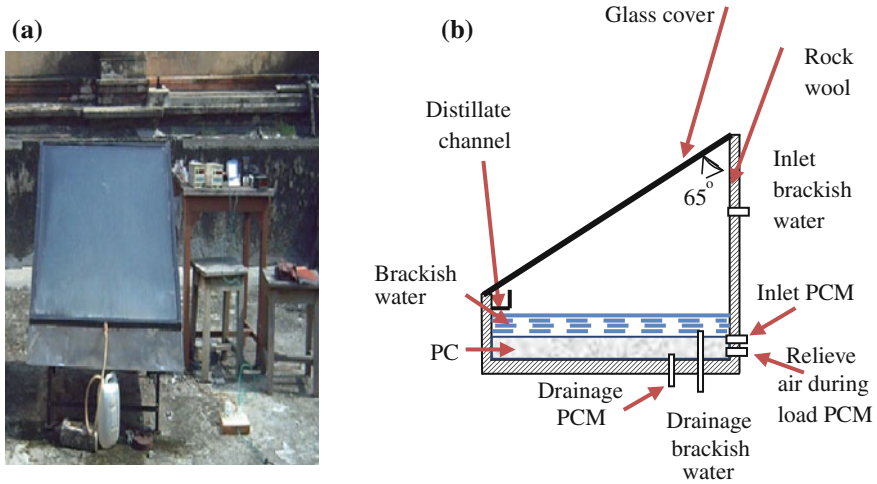
**Fig. 2.4** a Schematic diagram of passive solar still. b Photograph of passive solar still

For calculating the convection losses from the glass cover to the ambient air, wind heat transfer coefficient, may be expressed as a linear function of wind velocity [9]. However, as pointed out by Shukla et al. [12, 13] the heat transfer coefficient in natural environment can be larger by a factor of two or three.

## 2.2 Design of Solar Stills with Phase Change Material Storage

### 2.2.1 Conventional Solar Stills

A schematic diagram and photograph of the solar still is given in Fig. 2.4., installed at Renewable Energy Laboratory, Indian Institute of Technology, Banaras Hindu University, Varanasi, India. The basin is fabricated from black painted 2 mm thick mild steel having an area of 1 m<sup>2</sup> each. A vertical gap beneath the horizontal portion of the basin liner is provided to upload and/or unload the PCM through a PVC pipe which takes care of the volumetric expansion of the melting PCM as well. The operational and melting temperature of PCM, in fact, governs the applicability of different types of PCMs. The myristic acid relates to the class of fatty acids that have superior properties, such as, melting congruency, better chemical stability



**Fig. 2.5** A photograph of solar stills with PCM (a) A schematic diagram of the single slope-single basin solar still with the PCM (b)

non-toxicity and better thermal reliability over many other PCM [17–19] Schematic diagram and photograph of the conventional single slope solar still are shown in Fig. 2.4a, b, respectively.

### 2.2.2 Solar Still with Phase Change Material (*Myristic Acid*) Storage

A photograph and a schematic diagram of the single slope solar still with phase change material (PCM) as a storage medium, is given in Fig. 2.5. The basin, fabricated from a black painted mild steel sheet of thickness 2 mm, has an area of 1 m<sup>2</sup> each.

A vertical gap (0.05 m) beneath the horizontal portion of the basin liner is provided to upload and/or unload the PCM through a pipe which takes care of the volumetric expansion of the melting PCM as well. The operational and melting temperature of PCM, in fact, governs the applicability of different types of phase change materials. The bottom and sides of the basin are insulated by 5 cm thick layer of rock wool contained in an aluminium tray. The top cover of the still is made up of 4 mm thick window glass which inclines at an angle of 25° with horizontal, and has an average transmissivity ( $\tau$ ) of value 0.88. A U-shaped channel is used to collect the condensate from the lower edge of glass cover and carry it to storage. A temperature scanner (Altop Industries ltd, Sn.1005164, model ADT 5003) with resolution 0.1 °C and HTC DT-8811 Infrared thermometer range: –20 to 450 °C Spectral response: 6–14  $\mu\text{m}$  CE ASCO products had been used to



record the temperature with k-type thermocouples in solar stills. The solar radiation was measured by using the daystar meter Watts/m<sup>2</sup> ASCO products, the wind speed was observed by HTC Instruments AVM-07 Anemometer vane probe CE.

### 2.2.3 Energy and Exergy in a Solar Still with PCM

The energy and exergy of solar still with PCM are:

The energy and exergy of solar still with PCM are:

The energy efficiency of the solar still is:

$$mew_d = \sum_{24hr} mew_h \quad (2.10)$$

$$\eta_{en\_h} = \frac{mew_h \gamma}{A_b \sum I(t) * 3600} * 100 \% \quad (2.11)$$

$$\eta_{en\_d} = \frac{mew_d \gamma}{A_b I(t) * 3600} * 100 \% \quad (2.12)$$

The exergy balance of solar still with PCM is

$$\sum Ex_{lost} = \sum Ex_{in} - \sum Ex_{out} + \sum W \quad (2.13)$$

The exergy input is represent the exergy from sun

$$\sum Ex_{in} = A_b * I(t) * \left( 1 - \frac{4}{3} * \left( \frac{T_a + 273.15}{T_s + 273.15} \right) + \frac{1}{3} * \left( \frac{T_a + 273.15}{T_s + 273.15} \right)^4 \right) \quad (2.14)$$

The last term is the exergy of mechanical work. This term is negligible because no work in solar still.

$$\sum W = 0 \quad (2.15)$$

The exergy output or energy useful is due to evaporation heat transfer.

$$\sum Ex_{out} = \sum Ex_{useful} = \sum Ex_{ev} \quad (2.16)$$

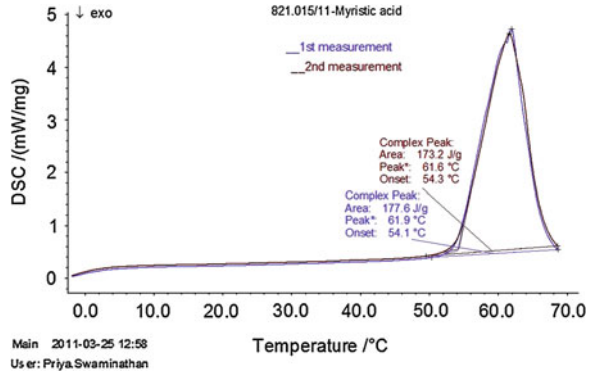
The exergy lost in solar still water is given by

$$\sum Ex_{lost} = mw * C_w * (T_w - T_a) * \left( 1 - \left( \frac{T_a + 273.15}{T_w + 273.15} \right) \right) \quad (2.17)$$

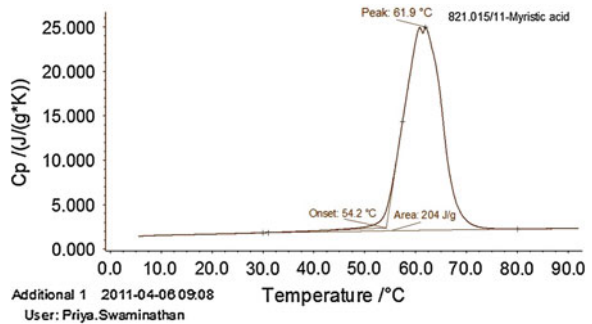
The second law efficiency as follows:

$$\eta_{EX} = Ex_{output} / Ex_{in} = 1 - Ex_{lost} / Ex_{in} \quad (2.18)$$

**Fig. 2.6** DSC curve for melting temperature and heat of fusion for myristic acid



**Fig. 2.7** DSC curve for specific heat ( $C_p$ ) of myristic acid



### 2.2.4 Selection of Phase Change Material

The measurement of PCM was done in Netzsch Technologies India Pvt Ltd, Chennai, India by using the differential scanning calorimetry (DSC) technique. The measurements were carried out under the following conditions for the evaluation of melting and heat of fusion as shown in Fig. 2.6:

- Temperature range: 0–80 °C.
- Heating rate: 10 K/min.
- Atmosphere: Static Air.

The measurements conditions of heat capacity of Myristic acid are as follows as shown in Fig. 2.7:

- Temperature range: 0–80 °C.
- Heating rate: 10 K/min.
- Atmosphere: Nitrogen.

The measurements were repeated to check the reproducibility of the results. Therefore, myristic acid, a by product of milk, has been used as a latent heat storage material due its low cost and easy availability. Table 2.1 summarizes the thermo-physical properties of myristic acid used in the experiment [19–23] (Table 2.2).

**Table 2.1** Thermo-physical properties of Myristic acid

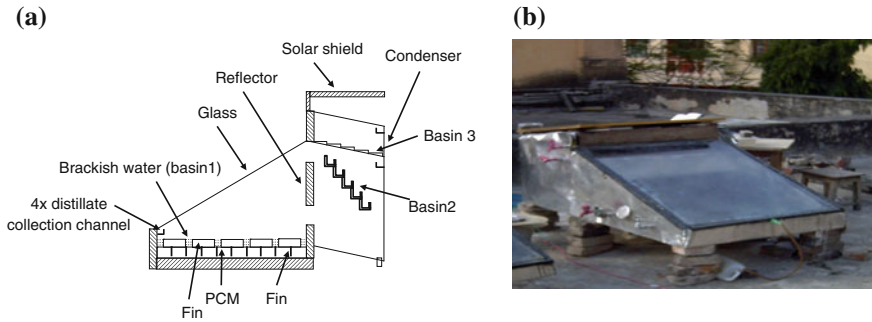
S. No	Properties	Value
1.	Melting point	50–54 [°C]
2.	Latent heat of fusion	177 [kJ kg <sup>-1</sup> ]
3.	Thermal conductivity	0.25 [22] [Wm <sup>-1</sup> °C <sup>-1</sup> ]
4.	<i>Specific heat</i>	
	Solid at 35 °C	1700 [Jkg <sup>-1</sup> °C <sup>-1</sup> ]
	Liquid at 55 °C	2040 [Jkg <sup>-1</sup> °C <sup>-1</sup> ]
5.	<i>Density</i>	
	Solid	990 [kgm <sup>-3</sup> ]
	Liquid	861 [kgm <sup>-3</sup> ]

**Table 2.2** Design parameters of solar still with and without PCM

S. No	Parameter	Values, units
1.	A <sub>b</sub>	1 [m <sup>2</sup> ]
2.	A <sub>g</sub>	1.1 [m <sup>2</sup> ]
3.	α <sub>b</sub>	0.95
4.	α <sub>g</sub>	0.05
5.	α <sub>w</sub>	0.05
6.	R <sub>g</sub>	0.05
7.	h <sub>2</sub>	5.7 + 3.8 v, [Wm <sup>-1</sup> °C <sup>-1</sup> ]
8.	K <sub>g</sub>	0.035 [Wm <sup>-2</sup> °C <sup>-1</sup> ]
9.	L <sub>g</sub>	0.004 [m]
10.	R <sub>w</sub>	0.05
11.	ε <sub>g</sub>	0.95
12.	ε <sub>w</sub>	0.95
13.	m <sub>w</sub>	30–50 [kg]
14.	m <sub>pcm</sub>	10–30 [kg]
15.	L <sub>ins</sub>	0.05 [m]
16.	L <sub>b</sub>	0.002 [m]
17.	K <sub>b</sub>	43 [Wm <sup>-2</sup> °C <sup>-1</sup> ]

### 2.3 Case Study of a New Design

Schematic diagram and photograph of the new design single slope solar still are shown in Fig. 2.8a, b respectively. The new design solar still and conventional solar still were designed based on the optimum inclination through the year for Varanasi city in India. The major components of this solar distillation system were evaporator, condenser and PCM chamber units. The two sides and back walls are covered with aluminium sheet plate, so these walls serve as the internal reflectors. This solar still forms three basins with saline water. Basin 1 was in the evaporator unit while basins 2 and 3 were stacked inside the condenser unit to recover heat from the first effect (Fig. 2a). Basin 1 of (1 × 1 m) of the test still (AR = 1) was constructed from mild steel sheet (0.002 m thick), painted black on the inner

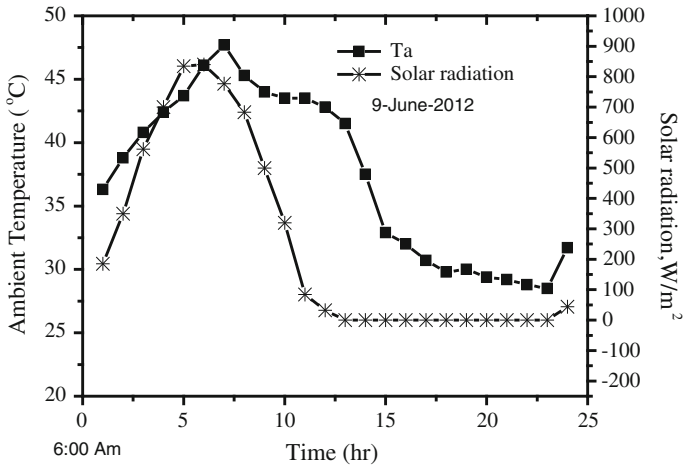


**Fig. 2.8** **a** Schematic diagram of new design solar still. **b** Photograph of a new design solar still

**Table 2.3** Thermo-physical properties of Lauric acid

S. No	Properties	Value
1.	Melting point	40–43.9 °C
2.	Latent heat of fusion	180 kJ/kg
3.	Thermal conductivity	0.16 Wm <sup>-1</sup> °C <sup>-1</sup>
4.	<i>Specific heat</i>	
	Solid at 25 °C	2.1 kJkg <sup>-1</sup> °C <sup>-1</sup>
	Liquid at 44 °C	3 kJkg <sup>-1</sup> °C <sup>-1</sup>
5.	<i>Density</i>	
	Solid	1007 kg/m <sup>3</sup>
	Liquid	862 kg/m <sup>3</sup>

surface to optimize absorption of solar radiation. Basin liners 2 and 3 were made from iron sheet (0.0015 m thick) but they were not painted to reduce resistance to heat conduction. Basin 2 includes stepped basin with 5 steps, with area of 0.3 m<sup>2</sup>. While basin 3 was inclined at 12° to enable distillate flow downward into the collection channels. The fins were added in basin 1 of NDSS to decrease the preheating time required for evaporating the still basin water. While using fins in the solar still, the area of the absorber plate increased. Hence, absorber plate temperature and saline water temperature increased. As the temperature difference between water and glass increases, productivity increased. In this work, five circular fins with height and diameter 35, 155 mm, respectively were used. Slender shaped fins were welded on the upper of PCM's chamber. The operational and melting temperature of PCM, in fact, governs the applicability of different types of phase change materials. The lauric acid relates to the class of fatty acids that have superior properties such as melting congruency, good chemical stability and non-toxicity, good thermal reliability over many other PCMs [39, 40]. The Table 2.3 shows thermo-physical properties of Lauric acid. A temperature scanner (Altop Industries ltd, Sn.1005164, model ADT 5003) with resolution 0.1 °C has been used to record the temperature with k-type thermocouples in both system for different location. The solar radiation passes through the glass cover to heat saline



**Fig. 2.9** Daily variation of ambient temperature and solar intensity

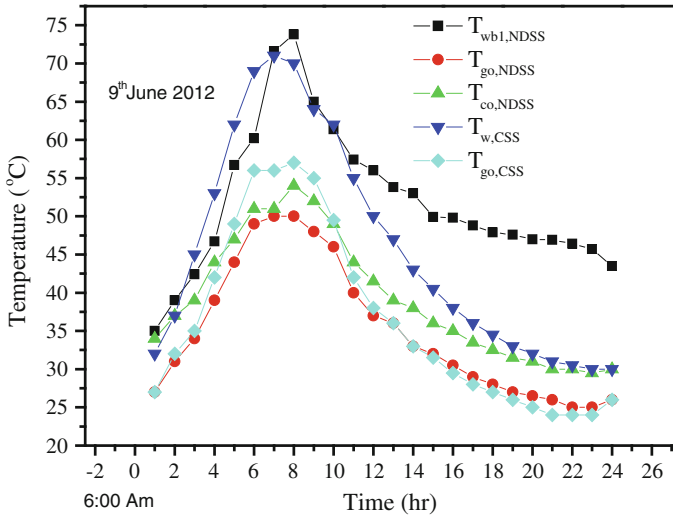
water in basin 1 (first effect). Then, vapour from the first effect flows upward and condenses when it gets into contact with the inner side of the glass cover at lower temperature while part of the vapour flows into the condensing chamber to heat water in basin 2 (second effect) and basin 3 (third effect). The transfer of water vapour from the still to the condenser could be done through one or more of the following mass transfer modes: diffusion, purging, and natural circulation.

## 2.4 Result and Discussion

An experimental study during 5th June 2012–18th June 2012 days were carried out to compare the thermal performance of NDSS and CSS. The mass water 10, 15, 20 and 30 kg are used for basin1. The mass water of basin 2 and basin 3 is 4.3 and 7.4 kg respectively. The mass of PCM is varies between 10 to 30 kg for NDSS. The mass water 10, 15, 20 and 30 kg are used for the CSS. Figure 2.9 shows the solar intensity and the hourly yield produced on 9th June 2012.

### 2.4.1 Temperature of System Components

Figure 2.10 shows the variation of the observed temperature of glass cover ( $T_{gc}$ ), saline water ( $T_{wb1}$ ), condenser cover ( $T_{co,NDSS}$ ), glass cover ( $T_{gc,CSS}$ ), saline water ( $T_{w,CSS}$ ) on a 9th June 2012. It is observed that the values of  $T_{gc}$  for the CSS ( $T_{gc,CSS}$ ) are higher than those of the NDSS ( $T_{gc,NDSS}$ ) from about 10:00 to 18:00 h, with maximum values of  $T_{gc,CSS} = 57^\circ\text{C}$  and  $T_{gc,NDSS} = 50^\circ\text{C}$ .



**Fig. 2.10** Variation of experimental temperature of saline water in basin1 of CSS and NDSS on 9th June 2012

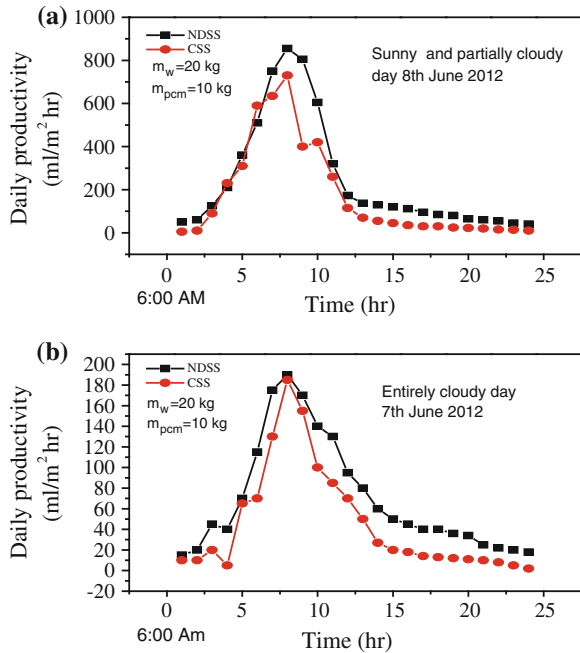
In addition, the temperature of water in basin for the CSS ( $T_{w,CSS}$ ) is higher than that of the NDSS ( $T_{wb1,NDSS}$ ) from about 8:00 to 13:00 h. But after that the temperature of water in basin for the CSS ( $T_{w,CSS}$ ) is lower than that of the NDSS ( $T_{wb1,NDSS}$ ). It should be mentioned that part of the heat from the evaporator basin flows into the condenser chamber by purging, diffusion and circulation which would tend to lower the glazing temperature of the NDSS [24, 25].

### 2.4.2 Sunny and Entirely Cloudy Day

An experimental study in sunny and partially cloudy 8th June 2012 and entirely cloudy 7th June 2012 days were carried out to compare the thermal performance of NDSS and CSS. Figure 2.11a, b illustrates the daily productivity of NDSS and CSS in typical sunny and cloudy days, respectively. The mass water of CSS is 20 kg, but for NDSS water masses are 20, 4.3 and 7.4 kg for basin1, basin2 and basin3 respectively.

Figure 2.11a indicates that hourly productivity in NDSS at morning is higher because NDSS has more area of water surface evaporation. But when the height of solar radiation about 9–11am the CSS is reading are higher than NDSS because, some of the absorbed solar energy is used to increase the PCM temperature and the solid fins of basin1, the absorber temperature is higher than the PCM and this trend is inverted in the low solar radiation intensity. It is clear from Fig. 5a, b that the hourly productivity at night is significant for NDSS because of using the stored

**Fig. 2.11** Variations of hourly productivity with time for NDSS and CSS, **a** typical sunny day, 8th June 2012 and **b** entirely cloudy day 7th June 2012

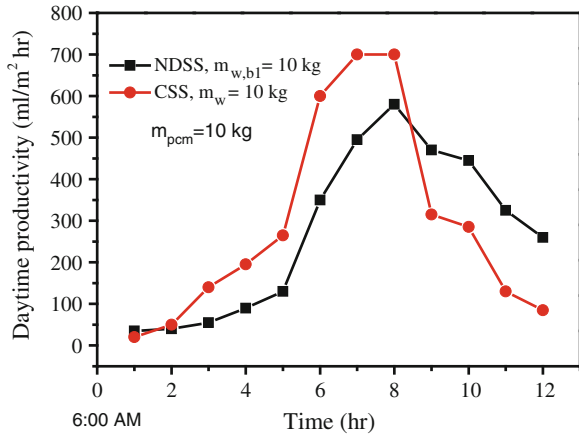


energy. Total productivity for NDSS and CSS is 5710 and 4295 ml/m<sup>2</sup> day for sunny day (8th June 2012) and 1675 and 1095 ml/m<sup>2</sup> day for entirely cloudy day (7th June 2012), respectively. It is also observed that there is a significant difference in the total productivity for the sunny day and entirely cloudy day.

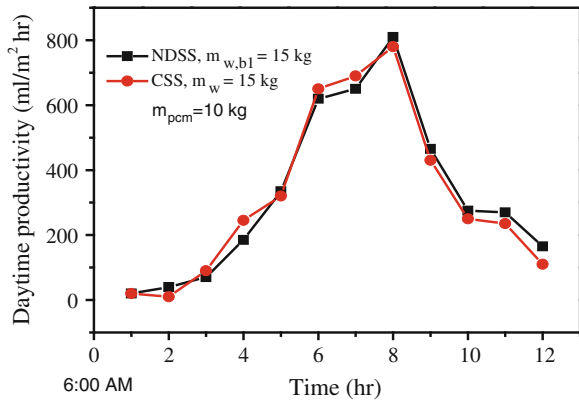
### 2.4.3 Daytime Productivity

Figures 2.12, 2.13, 2.14 and 2.15 reveals the variation of the daytime productivity with variation the mass of the mass of water 10, 15, 20 and 30 kg for basin1 in the NDSS and for the basin of CSS with 10 kg of PCM. The daytime productivity of CSS is higher than NDSS with 10 kg of water (see Fig. 2.12). The reason of this higher readings of CSS is because the little mass of water (2 cm depth of basin water), thus water immediately is heated and evaporated. But with NDSS the system is gradually heated; part of vapour will circulate inside the integrated condenser. These higher readings of CSS get reduced with mass of 15, 20 and 30 kg, see Figs. 2.12, 2.13 and 2.14. Figure 2.15 shows higher performance of NDSS. Furthermore NDSS yields better results in all figures after 2 pm.

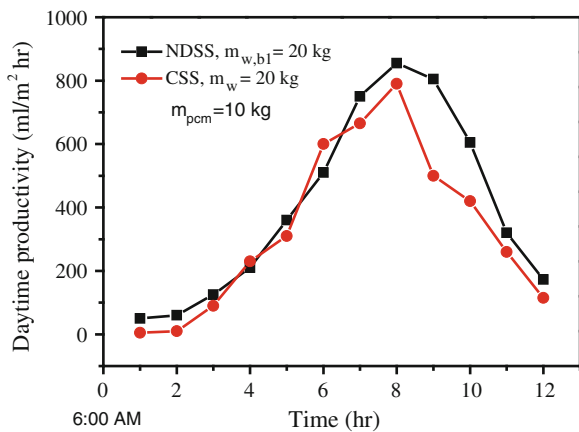
**Fig. 2.12** The daytime productivity of NDSS and CSS for  $m_{w,b1} = 10$  kg,  $m_{pcm} = 10$  kg and  $m_w = 10$  kg



**Fig. 2.13** The daytime productivity of NDSS and CSS for  $m_{w,b1} = 15$  kg,  $m_{pcm} = 10$  kg and  $m_w = 15$  kg

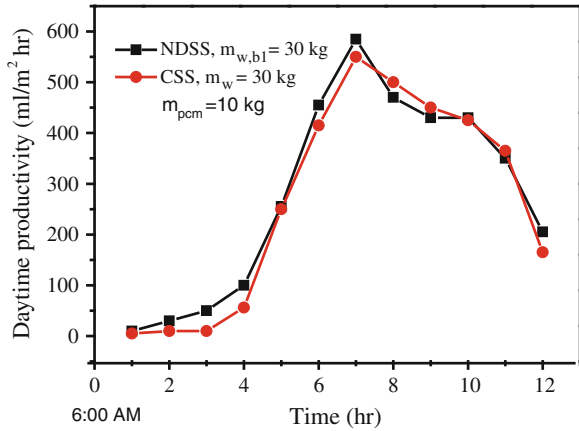


**Fig. 2.14** The daytime productivity of NDSS and CSS for  $m_{w,b1} = 20$  kg,  $m_{pcm} = 10$  kg and  $m_w = 20$  kg





**Fig. 2.15** The daytime productivity of NDSS and CSS for  $m_{w,b1} = 30$  kg,  $m_{pcm} = 10$  kg and  $m_w = 30$  kg



### 2.4.4 Variation of Daytime Productivity with Altered Mass of PCM

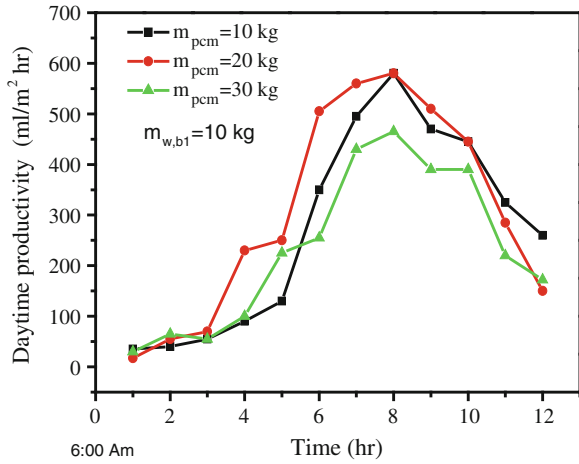
Figures 2.16, 2.17 and 2.18 present variations of the daytime productivities with  $m_{pcm}$  for different masses of basin1 water  $m_{w,b1}$  for NDSS. Assessment of Figures indicated that daytime productivity increases with time. Also the productivity of NDSS of  $m_{pcm} = 10$  kg with  $m_{w,b1} = 20$  kg is higher value as shown in figures. The reason may be that when the mass of PCM increases, the energy absorbed from basin liner by the PCM increases. Another reason may be the diffusion, purging and circulation inside the integrated condenser increase the convection, evaporating and condensation of vapor water, with the help of the cooler temperature of integrated condenser. The low glass temperature will increase the condensation of water vapor.

### 2.4.5 Variation of Productivity on Night with Altered Mass of PCM

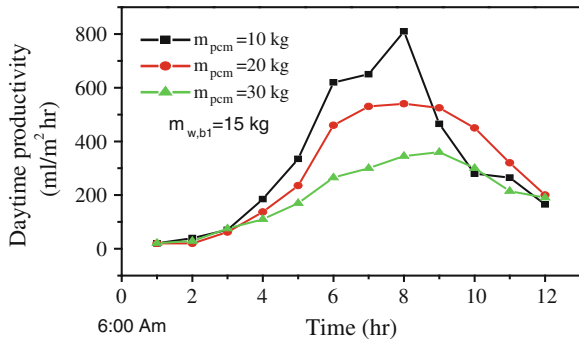
The key point of using PCM is to increase the productivity at night. Figure 2.19 depicts the influence of adding mass of PCM on the productivity during night time; Productivity increases with the increase of the mass of PCM.

The fluctuation in curves is because the difference in the solar radiation and the cloudy day. Table 2.4 depicts the productivity of 10, 20 and 30 kg of PCMM with 20 kg of water. The difference in productivity for 30 kg PCM with 10 kg PCM is nearly 193 ml/m<sup>2</sup> kg. This value if we compared with the daytime productivity of NDSS that has 20 kg of basin1 water with 10 kg of PCM. But the daytime productivity of NDSS that has 20 kg of basin1 water with 10 kg of PCM is greater

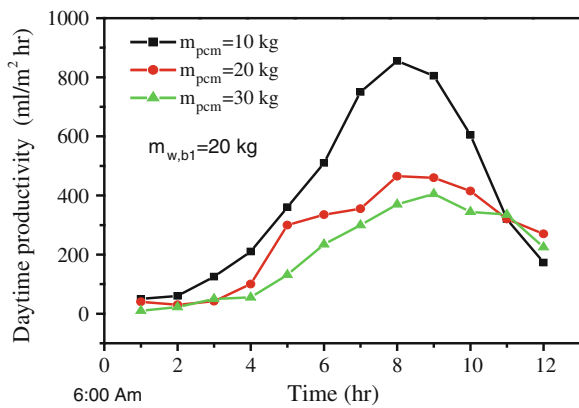
**Fig. 2.16** Variation of daytime productivity of NDSS for  $m_{pcm} = 10, 20$  and  $30$  kg with  $m_w = 10$  kg



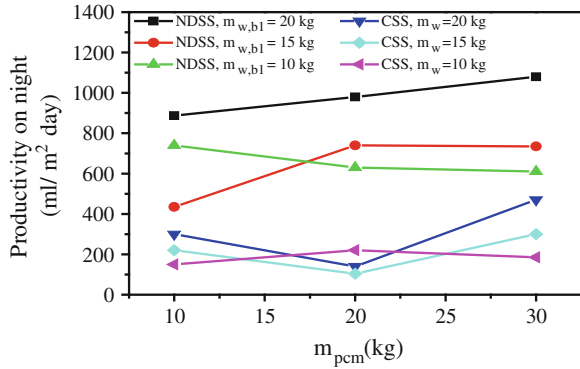
**Fig. 2.17** Variation of daytime productivity of NDSS for  $m_{pcm} = 10, 20$  and  $30$  kg with  $m_w = 15$  kg



**Fig. 2.18** Variation of daytime productivity of NDSS for  $m_{pcm} = 10, 20$  and  $30$  kg with  $m_w = 20$  kg



**Fig. 2.19** The productivity on night of NDSS and CSS for  $m_{pcm} = 10, 20$  and  $30$  kg with  $m_w = 10, 15$  and  $20$  kg



**Table 2.4** The daytime and on night productivity of NDSS of 10, 20 and 30 kg of PCM with 20 kg of water

Mass of PCM kg	Mass of water kg	Productivity on night, kg/m <sup>2</sup> day	Daytime productivity on kg/m <sup>2</sup> day
10	20	887	4.823
20	20	980	3.132
30	20	1,080	2.484

2.339 kg/m<sup>2</sup> day compared with 20 kg of basin1 water with 30 kg of PCM. Thus 10 kg of PCM is suitable for storage system. However the price of 30 kg of PCM is costly if compared with 10 kg of PCM.

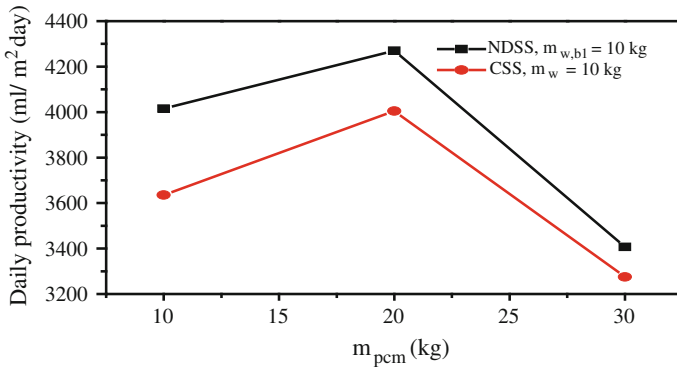
### 2.4.6 Variation of Daily Productivity with Altered Mass of PCM

The daily productivity is represented in Figs. 2.20, 2.21 and 2.22. It is clear that the highest productivity within 20 kg of water basin1 with 10 kg of PCM fee Fig. 2.16.

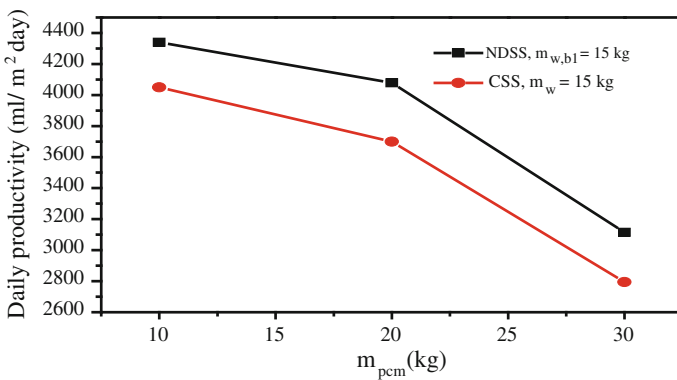
In order to have a comparison, amounts of total productivity of the designed stills and other still configurations are provided in Table 2.5.

## 2.5 Statistical Analysis

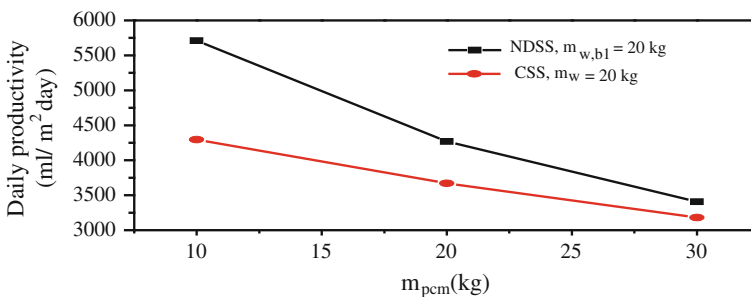
The productivity of NDSS was optimized using response surface methodology (RSM) provided by STATISTICA 8 software. A standard RSM design tool known as Central Composite Design (CCD) was applied to study the parameters affecting the performance of NDSS. The central composite experimental design (CCD) is a



**Fig. 2.20** Variation of daily productivity of NDSS with CSS for  $m_{w,b1} = 10$  kg,  $m_{pcm} = 10, 20$  and  $30$  kg and  $m_w = 10$  kg



**Fig. 2.21** Variation of daily productivity of NDSS with CSS for  $m_{w,b1} = 15$  kg,  $m_{pcm} = 10, 20$  and  $30$  kg and  $m_w = 15$  kg



**Fig. 2.22** Variation of daily productivity of NDSS with CSS for  $m_{w,b1} = 20$  kg,  $m_{pcm} = 10, 20$  and  $30$  kg and  $m_w = 20$  kg

**Table 2.5** Comparison between total productivity of the designed still and other still configurations

Still type	Date	Productivity (kg/m <sup>2</sup> day)
Designed NDSS (present work)	08/06/2012	5.71
Designed still with LHTESS [26]	23/05/2009	4.85
Designed still with a separate condenser [26]	2009	4.59
Inclined type with black fleece [12]	May, 2004	2.995
Basin type only [27]	16/08/2006	1.88
Basin type with sponge [27]	13/08/2006	2.26
Basin type with wick [27]	06/04/2006	4.07
Basin type with fin [27]	28/08/2006	2.81

**Table 2.6** Independent variables and levels used for experimental design

Independent variables	Codes	Variable levels		
		-1	0	+1
Temperature difference between water and glass cover (°C)	X <sub>1</sub>	6	13	20
Ambient temperature (°C)	X <sub>2</sub>	20	35	50
Solar radiation W/m <sup>2</sup>	X <sub>3</sub>	260	630	1,000

suitable design for sequential experiments to obtain appropriate information for testing lack of fit (Table 2.6).

### 2.5.1 Optimization of Parameters

The response surface methodology was used for the optimization of parameters. Among the models that can be fitted to the response (linear, two factor interaction (2FI) and quadratic polynomial), the quadratic model was selected as it is the best model due to its highest order polynomial with significance of additional terms. The model equation based on the coded values (X<sub>1</sub>, X<sub>2</sub> and X<sub>3</sub> as temperature difference between water and glass cover, ambient temperature and solar radiation respectively) for the productivity of NDSS was expressed by Equation below.

$$Y = 378.97 + 84.99X_1 - 50.65X_1^2 + 38.7X_2 - 31.9X_2^2 + 171.1X_3 - 18.7X_3^2 - 6.8X_1X_2 + 30.6X_1X_3$$

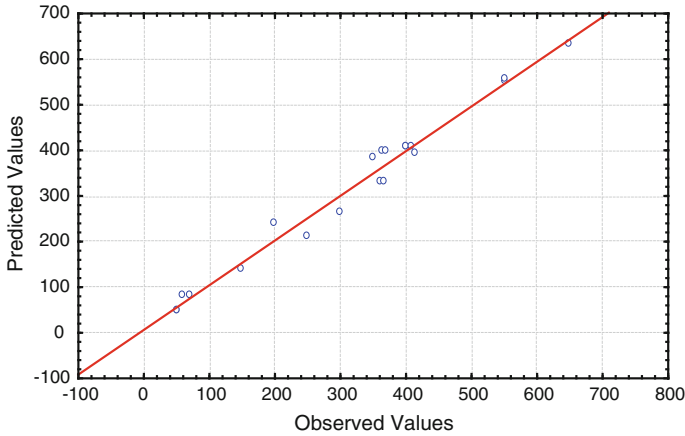
The result of statistical analysis of variance (ANOVA) was carried out to determine the significance and fitness of the quadratic model as well as the effect of significant individual terms and their interaction on the chosen responses. The *p* value (probability of error value) is used as a tool to check the significance of each regression coefficient, which also indicates the interaction effect of each cross product. The smaller the *p*-value, the bigger is the significance of the corresponding coefficient [28] In the case of model terms, the *p*-values less than 0.05

indicated that the particular model term was statistically significant. From the ANOVA results, the main model terms suggested that variables with significant influence on productivity of NDSS response were temperature difference between water and glass cover ( $X_1$ ) ambient temperature ( $X_2$ ), solar radiation ( $X_3$ ), and the interaction terms were found to exist between the main factors ( $X_1 X_2$  and  $X_1 X_3$ ), while the significant quadratic terms were temperature difference between water and glass cover ( $X_1^2$ ), ambient temperature ( $X_2^2$ ) and solar radiation ( $X_3^2$ ). The lack of fit test with p-value of 0.0521, which is not significant (p-value > 0.05 is not significant) showed that the model satisfactorily fitted to experimental data. Insignificant lack of fit is most wanted as significant lack of fit indicates that there might be contribution in the regress or-response relationship that is not accounted for by the model [38]. The predicted values versus actual values for productivity of NDSS with adjusted  $R^2$  value of 0.95 indicated that the predicted values and experimental values were in reasonable agreement Fig. 2.23. It means that the data fit well with the model and give a convincingly good estimate of response for the system in the range studied. The developed second-order regression model is complex with many variables. It is difficult to understand the effect of different independent variables from the regression model, but graphical representations are easier to interpret. Contour and response surface plots were drawn (not shown in the paper) to observe the effect of solar radiation, ambient temperature, difference temperature of water and glass on productivity of NDSS. These plots were generated by holding one of the variables at its mid-point and varying the other two variables to obtain the response. The elliptical shape of the curves indicated a strong interaction between the variables. Therefore, it is concluded that the generated model showed reasonable predictability and sufficient accuracy for the productivity of NDSS in the experimental conditions used.

## 2.6 Conclusions

Two solar stills, NDSS and CSS were constructed for comparing the performance of the stills productivity in sunny and cloudy days. The effect of mass water and mass of PCM were also investigated on the total productivity of stills. The experiments were conducted in typical days under weather conditions of Varanasi, India. The concluded results are presented as follows:

Using integrated condenser will reduce the glass temperature. The saline water in basin1 and basin 2 will also be heated and evaporated to increase the productivity of the solar still because the temperature of integrated condenser is lower than evaporator. The shielded condenser will keep the outer wall of condenser cool. To get benefit from circulation of the water vapor inside integrated condenser, the stepped basin 2 is housed in the condenser from left and right side and free from other side to let the mixture to circulate.



**Fig. 2.23** Predicted versus experimental productivity of NDSS

- The daily productivity of NDSS is slightly higher than the CSS in all days  $5.71 \text{ kg/m}^2 \text{ day}$  for NDSS and  $4.295 \text{ kg/m}^2 \text{ day}$  for CSS. Thus, NDSS is favored for sunny and partially cloudy days due to the higher productivity.
- The productivity during night time of NDSS is mostly higher than the CSS in all days. This increase is due to the increase of the mass of PCM. Thus it is necessary to consider using it in solar still system. Furthermore, the transferred heat from the PCM to the saline water during discharge process is enough to produce high amount of distilled water because of decreasing in operating temperature which is compared with low ambient temperature at night condition.

## Appendix

### Acronyms

<i>SSWLA</i>	Solar still with lauric acid
<i>SSWMA</i>	Solar still with myristic acid
<i>LA</i>	Lauric Acid
<i>MA</i>	Myristic Acid
<i>UNEP</i>	United Nations Environment Programme
<i>WHO</i>	World Health Organization
<i>WMO</i>	World Meteorological Organization

### Nomenclature

<i>A</i>	Area, $\text{m}^2$
<i>C<sub>p</sub></i>	Specific heat, $\text{J/kg } ^\circ\text{K}$
<i>dt</i>	Time interval, s

$Gr$	Grashof number
$h$	Heat transfer coefficient, $W/m^2 \text{ } ^\circ K$
$h_{cw}$	Convective heat transfer coefficient from water surface to the glass cover, $W/m^2 \text{ } ^\circ K$
$h_{ew}$	Evaporative heat transfer coefficient from water surface to the glass cover, $W/m^2 \text{ } ^\circ K$
$h_{rw}$	Radiative heat transfer coefficient from water surface to the glass cover, $W/m^2 \text{ } ^\circ K$
$h_1$	Total heat transfer coefficient from water to glass cover, $W/m^2 \text{ } ^\circ K$
$h_2$	Convective heat transfer coefficient from glass to ambient, $W/m^2 \text{ } ^\circ K$
$h_3$	Convective heat transfer coefficient from basin liner to water, $W/m^2 \text{ } ^\circ K$
$i$	Current, ampere
$I(t)$	Solar flux on an inclined collector, $W/m^2$
$m$	Mass, kg
$mew$	Productivity, $ml/m^2 hr$
$P$	Partial pressure, $N/m^2$
$Pr$	Prandtl number
$q$	Heat transfer, $W/m^2$
$Ra$	Rayleigh number
$T$	Temperature, $^\circ C$
$U_{ga}$	Overall heat transfer coefficient from inner glass to ambient, $W/m^2$
$U_b$	Overall bottom loss coefficient, $W/m^2$
$U_t$	Top loss coefficient, $W/m^2$
$U_L$	Overall heat transfer coefficient, $W/m^2$
$K$	Thermal conductivity, $W/mK$
$L_x$	Thickness, m
$l$	Liquid
$s$	Solid
$v$	Wind speed, m/s

*Greek symbols*

$\alpha$	Absorptivity
$\alpha'$	Fraction of solar energy absorbed
$\varepsilon$	Emissivity
$\delta$	Incremental rise, $^\circ K$
$\sigma$	Stefan-Boltzmann constant, $W/m^2 \text{ } ^\circ K^4$
$\tau$	Transmittance coefficient
$\eta_i$	Instantaneous efficiency (%)
$\Delta T$	Effective temperature difference ( $^\circ K$ )
$\gamma$	Latent heat of vaporization $J/kg$

*Subscripts*

$a$	Ambient
$b$	Basin
$c$	Convective



<i>cond</i>	Conduction
<i>e</i>	Evaporative
<i>eff</i>	Effective
<i>f</i>	Film temperature
<i>g</i>	Glass
<i>in</i>	Inner
<i>ins</i>	Insulation
<i>mt</i>	Melting
<i>o</i>	Outer
<i>pcm</i>	Phase change material
<i>r</i>	Radiation
<i>w</i>	Water

## References

1. Shukla SK, Sorayan VPS, Gupta SK (2004) Parametric studies of passive/active solar stills by using modified convective mass transfer relations. *Int J Ambient Energy* 25(3):212–232
2. Shukla SK (2003) Computer modeling of passive solar still by evaluating absorptivity of the basin liner. *Int J Ambient Energy* 24:123–132
3. Malik MAS, Tiwari GN, Kumar A, Sodha MS (eds) (1982) *Solar distillation a practical study of a wide rang of stills and their optimum design, construction and performance*. Pergamon Press, New York
4. Fath H (1998) Solar distillation: a promising alternative for water provision with free energy, a simple technology and a clean environment. *Desalination* 116:45–56
5. Dunkle RV (1961) Solar water distillation: the roof of type still and multiple effect diffusion still, international developments in heat transfer. In: A.S.M.E, *Proceeding of International Heat Transfer, part V*, University of Colorado, p 895
6. Cooper PI (1973) Digital simulation of experimental solar still data. *Sol Energy* 14:451
7. Abdallaha S, Abu-Khader MM, Badran B (2009) Effect of various absorbing materials on the thermal performance of solar stills. *Desalination* 242:128–137
8. Abdulhaiy MR (2005) Transient performance of a stepped solar still with built-in latent heat thermal energy storage. *Desalination* 171(1):61–76
9. Agyenim F, Eames P, Smyth M (2009) A comparison of heat transfer enhancement in a medium temperature thermal energy storage heat exchanger using fins. *Sol Energy* 83(9):1509–1520
10. El-Bahi A, Inan D (1999) Analysis of a parallel double glass solar still with separate condenser. *Renew Energy* 17(4):509–521
11. Al-Hamadani AAF, Shukla SK (2011) Water distillation using solar energy system with lauric acid as storage medium. *Int J Energy Eng* 1(1):1–8
12. Aybar HŞ, Egelioglu F, Atikol U (2005) An experimental study on an inclined solar water distillation system. *Desalination* 180(1–3):285–289
13. Dwivedi VK, Tiwari GN (2009) Comparison of internal heat transfer coefficients in passive solar stills by different thermal models: an experimental validation. *Desalination* 246(1–3):304–318
14. El-Sebaili AA, Al-Ghamdi AA, Al-Hazmi FS, Faidah AS (2009) Thermal performance of a single basin solar still with PCM as a storage medium. *Appl Energy* 86(7–8):1187–1195

15. El-Swify ME, Metias MZ (2002) Performance of double exposure solar still. *Renew Energy* 26(4):531–547
16. Fath HES, Elsherbiny SM (1993) Effect of adding a passive condenser on solar still performance. *Energy Convers Manage* 34(1):63–72
17. Fath HES, Hosny HM (2002) Thermal performance of a single-sloped basin still with an inherent built-in additional condenser. *Desalination* 142(1):19–27
18. Garg HP, Mann HS (1976) Effect of climatic, operational, and design parameters on the year round performance of single-sloped and double-sloped solar still under Indian arid zone conditions. *Sol Energy* (United States) 18(2):159–164 (Medium: X)
19. Kalogirou S (2009) *Solar energy engineering: processes and systems*, 1st edn. Academic Press is an imprint of Elsevier, USA
20. Lacroix M, Benmadda M (1997) Numerical simulation of natural convection-dominated melting and solidification from a finned vertical wall. *Numer Heat Transf, Part A: Appl* 31(1):71–86
21. Madhlopa A (2009) Development of an advanced passive solar still with separate condenser Ph.D. Strathclyde, Glasgow
22. Madhlopa A, Johnstone C (2009) Numerical study of a passive solar still with separate condenser. *Renew Energy* 34(7):1668–1677
23. Maroo SC, Goswami DY (2009) Theoretical analysis of a single-stage and two-stage solar driven flash desalination system based on passive vacuum generation. *Desalination* 249(2):635–646
24. Naim MM, El-Kawi MAA (2002) Non-conventional solar stills Part 2. Non-conventional solar stills with energy storage element. *Desalination* 153(1–3):71–80
25. Noordin MY, Venkatesh VC, Sharif S, Elting S, Abdullah A (2004) Application of response surface methodology in describing the performance of coated carbide tools when turning AISI 1045 steel. *J Mater Process Technol* 145:46–58
26. Tabrizi FF, Dashtban M, Moghaddam H (2010) Experimental investigation of a weir-type cascade solar still with built-in latent heat thermal energy storage system. *Desalination* 260(1–3):248–253
27. Velmurugan V, Gopalakrishnan M, Raghu R, Srithar K (2008) Single basin solar still with fin for enhancing productivity. *Energy Convers Manage* 49(10):2602–2608
28. Sarı A, Kaygusuz K (2002) Thermal and heat transfer characteristics in a latent heat storage system using lauric acid. *Energy Convers Manage* 43(18):2493–2507
29. Montgomery DC (2001) *Design and analysis of experiments*, 5th edn. Wiley, New York
30. Myers RH, Montgomery D (2000) *Response surface methodology: process and product optimization using designed experiments*, 2nd edn. Wiley, USA
31. Nafey AS, Abdelkader M, Abdelmotalip A, Mabrouk AA (2002) Enhancement of solar still productivity using floating perforated black plate. *Energy Convers Manage* 43(7):937–946
32. Sarı A (2003) Thermal reliability test of some fatty acids as PCMs used for solar thermal latent heat storage applications. *Energy Convers Manage* 44(14):2277–2287
33. Shukla SK, Sorayan VPS (2005) Thermal modeling of solar stills: an experimental validation: *renewable Energy. Fuel Energy Abstr* 30(5):683–699

# Chapter 3

## Magneto-Rheological Fluid Technology

Ali A. Alghamdi, Ruben Lostado and Abdul-Ghani Olabi

**Abstract** Increasingly, Magneto-rheological (MR) fluid technology has been successfully employed in various applications across various fields. This technology has received significant attention due to its adaptability in the operation of semi-active control systems requiring small power sources. It can potentially deliver highly reliable mechanical operations, managed by a magnetic field as the external operating power. To summarize current magneto-rheological technology, MR fluid can be described as a controllable material that is included in the group of smart materials that have the unique ability to change yield stress. This property can be used in MR devices to generate and control force. The aim of this chapter is to review recent research into MR fluid technology by describing the important factors affecting MR devices design, such as MR fluid properties, operational modes, magnetic materials, and magnetic circuits.

### 3.1 Introduction

Developments in science and technology have generated a new family of materials call smart materials. These kinds of materials have remarkably improved the design of devices by joining mechanical and electronics effect. Intelligent materials have the ability to change their shape, size, or state from liquid to solid, are

---

A. A. Alghamdi  
School of Mechanical and Manufacturing Engineering, Dublin City University,  
Dublin 9, Ireland

R. Lostado  
Mechanical Engineering Department, University of La Rioja, La Rioja, Spain

A.-G. Olabi (✉)  
School of Engineering, Institute of Engineering and Energy Technologies,  
University of the West of Scotland, Paisley, PA1 2BE, Scotland  
e-mail: Abdul.olabi@uws.ac.uk

**Table 3.1** Main properties of MR fluids [8]

Property	Typical value
Yield stress	50–100 kPa
Magnetic field strength	≈ 250 kA/m
Viscosity	0.1–1.0 Pa·s
Operation temperatures range	–40–150 °C
Response time	Milliseconds
Density	3–4 g/cm <sup>3</sup>
Power supply	2–25 V @ 1–2 A

controlled by external power and can be used in many applications [1]. Their properties can be dramatically altered and controlled by means of external influences such as temperature, and electric or magnetic fields. Individual types of smart material have a property that can be significantly altered, such as conductivity, volume, or viscosity. This property that can be changed influences the type of applications for which the smart material can be used. There are different types of smart materials such as piezoelectric materials (PZT), electro-rheological fluids (ERF), and magneto-rheological fluids (MRF) [2]. This chapter focuses on MR fluid material. MR fluid can be included in two general subjects areas: in the scientific area of conventional Magnetism, and also in the engineering area of Rheology. The relationship between MR fluids and a magnetic field is that the magnetic field can change the rheological properties of an MR Fluid [3]. MR fluids are suspensions of micron-sized magnetizable soft particles suspended in a carrier fluid such as silicon oil, mineral oil, or hydrocarbon oil. The carrier liquid acts as a dispersant medium and ensures the homogeneity of the particles in the fluid. Important characteristics of the magnetic dynamically diffused particles are particle shape, size, density, saturation magnetization, particles suspension and distribution [4, 5]. Therefore, a variety of additives are added to MR fluids such as those that affect the polarization of the particles, and those that stabilize the particles suspension in the MR fluid and stabilize the MR fluid structure by prevented gravitational settling and reduction of friction between particles [6, 7]. These additives are used in order to keep the particles suspension in the fluid stable. Table 3.1 shows the main properties of MR fluids.

A common problem in MR fluids technology applications is the inclination of the active magnetic particles to settle down, which disturbs the homogeneity of the MR fluid and could influence its properties. This is important in MR fluids technology because the magnetic particles can be denser than the liquid, and they can settle under gravity to form a hard “cake”, which makes them difficult to redistribute.

This chapter presents a state-of-the-art review on MR fluid technology, and focuses on the following:

- MR fluid technology, in particular MR fluid modes as well as models.
- Magnetism and magnetic materials, in particular the fundamentals of electromagnetism and types of magnetic materials.
- MR fluid technology, in particular the effect of different additive techniques.

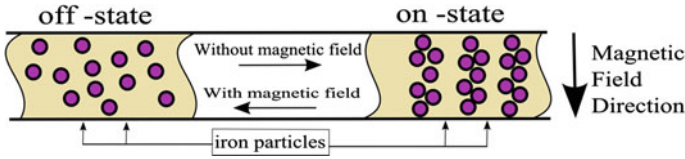


Fig. 3.1 Illustrate off-state and on-state

## 3.2 Magneto-Rheological Fluid

The MR fluid effect was discovered in the 1940s by Jacob Rabinow at the US National Bureau of Standards. In 1948, Rabinow applied MR fluids as a new material to clutch technology [9]. MR fluids contain magnetizable particles, non-magnetizable suspending fluids, and additives. The most popular materials used as particles are carbonyl iron, or iron powder, because they have a high level of saturation magnetization [10]. An MR fluid is a fluid that exhibits rheological behaviour. Rheology is the study of flow and deformation that responds to an applied magnetic field and has the ability to change reversibly from a free flowing linear viscous fluid to a semi-solid state. The behaviour of MR fluid can be described by elastic and plastic deformation. When a magnetic field is not present, an MR fluid behaves like a Newtonian liquid. This state is called off-state as shown in Fig. 3.1. It accepts low viscosity as shown in Fig. 3.2, and is characterized by Newton's law:

$$\tau = \eta \dot{\gamma} \quad (3.1)$$

where  $\tau$  is shear stress (units:  $\text{N/m}^2$ ),  $\eta$  is dynamic viscosity (units:  $\text{Pa} \cdot \text{s}$ ), and  $\dot{\gamma}$  is shear rate (units:  $1/\text{s}$ ). Water and oil examples of Newtonian fluid. The dynamic viscosity in newtonian fluid has a constant value.

When MR fluid behaves like a Newtonian fluid, the most important property is the viscosity as shown by Newton's law. The viscosity changes with temperature. Because of this, temperature would normally be considered as an uncontrollable feature. There are two ways of expressing the viscosity-dynamic and kinematic viscosity. Dynamic viscosity is defined by:

$$\eta = \frac{\tau}{\dot{\gamma}} \quad (3.2)$$

Kinematic viscosity is express as:

$$\nu = \frac{\eta}{\rho} \quad (3.3)$$

where  $\rho$  is density (units:  $\text{kg/m}^3$ )

**Fig. 3.2** Shear stress versus speed for Newton and Bingham model of MRF

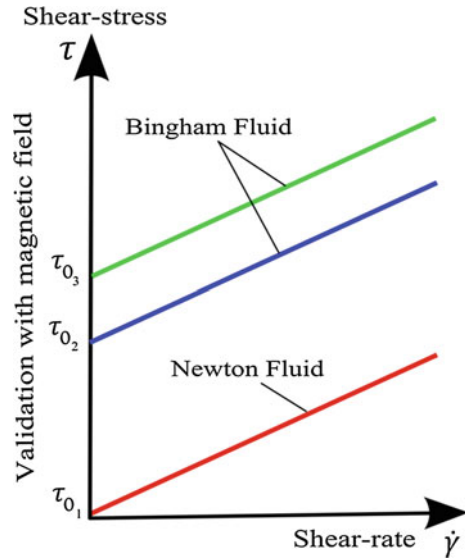


Figure 3.2 represents the relationship between shear stress and shear rate for a Newtonian and a Bingham fluid. The Bingham plastic model is used to describe the rheological characteristics of an MR fluid [8]. The Bingham plastic model can maintain a certain quantity of yield stress without having the corresponding shear rate as shown in Fig. 3.2.

The MR fluids deform with a shear rate until shear stress goes beyond the highest yield stress of MR fluid material [11, 12]. The maximum yield stress can be applied without causing continuous movement of the magnetizable particles. The yield stress can be controlled, increasing, or decreasing through the strength of the magnetic field applied to MR fluid, the fluid under an applied external magnetic field shows a characteristic of the Bingham plastic model [13] and is called on-state as shown in Fig. 3.1. If the stress has reached its maximum value, the chains are going to break and the fluid will flow even if the magnetic field is still in effect on the MR fluid [14]. The maximum value of yield stress can be controlled through the value of the magnetic field applied to the MR fluid [8] as:

$$\tau = \tau_y H + \eta \dot{\gamma} \quad (3.4)$$

where  $\tau_y$  is yield stress at zero shear rates (units:  $\text{N/m}^2$ ),  $H$  is the magnetic field intensity (unit:  $\text{A/m}$ ).

In a magnetic field each metal particle of the MR fluid becomes a dipole and their ferromagnetic properties attract the neighbouring particles to make chains that are called the MR fluid structure [12]. Since they aligned in this manner, the particles are restrained from moving away from their respective “fluxes” lines, and act as a barrier opposing any external force [15].

### 3.2.1 Shear Yield Stress

There are two ways of expressing the yield stress, the dynamic and the static yield stress. The dynamic yield stress of an MR fluid is usually described as the zero-rate intercept determined through a linear regression curve fitted to the flow data as shown in Fig. 3.2. The static yield stress identifies to the shear stress essential to zero flow of shear-rate. The shear yield stress [16] can be calculated from Eq. (3.3):

$$\tau_y = F_y / \pi r^2 \quad (3.5)$$

where,  $\tau_y$  is shear yield stress,  $r$  is radius of the pipe where the MR fluid under magnetic field effect (units: m),  $F_y$  is the press force (units: N).

## 3.3 MR Fluid Modes

MR fluid devices can be operated in several modes depending on the function and the type of deformation employed. Usually they are designed to operate either in Valve mode, Shear mode, or Squeeze mode [8, 17].

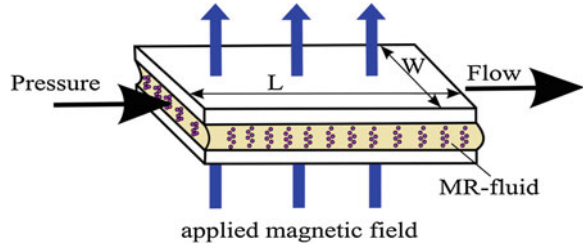
### 3.3.1 Valve Mode

In valve mode, as shown in Fig. 3.3, the MR fluid flow is located between two static plates forming an orifice. The force generating through the MR fluid is created by a pressure drop. The magnetic field, which is applied perpendicular to the direction of the flow, is used to change the viscosity of the MR fluid in order to control the flow. This pressure drop is fundamentally in a device, which resists an output force attacking the device. The pressure drop developed in a valve mode device can be created by two independent viscous components, i.e.:  $\Delta P_\eta$  as the pure rheological component, and  $\Delta P_{mr}$  as the magneto-rheological component that is dependent on the magnetic field [18]. The pressure value drop can be expressed as:

$$\Delta P = \Delta P_\eta + \Delta P_{mr} = \frac{\eta}{\tau^2} \left[ \frac{12}{f^2} \right] \left[ \frac{\Delta P_{mr}}{\Delta P_\eta} \right] Q \cdot mr \quad (3.6)$$

where  $\eta$  (units: Pa-s) is the dynamic viscosity,  $\tau_{mr}$  (units: N/m<sup>2</sup>) is the yield stress variable in response to an applied magnetic field,  $Q$  (units: m<sup>3</sup>-s) is the volumetric flow rate of the MR fluid, while  $L$ ,  $g$ , and  $w$  (units: m) are the length, fluid gap, and the width of the flow orifice as shown in Fig. 3.3 [19, 20].  $f$  (no unit) is an empirical factor that is identified experimentally. This is dependent on the ratio of the pressure drop relating to the magneto-rheological response factor, and the pressure drop relating to the natural viscosity state of the fluid. The value of

**Fig. 3.3** Valve mode with an applied magnetic fluid



constant  $f$  ranges between 2 and 3 depending on the value of the  $\Delta P_{mr}/\Delta P_{\eta}$  ratio for the device. If the  $\Delta P_{mr}/\Delta P_{\eta}$  ratio is equal or less than 1, the value of  $f$  is likely to be 2, and for a  $\Delta P_{mr}/\Delta P_{\eta}$  ratio equal or larger than 100 the value of  $f$  is likely to be 3. An MR fluid device designed to operate in valve mode can be described by Eq. (3.6) [8]. This equation can be considered in terms of volume, and can be rewritten in terms of the minimum volume,  $V$  (unit:  $m^3$ ), of an active fluid:

$$V = L.w.g = \frac{\eta}{\tau^2} \left[ \frac{12}{f^2} \right] \left[ \frac{\Delta P_{mr}}{\Delta P_{\eta}} \right] Q \cdot \Delta P_{mr} \quad (3.7)$$

Equation (3.7) stands for the minimum volume of an activated fluid required to reach the MR effect at a given flow rate  $Q$ , for a given pressure drop. The valve mode as an operational mode is used in many applications, particular in dampers. They have been studied and developed for commercial applications [18].

### 3.3.2 Shear Mode

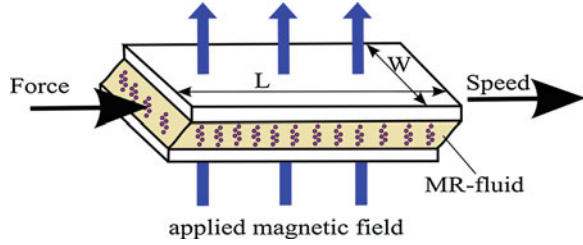
Shear mode can also be referred to as direct shear mode, and is shown in Fig. 3.4. In shear mode the MR fluid flows is located between two surfaces moving in relation to each other, with the magnetic field flowing perpendicularly to the direction of motion of these surfaces. The magnetic field is used to change the viscosity of the MR fluid in order to control the flow and to generating force between the two surfaces [14]. Shear mode can generate force from two independent viscous components,  $\Delta P_{\eta}$  as the pure rheological component, and  $\Delta P_{mr}$  as the magneto-rheological component that is dependent on the magnetic field, similar to valve mode [8]. The total force in shear mode can be expressed as:

$$F = F_{\eta} + F_{mr} = \left[ \frac{\eta.S.A}{g} \right] + \tau_{mr}.A \quad (3.8)$$

where  $S$  is the relative speed of the two surfaces,  $\eta$  is the dynamic viscosity,  $g$  is gap size of the flow channel,  $\tau_{mr}$  is the yield stress developed in response to an applied magnetic field, and  $A$  is the surfaces area of the activated fluid and can be define as shown in Fig. 3.4. The surfaces area can be expressed as:



**Fig. 3.4** Shear mode with an applied magnetic fluid



$$A = L.w \quad (3.9)$$

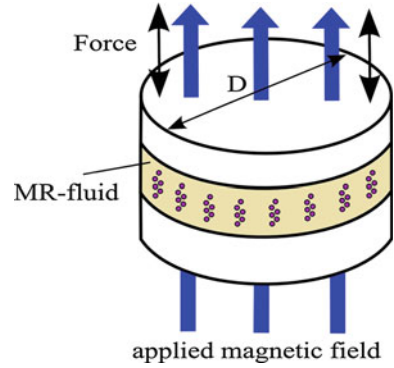
where  $L$  and  $w$ , are the length and width of the flow channel as shown in Fig. 3.4. Equation (3.8) can be used for the design of MRF applications that operate in shear mode. The  $\Delta F_{mr}/\Delta F\eta$  ratio indicates the range of force adjustability inherent in the MR fluid device in shear mode [8]. A larger ratio indicates that the device is capable of considerable changes of force from off-state to on-state, as shown in Fig. 3.1. By rewriting Eq. (3.8), the minimum volume,  $V$ , of activated fluid can be calculated as

$$V = L.w.g = \frac{\eta}{\tau^2} \left[ \frac{F_{mr}}{F_\eta} \right] F_{mr}.S \quad (3.10)$$

### 3.3.3 Squeeze Mode

Squeeze mode is shown in Fig. 3.5, and operates when a force is applied to the plates in the same direction as the magnetic field to reduce or expand the distance between the parallel plates, causing a squeeze flow. The magnetic field flowing perpendicularly to the direction of MR fluid motion is used to change the viscosity of the MR fluid in order to control the flow and to generating force. In this mode the acting force is in line with the magnetic flux lines and the particle chains [18, 21]. MR fluid devices in squeeze mode usually have little or no flow of MR fluid depend on the devices structure. In squeeze mode, the force is dependent on the mechanical properties of the carbonyl iron particles chains, rather than the viscosity changes of the MR fluid. This mode has been used in the control of small amplitude vibration isolators' application [22]. Larger displacement applications designed to operate in squeeze mode have yet to be investigated, and there may be some novel device designs that are currently being explored to develop this type of squeeze mode operation. Few squeeze mode studies were found investigating squeeze mode operation that is capable of producing compression and tensile stresses at same time. This would create more force compared to valve and shear operational modes [8, 21].

Fig. 3.5 Squeeze mode



### 3.4 MR Fluid Modeling

One of the challenging aspects in developing devices to achieve high performances is the development of models that can accurately describe their unique characteristics. Models of MR fluids play an important role in the development of MR fluid devices. A wide variety of nonlinear models have been used to characterize MR fluids. Three kinds of models were established for MR devices. These are the Bingham plastic model [17], the bi-viscous model [23], and Herschel-Bulkley model [24] for MR fluid applications.

#### 3.4.1 Bingham-Plastic Model

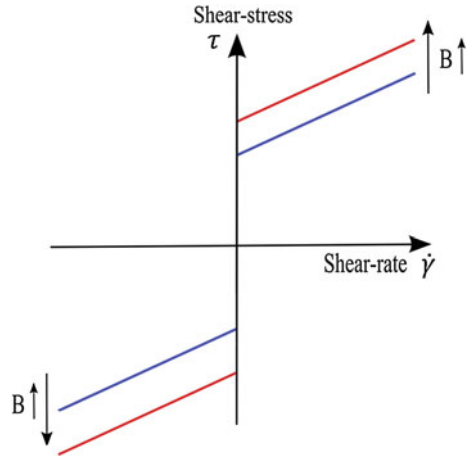
The Bingham plastic model is normally used to describe MR fluid flow behaviour, particularly in damper design [25]. The MR fluid material can absorb a certain level of shear stress without any change in viscous behaviour in the Bingham plastic model. The MR fluid begins to flow when the generating stress level is higher than the yield shear stress [26] as illustrated in Fig. 3.6. The MR fluid flows in the case of a post-yield of viscosity. The yield stress is a function of magnetic field, and can be increased if the magnetic field is increased. The Bingham plastic model of the stress–strain constitutive relationship can be expressed as [17]:

$$\tau = \begin{cases} \tau_y + \eta \dot{\gamma} & \tau > \tau_y \\ 0 & \tau \leq \tau_y \end{cases} \quad (3.11)$$

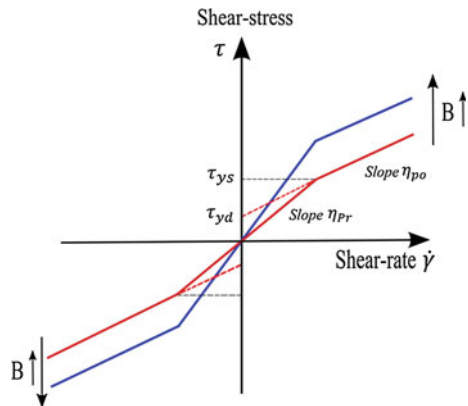
#### 3.4.2 Bi-viscous Model

The Bi-viscous Model has two viscosities. The first viscosity is called the pre-yield viscosity,  $\eta_{pr}$ . The second viscosity is called the post-yield viscosity,  $\eta_{po}$ . Also, there are two yield stresses. The first yield stress is called the dynamic yield stress,

**Fig. 3.6** Bingham plastic model of MR fluid



**Fig. 3.7** Biviscous model of MR fluid



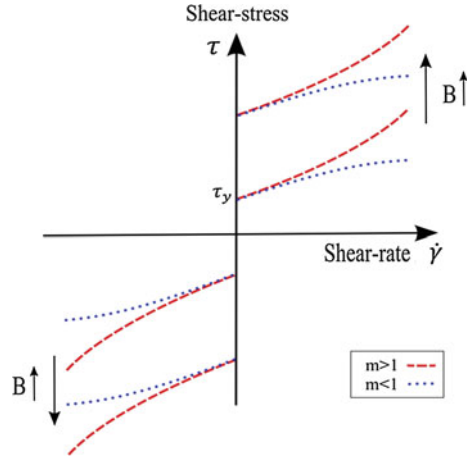
$\tau_{yd}$ . The second yield stress is called static yield stress,  $\tau_{ys}$  [27]. The bi-viscous model is demonstrated in Fig. 3.7. The shear stress equation is expressed as:

$$\tau = \begin{cases} \pm\tau_{yd}(B) + \eta_{po}\dot{\gamma} & \tau > \tau_{ys} \\ \eta_{pr}\dot{\gamma} & \tau \leq \tau_{ys} \end{cases} \quad (3.12)$$

### 3.4.3 Herschel-Bulkley Model

The Herschel-Bulkley model is similar to the Bingham plastic model [28]. However, Herschel-Bulkley’s post yield property of material exists in two states as illustrated in Fig. 3.8. The first state is shear thickening, ( $m > 1$ ), and the second state is shear thinning, ( $m < 1$ ) [29]. Herschel-Bulkley can be expressed as:

**Fig. 3.8** Herschel-Bulkley model of MR fluid



$$\tau = \pm \tau_y(B) + \eta(\eta)^{\frac{1}{m}} \tag{3.13}$$

### 3.5 Types of Magnetic Materials

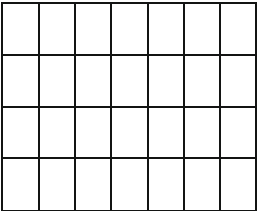
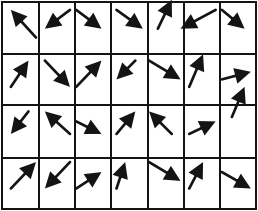
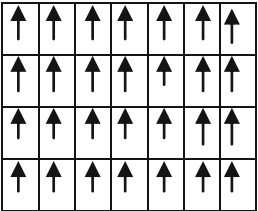
Magnetic materials can be classified as diamagnetic materials, paramagnetic materials, and ferromagnetic materials. The classification criterion of magnetic materials is based on the variation between the external applied magnetic field,  $H$ , and the internal induced magnetic flux density,  $B$  [30]. In general, the magnetic materials' characteristics can be obtained from a graph of the magnetization curve of magnetic flux density  $B$ , versus magnetic field strength  $H$ . This is called the  $BH$  curve, and the slope of the curve is the magnetic permeability of the material.

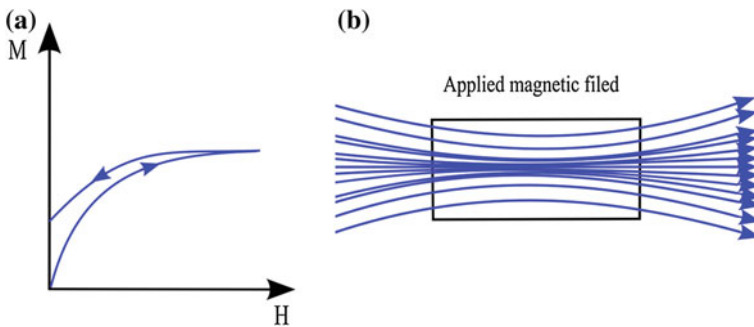
The behaviour of diamagnetic materials, paramagnetic materials, and ferromagnetic materials under the influence of a magnetic field is illustrated in Table 3.2 for each class of magnetic materials, with the magnetic behaviour, magnetic susceptibility, and examples of the materials.

The rheological effect of a magneto-rheological fluid is mainly attributed to the carbonyl iron particles. Iron is a type of ferromagnetic material, so MR fluids properties can be controlled and managed like any other material with ferromagnetic properties [8, 21].

Ferromagnetic materials under the influence of a magnetic field show that this influence is concentrated inside the materials. That creates strong attractions between the ferromagnetic material and the magnetic field, which acts in all magnetic field directions, and the magnetic permeability and the magnetic susceptibility are very large. The magnetization curve of the material reaches saturation when the magnetic field reaches a certain value, as illustrating in Fig. 3.9 showing the behaviour of ferromagnetic materials under a magnetic field.

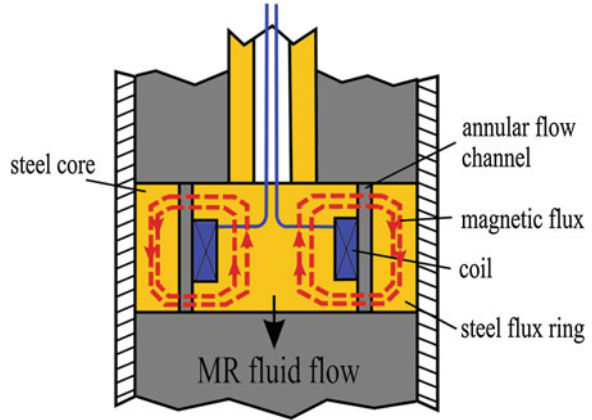
**Table 3.2** Different of magnetic behaviour

Type of magnetism	Magnetic behaviour	Magnetic susceptibility	Examples
Diamagnetic		Small and negative	Copper, silver, gold, and alumina
Paramagnetic		Small and positive	Aluminum, titanium and alloys of copper
Ferromagnetic		Very large and positive, function of applied field, microstructure dependent	Iron, nickel and cobalt



**Fig. 3.9** Ferromagnetic materials behaviour under magnetic field. **a** Magnetization curve for ferromagnetic materials. **b** The schematic diagram about the field difference between the external and internal of the ferromagnetic material under magnetic field

**Fig. 3.10** Magnetic circuit design for MR device

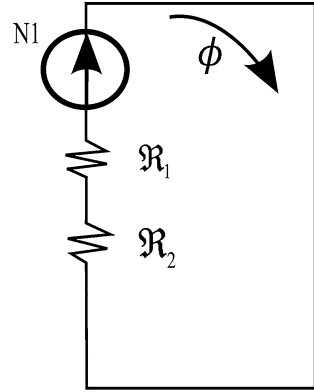


### 3.6 Magnetic Circuit

Magnetic circuit plays an important role in MR device design, in order to generate sufficient force and improve MR device performance. Certain structural characteristics are important, such as the path of the magnetic flux lines in the fluid resistance space, geometry of flow area, selection of the materials in magnetic circuit structure, and effective surface area of the magnetic field in the device [31].

Figure 3.10 shows a typical Magnetic circuit design for MR device. The Magnetic circuit design usually contains a magnetic flux lines board, a magnetic core, an electromagnetic coil, and an MR fluid flow channel located between the outside of the magnetic core and the inside of the magnetic board [32]. The magnetic circuit design guides and focuses the magnetic flux lines through the active surface magnetic area of the MR device to control the viscosity of MR fluid, and in turn, the dynamic yield stress of the MR Fluid is changed by the magnetic field intensity generated by the electromagnetic coil. As a result, the force of MR device is activated and aligned with the MR fluid motion, which is under the control of the magnetic field intensity value. Without the magnetic field, the MR device force is solely due to the viscosity of the MR Fluid itself [33]. Good magnetic circuit design has a small cross section of steel to keep the magnetic field in the steel very low, and also minimizes the total amount of steel in the magnetic circuit design and in the MR device. It is usually composed of low carbon steel, as it has a high magnetic permeability and saturation boundary. This gives an opportunity to guide magnetic flux line into the MR device's magnetic area.

**Fig. 3.11** Reluctances in the magnetic circuit



### 3.6.1 Electromagnetic Circuit Design Calculation

The mathematical calculations of the magnetic flux lines  $\phi$  in a magnetic circuit is based on the Eq. (3.14) and is similarity to Ohm's law [34]:

$$\phi = \frac{F}{\sum \mathfrak{R}} \quad (3.14)$$

where  $F$  is a magnetic movement force in the circuit, and  $\sum \mathfrak{R}$  is the summation of the magnetic reluctance of each material used as part of magnetic circuit. The electromagnetic circuit is usually designed with a low reluctance flux conduit by using steel to guide and to focus the magnetic flux density into the MR Fluid of the cylinder of the MR device. Therefore, any magnetic circuit design for an MR device has two reluctances as shown in Fig. 3.11, where  $\mathfrak{R}_1$  is the reluctance of the steel and  $\mathfrak{R}_2$  is the reluctance of the MR fluid.

Equation (3.15) can be used to calculate the reluctance in each material utilized in the magnetic circuit:

$$\mathfrak{R} = \frac{l}{\mu A} \quad (3.15)$$

where  $l$  is the length of magnetic path for each material used in the magnetic circuit,  $A$  is the cross-sectional area of the flux path for each material used in the magnetic circuit, and  $\mu$  is the permeability of each material used in the magnetic circuit.

The permeability  $\mu$ , of each material can be found out through  $BH$  curve and calculated by Eq. (3.16) for each material of the magnetic circuit.

$$\mu = \frac{B}{H} \quad (3.16)$$

The saturation point is shown in the  $BH$  curve for each material of the MR damper design.

### 3.7 Magneto-Rheological Fluid Properties

The most important properties of MR fluids are the particle density and size. These play an important role in the capability of the MR fluid. Jolly and Carlson [19] studied the magnetic properties of MR fluids, and their theoretical model results indicate that these fluctuates significantly depending on the particle density loading in the MR Fluid. The study examined three different samples with iron loadings, 10, 20, and 30 % by volume. The magnetic induction increased with field strength until the saturation boundary of the fluid was reached. Also, their results indicated that when the iron particles loading by volume was increased, the inherent induction of the magnetic flux density also increased due to the permeability being high due to the high volume [35]. The intrinsic induction or the polarization density of an MR fluid at complete saturation was equal to  $\varphi \times J_s$  Tesla, where  $\varphi$  is the percentage volume of the particles in the fluid (unitless), and  $J_s$  is the saturation polarization of the particles' material (Tesla). For instance, with an MR fluid containing 40 % iron and  $J_s$  equal 2.1 Tesla, it is expected to become saturated at about  $0.40 \times 2.1 = 0.84$  Tesla.

Chiriac and Stoian [36] identified a relationship between particle size and yield stress. This relationship could explain the greater magneto-static interaction between larger magnetic particles. Increasing the particles' dimension contributes towards improving the MR fluid yield stress in a magnetic field [36]. The conclusion of that study was that greater yield stress corresponds to bigger particles size.

MR Fluids usual utilize micrometer scale carbonyl iron particles as solids, with contents by weight of up to (80 %), with a high yield stress domain range of (50–100 kPa). The higher the solids contents by weight, the higher the yield stress, which can be attributed to the higher magnetic field density at saturation.

However, the problem of high density and large particles size becomes apparent in the absence of a magnetic field or without constant motion of the particles in MR fluid. Over time suspended micro-particles slowly sediment, due to gravitational forces acting on the particles that have a higher density compared to the carrier fluid, and have a larger particle size. This problem leads to the particles suspension becoming no longer effective because the particles collect in lumps on the bottom of the device container. High density and large particles size leads to the difficulty of keeping the particles in a state of suspension.

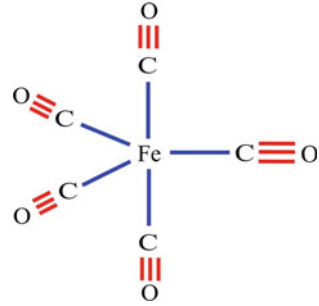
### 3.8 Additives

Additives are used as a solution for alleviating sedimentation of particles and keeping particles in suspension. There are two additive techniques using synthesized particles and syntheoil carrier fluid. The synthesized particle method involves mixing micro-particles with nano-particles in the MR fluid [37].

Nano-particles in MR fluid result in thermodynamic forces which tend to neutralize gravitational forces, thereby delaying the onset of the sedimentation



**Fig. 3.12** Construction of pentacarbonyl



problem. Sedimentation rate,  $R$ , can be determined by placing MR fluid in a vertical cylindrical container at room temperature [37]:

$$R = \Delta h/h \quad (3.17)$$

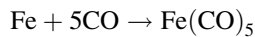
where  $\Delta h$  is the length of the turbid fluid and  $h$  is the whole length of the fluid.

### 3.8.1 Synthesized Particles Additives

Carbonyl iron nano-particles additive is synthesized and then added to the MR fluid to improve sedimentation stability. There are different synthesized particles suspensions in MR fluid, such as pentacarbonyl [38] mixed with kerosene [37].

#### 3.8.1.1 Pentacarbonyl Compound

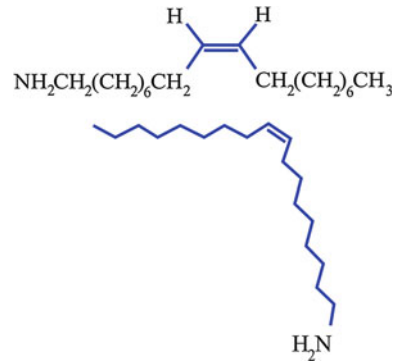
Pentacarbonyl is a compound with the formula  $\text{Fe}(\text{CO})_5$  and its formulation is shown in Fig. 3.12. This compound is synthesized from iron metal heated in the presence of carbon monoxide (CO), at a pressure of greater than 50 bars, and at temperatures in the range of 100–200 °C. The highly volatile organo-metallic compound, iron pentacarbonyl,  $\text{Fe}(\text{CO})_5$  is formed by the reaction:



#### 3.8.1.2 Oleylamine

The oleylamine is an amine of the fatty acid, oleic acid, and its formulation is shown in Fig. 3.13. It is a reagent in the chemical synthesis of nano-particles, and functions as coordinating agent to stabilize the surface of the particles. Most commercially available oleylamine is technical grade, with only 70 % clarity. The oleylamine is formed by the reaction:

**Fig. 3.13** Construction of oleylamine



### 3.8.1.3 Synthesis of Carbonyl Iron Nano-Particles

Carbonyl iron can be synthesized for use as magnetic nano-sized particle additives, as follows. Oleylamine (Tokyo Chemical, Japan) and kerosine (Yakuri Pure Chemical, Japan) are mixed in a 3-neck flask and heated about  $150\text{ }^\circ\text{C}$ . Then pentacarbonyl (Aldrich, USA) is added to the flask while stirring vigorously. After 3 h, the mixture is cooled to  $20\text{ }^\circ\text{C}$ . The magnetic particles are precipitated by adding ethanol to the flask. The supernatant is removed by centrifugation. For dispersion stability, the mixture is resuspended in hexane with oleylamine. After resuspension in hexane, the mixture is dried in a vacuum oven [37, 39].

### 3.8.1.4 Synthesized Effecting

Ngatu and Wereley [40], studied the MR fluid density effect on sedimentation and yield stress by mixing micro-particles and nano-particles. The results indicated that mixing reduces sedimentation and improved particles suspensions in MR fluids with a small reduction in yield stress. The study examined four different samples with solids contents loadings by weight of 50, 60, 70, and 80 %. Table 3.3 shows the approximate sedimentation velocities for all the MR fluid samples. For instance, for 70 wt% particles load in MR fluid, the sedimentation rate dropped from 0.07 to 0.006 m/s with partial substitution of 10 wt% of the total micro-particles with nano-particles. Table 3.4 shows the yield stress for the MR fluids. There was a significant decrease in yield stress as wt% of the nano-particles increased. Therefore, results from the two tests show that sedimentation can be mitigated with a relatively small yield stress penalty by substituting 10 wt% of the total micro-particles with nano-particles at higher iron particles loading of 70 and 80 wt%.

**Table 3.3** Sedimentation velocities for all the MR fluids samples

Synthesized particles size composition	Iron particles loading by %wt			
	50	60	70	80
	Sedimentation rate ( $\mu$ m/s)			
100 % micron particles size	10	8	0.07	0.03
90 % micron mixing with 10 % nano particles size	0.02	0.01	0.007	0.003
85 % micron mixing with 15 % nano particles size	0.006	0.005	0.006	–

**Table 3.4** The yield stress for all the MR fluids samples

Synthesized particles size composition	Iron particles loading by %wt			
	50	60	70	80
	Yield stress (KPa)			
100 % micron particles size	15	20	35	55
90 % micron mixing with 10 % nano particles size	11	17	26	47
85 % micron mixing with 15 % nano particles size	5	13	24	–

### 3.8.2 Synthetic Oil

Some researchers have tested various additive techniques. Materials were added to the formulation of the MR carrier fluid, in order to improve the stability of MR fluid, and reduce sedimentation of particles. There are three types of carrier fluids usually used in MR fluid composition, silicone oil (Oks 1,050), synthetic oil (Oks 352), and mineral oil (Oks 600), and these are mixed with carbonyl iron particles. In order to reduce sedimentation, Arsil 1,100, Arabic gum, and Aerosil 200 and 972 are materials added to stabilize the MR fluid.

#### 3.8.2.1 Synthetic Oil Effect

In a study carried out by Turczyn and Kciuk [41] of the sedimentation of the carbonyl iron particles, additional components were added to the fluid of between 1 and 2 % of the carbonyl iron particles content. Depending on the carrier medium, different stabilizers, such as Aerosil 200 and 972 (Degussa AG) and Arsil 1,100 (Rudniki S.A.), were used for silicone and mineral oil, and Arabic gum (Sigma-Aldrich GmbH) was chosen for synthetic oil suspensions.

These additives have a most important effect on the magneto-rheological properties of the MR fluid. They stabilise the MR fluid, and create a more homogeneous suspension of iron particles in the carrier fluid. This improves the MR fluid technology as it results in faster response times to changes in magnetic field at lower magnetic field value [6].

Turczyn and Kciuk's study showed that the properties of such substances are very important. In case of Arsil 1,100 the sedimentation rate was higher, which can

be explained by the state of the silica surface and its modification [42] and it can be considered as ordinary stabilizer. However the Aerosil 972 showed a lower sedimentation rate, decreasing below 60 % after 35 h in the case of low viscosity mineral oil OKS 600 (only 7.3 mPa·s) containing 20 % of iron carbonyl particles [41].

### ***3.8.3 Others Technique***

Finally, various researchers have introduced other techniques, such as polymer coating the magnetic particles, in order to overcome sedimentation problems. Coating of polymeric shells onto the surface of iron carbonyl particles to reduce the density mismatch has been reported to be a complex process in practice because of various factors affecting the coating thickness such as temperature, molar ratio among reactive agents, and reaction time [43, 44]. For this reason, introduction of additives into iron carbonyl suspensions has been reported as the more usual method employed as it can prevent the physical contact of the magnetic particles, thus preventing hard baking [45].

## **3.9 Conclusion**

MR fluid technology has been widely investigated, is employed in many applications, and has rapidly entered the marketplace. This chapter has clarified the reasons for the extensive use of MR fluid material technology. The studies discussed above evaluated the parameters at the centre of MR fluid technology such as, MR fluid properties, operational modes, magnetic circuit, and additives. These parameters can be summarised as follows:

Knowledge of MR fluids can be gained through understanding MR fluid modes and models. These modes and models describe motion in a fluid that changes viscosity, and can provide solutions based on MR fluid technology. This knowledge can be used to design and develop MR device functionality.

The fundamentals of electromagnetism and types of magnetic materials should be understood, leading to electromagnetic coils that can generate a magnetic field of sufficient strength so that the saturation boundary of the MR fluid is reached. These coils can also control levels of yield stress by generating different values of magnetic field strength.

MR fluid properties of a material is dependent on particles size and density, and these parameters are used to increase yield stress by increasing the magneto-static force between the particles and the permeability of iron particles in the MR fluid

There are two additive techniques, these are synthesized particles and syntheoil carrier fluid. They have been investigated and are used to improve MR fluid properties and to reduce the particle sedimentation effect.

## References

1. Flatau AB, Chong KP (2002) Dynamic smart material and structural system. *Eng Struct* 24:261–270
2. Kciuk M, Turczyn R (2006) Properties and application of magnetorheological fluids. *J Achiev Mater Manuf Eng* 18:127–130
3. Rinaldi C, Chaves A, Elborai S, He X, Zahn M (2005) Magnetic fluid rheology and flows. *Curr Opin Colloid Interface Sci* 10:141–157
4. Aslam M, Xiong-liang Y, Zhong-Chao D (2006) Review of magnetorheological (MR) fluids and its applications in vibration control. *J Mar Sci Appl* 5:17–29
5. Goldsmith K (1950) Note on the rheological properties of elasto-plastic materials. *Br J Appl Phys* 1:107–109
6. Fang C, Zhao BY, Chen LS, Wu Q, Liu N, Hu KA (2005) The effect of the green additive guar gum on the properties of magnetorheological fluid. *Smart Mater Struct* 14:N1–N5
7. Nam TH, AHN KK (2009) New approach to designing an MR brake using a small steel roller and MR fluid. *J Mech Sci Technol* 23:1911–1923
8. Olabi AG, Grunwald A (2007) Design and application of magneto-rheological fluid. *Mater Des* 28:2658–2664
9. Rabiow J (1948) The magnetic fluid clutch. *AIEE Trans* 67:1308–1315
10. Kciuk M, Kciuk S, Turczyn R (2009) Magnetorheological characterisation of carbonyl iron based suspension. *J Achiev Mater Manuf Eng* 33:135–141
11. Flatau AB, Chong KP (2002) Dynamic smart material and structural systems. *Eng Struct* 24:261–270
12. Ginder MJ, Ginder, Davis CL (1994) Shear stresses in magnetorheological fluid: role of magnetic saturation. *Appl Phys Lett* 65:3410–3412
13. Spencer BF Jr, Dyke SJ, Sain MK, Carlson JD (1997) Phenomenological model of a magnetorheological damper. *J Eng Mech* 123:138–230
14. Grunwald A, Olabi AG (2008) Design of magneto-rheological (MR) valve. *Sens Actuators A* 148:211–223
15. Hagenbuchle M, Liu J (1997) Chain formation and chain dynamics in a dilute magnetorheological fluid. *Appl Opt* 36:7664–7671
16. Premalatha SE, Chokkalingam R, Mahendran M (2012) Magneto mechanical properties of iron based MR fluid. *Am J Polym Sci* 2:50–55
17. Carlson JD, Jolly MR (2000) M R fluid, foam and elastomer devices. *Mechatronics* 10:555–569
18. Jolly MR, Bender JW, Carlson JD (2013) Properties and application of commercial magnetorheological fluids. Thomas Lord Research Centre, Lord Corporation, 110 Lord Drive Cary, NC 27511. <http://www.coe.montana.edu/me/faculty/jenkins/Smart%20Structures/prop%20MRF.pdf>, 06/03/2013
19. Jolly MR, Carlson JD, Muñoz BC (1996) A model of the behaviour of magnetorheological materials. *Smart Mater Struct* 5:607–614
20. Zhu X, Jing X, Cheng L (2012) Magnetorheological fluid dampers: a review on structure design and analysis. *J Intell Mater Syst Struct* 23:839–873
21. Mazlan SA, Ekreem NB, Olabi AG (2007) The performance of magnetorheological fluid in squeeze mode. *Smart Mater Struct* 16:1678–1682
22. Kim K, Lee C, Koo J (2008) Design and modelling of semi-active squeeze film dampers using magneto-rheological fluids. *Smart Mater Struct* 17:1–12
23. Stanway R, Sproston JL, El-Wahed AK (1996) Applications of electro-rheological fluids in vibration control: a survey. *Smart Mater Struct* 5:464–482
24. Choi YT, Cho JU, Choi SB, Wereley NM (2005) Constitutive models of electrorheological and magnetorheological fluids using viscometer. *Smart Mater Struct* 14:1025–1036
25. Yang G, Spencer BF Jr, Carlson JD, Sain MK (2002) Large-scale MR fluid dampers: modelling and dynamic performance considerations. *Eng Struct* 24:309–323

26. Barnes HA (1999) The yield stress—a review or ‘panta roi’—everything flows? *J Nonnewton Fluid Mech* 81:133–178
27. Guo S, Yang S, Pan C (2006) Dynamic modeling of magnetorheological damper behaviors. *J Intell Mater Syst Struct* 17:3–14
28. Beaulne M, Mitsoulis E (1997) Creeping motion of a sphere in tubes filled with Herschel–Bulkley fluids. *J Nonnewton Fluid Mech* 72:55–71
29. Farjoud A, Cavey R, Ahmadian M (2009) Craft M. Magneto-rheological fluid behavior in squeeze mode. *Smart Mater Struct* 18:1–7
30. Spaldin N (2003) *Magnetic materials: fundamentals and device applications*. University Press, Cambridge
31. Ginder MJ, Davis CL (1994) Shear stresses in magnetorheological fluid: role of magnetic saturation. *Appl Phys Lett* 65:3410–3412
32. Carlson JD (2002) What makes a good MR fluid. *J Intell Mater Syst Struct* 13:431–435
33. Boese H, Ehrlich J (2010) Performance of magnetorheological fluids in a novel damper with excellent fail-safe behaviour. *J Intell Mater Syst Struct* 21:1537–1542
34. Kraus JD (1991). *Electromagnetics*. McGraw-Hill, Singapore
35. Gökürk HS, Fiske TJ, Kalyon DM (1993) Electric and magnetic properties of a thermoplastic elastomer incorporated with ferromagnetic powders. *IEEE Trans Magn* 29:4170–4176
36. Chiriac H, Stoian G (2010) Influence of particle size distributions on magnetorheological fluid performances. *J Phys Conf Ser* 200:1–4
37. Song KH, Park BJ, Choi HJ (2009) Effect of magnetic nanoparticle additive on characteristics of magnetorheological fluid. *IEEE Trans Magn* 45:4045–4048
38. Japka JE (1988) Microstructure and properties of carbonyl iron powder. *JOM* 40:18–21
39. Yang H, Hasegawa D, Sato OT, Takahashi M, Ogawa T (2008) Gram-scale synthesis of monodisperse Fe nanoparticle in one pot. *Scripta Mater* 58:822–825
40. Ngatu GT, Wereley NM (2007) Viscometric and sedimentation characterization of bidisperse magnetorheological fluids. *IEEE Trans Magn* 43:2474–2476
41. Turczyn R, Kciuk M (2008) preparation and study of model magnetorheological fluids. *J Achiev Mater Manuf Eng* 27:131–134
42. Rager A, Krysztafkiewicz A (1997) Effect of electrolytes and surfactants on physicochemical properties of hydrated silicas. *Colloids Surf A* 125:121–130
43. Jang IB, Kim HB, Lee JY, You JL, Choi HJ, Jhon MS (2005) Role of organic coating on carbonyl iron suspended particles in magnetorheological fluids. *J Appl Phys* 97:10Q912
44. Wu WP, Zhao BY, Wu Q, Chen LS, Hu KA (2006) The strengthening effect of guar gum on the yield stress of magnetorheological fluid. *Smart Mater Struct* 15:N94–N98
45. Lim ST, Choi HJ, Jhon MS (2005) Magnetorheological characterization of carbonyl iron-organoclay suspensions. *IEEE Trans Magn* 41:3745–3747

# Chapter 4

## Tribological Behaviour of Rare-Earth Lubricating Oils

Tianxia Liu, Xianguo Hu, Enzhu Hu and Yufu Xu

**Abstract** Exhaust gas recirculation (EGR) is being used widely to reduce and control the NO<sub>x</sub> emissions from internal combustion engines. However, the use of EGR leads to rise in soot emission and contamination of lubricating oil. Soot-contaminated lubricating oil has been shown to accelerate engine wear. At the same time, zinc dialkyldithiophosphate (ZDDP) which is widely used in lubricating oil as highly active extreme pressure and anti-wear additive may harmfully affect the internal combustion engine and its exhaust gas purifying installation. It is an important for the engine lubricating oil to look for new type lubricating oil additives without or low phosphorus. Rare-earth is the general name of 17 elements, including scandium, yttrium and lanthanide series. It is potential for the rare-earth compounds as new type and efficient lubricating oil additives because most of rare-earth compounds have hexagonal crystal layer structure. Current research results showed that the application prospects of some rare-earth compounds in the tribology field were extremely vast, such as oil, grease, water and coating lubrication systems etc. This chapter will mainly focus on the recent developments on the tribological behaviour of lubricating oils containing rare-earth elements, including the preparation and lubricating mechanism of rare-earth nanoparticles, surface modification for the rare-earth inorganic compound, rare-earth additives with potential application, synergistic effect of the rare-earth compounds and other additives, and the research tendency on the rare-earth lubricating additives, especially the influence of rare-earth elements on the tribological performance of carbon films and its friction-reduced catalyst function etc. which will be helpful to understand and develop the lubrication role of the rare-earth elements as additives in the lubricating oils.

---

T. Liu · X. Hu (✉) · E. Hu · Y. Xu  
Institute of Tribology, Hefei University of Technology, Hefei 230009, China  
e-mail: xghu@hfut.edu.cn

T. Liu  
School of Chemistry and Chemical Engineering, Beifang University of Nationalities,  
Yinchuan 750021, China

## 4.1 Introduction

Diesel engine is the most widely utilized chemical energy conversion device owing to its high energy conversion efficiency and fuel economy. But its effect on the environment has become an increasing concern, especially its exhaust emissions such as  $\text{NO}_x$  and particulate material. In order to meet more and more stringent vehicular exhaust emission norms worldwide, several exhaust pretreatment and post-treatment techniques have been employed in modern engines. EGR is being used to reduce and control the  $\text{NO}_x$  emissions from internal combustion engines. However, the use of EGR leads to rise in soot emission and contamination of lubricating oil. Soot-contaminated lubricating oil has been shown to accelerate engine wear. At the same time, the engine and lubricant manufacturers are expecting to lengthen service intervals and oil life in order to reduce lifetime vehicle costs for the customer.

On the other hand, ZDDP that is extensively used in lubricating oil as highly active extreme pressure anti-wear additive may generate harmful effects on the particle trap of diesel engine, because it might produce ash content blocking filtration system. ZDDP also loses its antioxidant and extreme pressure antiwear property due to rapid decomposition above 160 °C working condition [1]. The combination of mixed ZDDP and carbon black (used as soot surrogate) is strongly antagonistic in terms of wear [2]. In addition, the phosphor from ZDDP can poison the catalytic agent in the diesel engine tail gas after treatment system [3]. At the same time, the phosphor atom from ZDDP may form  $\text{Ca}(\text{PO}_4)_2$ ,  $\text{Ca}_5(\text{PO}_4)_3\text{OH}$  and  $\text{Ca}_5(\text{PO}_4, \text{CO}_3)_3\text{OH}$  nanocrystalline particles that might mechanically embed or chemically bond with the turbostratic soot structure in the lubricating oil during the friction process. The trapped diesel soot particles or agglomerates between two rubbing surfaces experiences repeated extreme local temperature and high contact pressure which might be responsible for particles harder under the engine operation conditions. The harder nanocrystalline particle might promote higher wear, abrasive in some cases [4]. Therefore, it is an important development direction for the engine lubricating oil searching for new type lubricating oil additives without phosphorus or low phosphorus, which may substitute for ZDDP.

In periodic table of the elements, rare earth is a total name including lanthanide and scandium, yttrium which are closely related to lanthanide, altogether 17 elements. Lanthanide include 15 elements: lanthanum (La), cerium (Ce), praseodymium (Pr), neodymium (Nd), promethium (Pm), samarium (Sm), europium (Eu), gadolinium (Gd), terbium (Tb), dysprosium (Dy), holmium (Ho), erbium (Er), thulium (Tm), ytterbium (Yb) and lutetium (Lu). There are abundant rare earth resources in China. The largest rare earth reserve is also in China all over the world, which takes up 80 % of the world's rare-earth resources. More and more scholars pay close attention to rare earth compound applied research in the field of lubricating material in recent years. In especial, the researchers in the field of materials science and technology keep a widespread watchful eye on its



mechanism of tribology action and application study. Their research and development possess vital theoretical significance and application value [5, 6].

Rare-earth and their compounds possess a lot of specific property because of 4f electron. Rare-earth element was named as “industry gourmet powder”. It is potential for the rare-earth compounds as new type and efficient lubricating oil additives because most of rare-earth compounds have hexagonal crystal layer structure. The chemical activity of rare-earth element is better, its atomic radius is large and its electronegativity is low as well. Solid solubility is very low on the rubbing surface. Adsorption capacity is quite powerful at crystal boundary. It is easy to enrichment on the friction surface. The Bell Telephone Laboratories and Lewis centre of NASA firstly researched the friction and wear and adhesiveness of rare-earth in USA. The results showed that adhere coefficient of rare-earth was low ( $<0.4$ ), hardness was also low. It was easy to be oxidized and produced rare earth oxide with certain lubricity at high temperature compared with other metal elements with same crystal structure. Thus it is of possibility for the rare-earth as high temperature lubricating additive. In vacuum and air circumstance, rare-earth friction coefficient and wear rate have obvious difference, which are resulted from 4f orbital electron of rare-earth that influences its chemisorption activity [7, 8].

Current research results showed that some application prospects in the tribology field were extremely vast, such as in oil lubrication, grease lubrication, water lubrication and coating lubrication etc. Rare-earth compounds can obviously prolong the using life of oil, enhance the antiwear ability of machine by 2–4 times, and improve load carrying capacity of lubricating grease 10–100 %. Moreover, the synergistic lubrication effect of the rare-earth compounds and other additives is more outstanding. This chapter will mainly focus on the recent developments and future tendency on the tribological behavior of engine lubricating oils containing rare-earth elements.

## 4.2 Preparation of Rare-Earth Nanoparticles

Main preparations of rare-earth nanoparticles include mechanical method, thermo-physical method and chemical method. The most popular preparation method is chemical method. Chemical method is to mix with ion or molecule from liquid phase or gaseous phase of feedback and let the rare earth elements disperse highly in the final product. Thus it may be prepared nanoparticles at lower reaction temperature or under mild chemical environment. There are several common methods, such as sol-gel method, microemulsion method, precipitation method, hydrothermal method, supercritical fluid desiccation, gaseous phase chemical synthesis and solid phase method etc. The preparation of rare-earth nanoparticles hold vital important role. When taking consider into the rare-earth nanoparticles as lubricating oil additives, suitable preparation method should be selected carefully based on the production cost, production condition, yield capacity, quality requirements and so on [9].

### ***4.2.1 Sol-Gel Method***

To prepare nanometer rare earth compound by sol-gel method includes two steps reactions of hydrolysis and polymerization. Basic process includes to select the suitable rare earth inorganic salt or rare earth metal alkoxide which is apt to hydrolyse gradually gelation by hydrolysis and polycondensation, then go through drying, sintering and other treatment processes and finally gain the requested material. Next clipping and controlling the materials microstructure can be completed in quite small size scale by ways of low-temperature chemical methods, which may carry out the uniformity with nano-sized level.

Compared with others, this method can synthesize inorganic material at low temperature, and carry out the destinations of design, control uniformity and particle size of materials in a molecular level, and finally obtain the nanoparticles with high purity. Using the sol-gel method, the fluorescence materials Eu-dope  $\text{La}_2\text{CaB}_{10}\text{O}_{19}$  were prepared by Li [10]. The single crystal phase Eu: $\text{La}_2\text{CaB}_{10}\text{O}_{19}$  were obtained after the precursor were sintered at 700 °C for 12 h. The phosphors with grain sizes about 100 nm were obtained after the precursor were sintered at 800 °C for 12 h. The intensity of emission of phosphors prepared by above sol-gel method was as much as the phosphors with 395 nm prepared by the solid-state reaction. The results also indicated that the sintering temperature had marked influence on the phase of the final powders.

### ***4.2.2 Microemulsion Method***

Microemulsion relatively recognized definition is two kinds of mutual exclusive liquids (for example: water and oil) spontaneous formative isotropy, low viscosity appearance transparent or translucence and thermodynamics stable disperse system [11]. Preparation nanometer rare-earth compound by microemulsion method included the following four steps.

- Firstly, one kind reactant solubilizes in water intranuclear, another reactant mixed with the former in aqueous solution form.
- Secondly, aqueous phase reactant permeated microemulsion interfacial film and entranced water intranuclear, reacted to another reactant.
- Thirdly, the crystal nucleus was forming and growing. The final product particle size was determined by the water nuclear size.
- Finally, after the formation of the nanoparticles, the system is divided into two phases. It can separate the nanoparticles from the microemulsion phase which contains generated particles through centrifuging separation or adding other solvent to the microemulsion system.

This method has several advantages: particle size uniformity, simple laboratory equipment, operation facility etc. Liu et al. [12] based on the investigation of the

influence factors, such as surfactant, co-surfactant and water phase, on the system containing base oil, span-80, cetyl trimethyl ammonium bromide, isobutanol and water phase, the optimum reaction conditions for forming a micro-emulsion system were proposed.  $\text{LaF}_3$  nanoparticles were synthesized in the formed micro-emulsion system. They obtained the fine nanoparticles with uniform diameter of 10–20 nm. Zhang et al. [13] prepared  $\text{LaF}_3$  nanocluster modified by the compound containing nitrogen by using microemulsion method. The particles size distribution is in the range of 10–30 nm.

### ***4.2.3 Precipitation Method***

Precipitation method includes direct precipitation, coprecipitation method, homogeneous precipitation and rare-earth alkoxide hydrolysis.

Direct precipitation prepares nanoparticles via a certain cationic chemical precipitation. Its advantage is easy to prepare high purity rare earth oxide nanoparticles.

Coprecipitation method is to add precipitating agent to multicomponent system solution containing two or more cationic and get precipitation product of ingredient uniform. This method is of slow reaction speed, easy control, more economic and suitable for mass production.

Homogeneous precipitation is a process in a homogeneous phase solution, in which it generates the precipitation ion and then precipitates slowly dense and heavier amorphous precipitation or granular crystal precipitation in the whole solution with the help of appropriate chemical reaction.

Rare-earth alkoxide hydrolysis is a preparation nanoparticles method to use some organic alkoxide characteristics that can dissolve in organic solvents, and as well take place hydrolysis and generate hydroxide or oxide. Because the reaction process only involves aqueous phase, the possibility of impurity is small. At the same time, alkoxide hydrolysis method separate and prepare the nanoparticles from the solution directly. Thus, the particles are of high purity, fine granularity and narrow particle size distribution. Jing et al. [14] prepared the rare earth oxides and hydroxide by using alkoxide hydrolysis method. The grain size of final product is 10–50 nm and its dispersion degree is high.

### ***4.2.4 Hydrothermal Method***

The hydrothermal method is an effective method which needs a specially-made reactor in which the aqueous solution is selected as reaction medium, and heats it to critical temperature and obtains a high pressure environment in the reaction system. The inorganic synthesis can be taken place in the above system and produced the material with nano/microstructure. The advantages of hydrothermal

method are easy to control the reaction condition, using the most cost-effective reaction medium, a typical environmentally friendly solution approach for the efficient and convenient preparation of various inorganic materials with diverse controllable morphologies and structures.

Jia et al. [15] synthesized nano/microstructure lanthanum borates and oleic acid-capped nano/microstructure lanthanum borates (OANLBs) by hydrothermal route. At the same time, the friction and wear properties of OANLBs as additives in poly-alpha-olefin (PAO) were measured for AISI 1045 steel and laser treated AISI 1045 steel. The results showed that the PAO containing OANLBs possessed much better tribological properties for steel/steel and steel/laser treated AISI 1045 steel sliding pairs than pure PAO.

Liu et al. [16] successfully synthesized  $\text{CeF}_3$  nanocrystals with plate-like and perforated morphologies via a facile hydrothermal route. Meanwhile, they also prepared the nanocrystals of  $\text{CeF}_3$ @silica ( $\text{SiO}_2$  doped  $\text{CeF}_3$ ) which could be dispersed in aqueous solution. They investigated the effects of fluoride sources on the morphology and microstructure of the nanocrystals. The results also indicated that the morphology of the rare earth compound nanocrystals could be tuned well by selecting proper fluoride sources.

#### ***4.2.5 Supercritical Fluid Desiccation***

Supercritical fluid is a kind of material state without difference of gas-liquid interface and with gas and liquid properties in which both temperature and pressure are over the critical points. Supercritical fluid has special solubility, easy modulation density, low viscosity and high mass transfer rate. It has a unique advantage and practical value as a solvent and drying medium. Depending on the medium difference, the supercritical fluid desiccation technology can be generally divided into high temperature supercritical fluid organic solvent desiccation, low temperature supercritical  $\text{CO}_2$  desiccation, low temperature supercritical  $\text{CO}_2$  extraction desiccation and so on.

Supercritical fluid desiccation technology can effectively overcome the surface tension effect that may cause the gel particle aggregate. The as-prepared aerogel powder is often constituted by ultrafine particle. Ye et al. [17] prepared several nanometric borates (lanthanum borate; cerium borate; nickel borate and copper borate) with a particle size 10–70 nm by using  $\text{CO}_2$  supercritical fluid desiccation technique. They also tested the conventional tribological behaviors of nanometric borate. The results showed that nanometric borate had little difference with ZDDP in the general tribological tests. But in the test specially designed by researchers, they found that the nanometric borate additives could increase base oil antiwear and load bearing ability obviously, which demonstrated the special tribological behavior of nanometric borate.

### ***4.2.6 Gaseous Phase Chemical Synthesis***

Gas phase chemical synthesis method is to use volatile rare earth compound steam by chemical reaction synthesis nanometer particle. This method usually used the raw materials such as rare-earth chloride, oxygen chloride, hydrocarbon compound, alcohol salt and carbon compound, and so on, which were easy to obtain, high vapour pressure and reactivity. This method is suitable for preparation some nanoparticles of rare earth nitride, carbide and boride, oxide and so on.

### ***4.2.7 Solid Phase Method***

The most common preparation method of rare earth nanometer lubricating grains is mechanical pulverizing among solid phase. It is a method that utilizes medium and material grind and impact to let material pulverize by using a variety of mill, such as ball mill, agitating mill, tower mill, etc. This method is of simple and easy operation, low cost, high yield, and can prepare high melting point metal or alloy. But its energy consumption is large, the distribution of particle size is uneven, the powder is not fine and the impurity content is high. The product performance can be greatly improved after reasonable control surface modification treatment and process conditions, so it is still a major method of preparation nanometer lubricating material. Ye et al. [18] used home-made ball mill machine and prepared some nanoparticle additives through controlling processing conditions. The nanoparticle additives had more excellent friction reduction effects than conventional additives. The antifriction mechanism of nanoparticle additives is attributed to the formation of a transferred film playing a role of rolling micro-bearing.

## **4.3 Lubrication Mechanism of Rare-Earth Nanoparticles**

When the structure of materials in nanometer scale modulation range, it will show special effect, mainly including the quantum size effect, surface effect, small size effect, macroscopic quantum tunnel effect and dielectric limit field effect [19]. Compared with the traditional materials, nanometer material has special physical and chemical properties because of its special structures. At present nanometer materials applied in the lubrication system is a brand new research field. When adding the nanometer lubricating material as additive to the lubricating oil, it has different antiwear and friction reduction effect than the traditional lubricant additives in the tribological system [20].

Gu et al. [21] tested the tribological properties of SAE 40 CD lubricating oil containing Cu and  $\text{La}_2\text{O}_3\text{-Ce}_2\text{O}_3$  nanoparticles by MRS-1 J four-ball tribometer. They also carried out a 300 h contrastive experiment in KDE5000E diesel engine.

The results showed that lubricating oils containing nanoparticles had optimal tribological properties, when  $W(\text{Cu}):W(\text{La}_2\text{O}_3\text{-Ce}_2\text{O}_3) = 2:1$ ,  $W(\text{Cu}^+)(\text{La}_2\text{O}_3\text{-Ce}_2\text{O}_3) = 0.8\%$ . The wear volumes of friction pair using SAE 40 CD lubricated with Cu and  $\text{La}_2\text{O}_3\text{-Ce}_2\text{O}_3$  nanoparticles were decreased by 88.6–100 %, compared with using pure SAE 40 CD lubricating oil. The consumptions of diesel fuel could be reduced by 9.29 %.

The mostly main principles which have been proposed to explain the lubrication mechanism of rare-earth nanoparticles can be summarized in the following sections.

### ***4.3.1 Permeation and Tribochemical Reaction Film Mechanism***

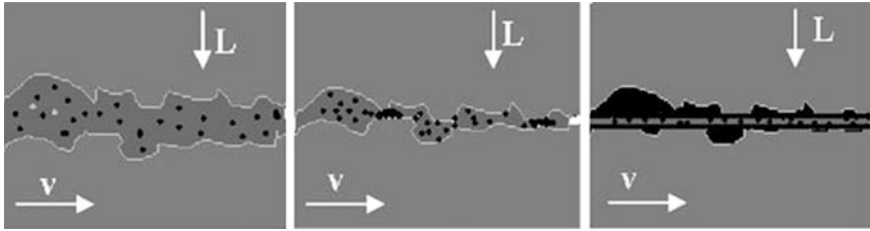
For the traditional organic compound AW/EP additives such as zinc dialkyl (or diaryl) dithiophosphate (ZDDP), tricresyl phosphate (TCP), trixylyl phosphate (TXP) and dilauryl phosphate, usually contain “active” elements such as sulphur, phosphorus or chlorine as well as polar group for strong adsorption. Under mild sliding conditions, they can orientate perpendicular to the rubbing material surface and thus form a film to withstand the local contact pressure. On the other hand, under harsh condition, those additives could react with the rubbing material surface to form so called “reaction film” providing good protection for tribo-pair. In other words, there occurs a continuously “corrosive wear” during these EP additives in which it took action to prevent further severe wear.

### ***4.3.2 Deposited Film Mechanism***

On the lubrication surface, borate additives can move to the surface, deposit on the friction surface and form amorphous state or no qualitative film because of its carrying electric charge. The film plays a role of antiwear and friction reduction.

Li et al. [22] studied the tribological properties of the nanoparticles as lubricant additives and found that the addition of nanoparticles as additives reduced wear and increased load-carrying capacity of base oil remarkably, indicating that the nanoparticles could be used as anti-wear and extreme-pressure additives with excellent performances. The boundary lubrication mechanism of nanoparticles was discussed that probably an extreme thin film was formed by the melting and elongation of nanoparticles under tribological heat and shear force, which was attributed to good tribological performance of nanoparticles as additives.

When the lubricant film between tribo-pairs becomes thinner and mixed lubrication or boundary lubrication occurs, the nanoparticles may carry a proportion of load and separate two rubbing surfaces to prevent their adhesion, thus it



**Fig. 4.1** Schematic representation of the mechanism of nanoparticles as lubricant additive (from Li [22])

is of benefit from the anti-wear properties. When the shearing is strong, the nanoparticles with core/shell structure may be destroyed, the surface capping layer desorbed and decomposed, the inorganic core may be melted and welded on the shearing surface, or adhered on the surface (adherent, laminar material such as lanthanum trifluoride), or reacted with rubbing surfaces to form a protective layer to provide good AW and EP properties. This process is illustrated in Fig. 4.1.

Accordingly, the nanoparticles in lubricants as additives could generate a tribofilm through an in situ deposition mechanism under mixed lubrication and boundary lubrication conditions.

### 4.3.3 “Rolling Ball” Mechanism

“Rolling ball” viewpoint thinks that the nano-lubricant owning good lubrication effect is due to the following three roles:

- The spherical nanoparticles can play a similar “micro bearing” role to improve the lubrication performance.
- Under heavy load and high temperature conditions, the spherical nanoparticles on the rubbing surface are flatten and form sliding system to reduce the friction and wear.
- The ultrafine nano-powder can fill the micro-pits and damaged place on the rubbing surface and play a repairing role.

The antiwear mechanism of hard powder as lubricating oil additives belongs to this idea [23].

### 4.3.4 Recombination Action, Mechanism

There are two or more factors playing a lubrication role. For example, the mechanism of nanometer lanthanum borate enhancing the antiwear property of lubricating oil includes two aspects. On the one hand, the nano-powder generates

physical deposited film on the cores or defects of frictional surfaces. On the other hand, the nanometer lanthanum borate takes place tribochemical reaction under rigor conditions and forms the special chemical reaction film [24].

#### 4.4 Surface Modification for Rare-Earth Inorganic Compound

Some rare-earth inorganic compound nanoparticles have good tribological properties, but they could not be used widely because of their poor dispersion stability in lubricating oil. After the rare-earth nanoparticles by surface modification, its surface activity may be enhanced and its lipophilicity may be also improved, and then its dispersion stability may be increased as well. For example, it was effectively prevented the  $\text{LaF}_3$  agglomerating between nanoparticles, the dispersion capacity of lanthanum oleate capped  $\text{LaF}_3$  nanoparticles in low polar or nonpolar organic solvent was also improved effectively after surface modification of lanthanum oleate surface capped  $\text{LaF}_3$  nanoparticles, because the modification agent lanthanum oleate reacted with the  $\text{LaF}_3$  nanoparticles. The lanthanum oleate capped  $\text{LaF}_3$  nanoparticles were distributed uniformly with average diameter of about 10 nm [25].

He et al. [26] synthesized surface-modified  $\text{Sm}_2(\text{CO}_3)_3$  nanoparticles in the mixture solvent of water–ethanol with the help of the surface modification agent of dialkylphosphate. The results indicated that the surface-modified nanoparticles of  $\text{Sm}_2(\text{CO}_3)_3$  could be dispersed in organic solvent and had lower friction factor, better antiwear and higher load-carrying capacity. Surface modification  $\text{Sm}_2(\text{CO}_3)_3$  nanoparticles used as lubricating oil additives was mainly organic fat chain antiwear in low load. When the load increases, the rare earth compound nanonucleus and friction surface film formed by active element sulfur and phosphorus via tribochemical reaction on the friction surface acted as antiwear and extreme pressure. Therefore, it may achieve continuous lubrication from the low temperature and load to the high temperature and load.

The surface modification methods for the nano rare-earth inorganic compound are mainly divided surface physical modification method and surface chemical modification method.

##### 4.4.1 Surface Physical Modification

Nanoparticles surface physical modification is a method that surface modification agent adsorb on the nanometer particle surface by physisorption in order to prevent nanoparticles agglomerating effectively. The nanoparticles prepared by this method are vulnerable to desorption under the condition of strong stirring and



agglomeration again, because the surface modification agent adsorption effect is not very strong. The surface physical modification method includes the following two factors:

#### **4.4.1.1 Surface Active Agent**

The surfactant molecule contains polar groups and nonpolar groups. When the surface modification nanoparticles disperse in the lubricating oil, the polar groups of surfactant interact with nanoparticles, and lipophilic groups dissolve in lubricating oil. Conversely, when the surface modification nanoparticles disperse in aqueous solution, the nonpolar groups of surfactant interact with nanoparticles, and the polar groups dissolve in water. Thus the inorganic nanometer particles can be dispersed in the lubricating oil well.

#### **4.4.1.2 Surface Sedimentation**

Surface sedimentation refers to a process bringing a kind of inert matter sedimentation to the nanometer particle surface and forms a cladding layer. It is required that the sediment can not take place chemical bonding with particles.

### ***4.4.2 Surface Chemical Modification***

The surface chemical modification plays a very important role in the surface modification of nanoparticles. In order to improve the dispersion stability of nanoparticles the nanoparticles surface structure and state are changed because adding the surface modification agent may modify the surface property due to the effect of chemical bonds. The surface chemical modification can be divided into the following three aspects:

#### **4.4.2.1 Surface Grafting Modification**

Surface grafting modification method is the macromolecule material through the chemical reaction to link to the surface of inorganic nanoparticles. This method can be divided into three types: particle surface polymerization growth grafting method, polymerization and surface modification synchronization method and coupling and grafting method.

#### 4.4.2.2 Esterification Reaction

Metal oxide and alcohol reactions can make the surface of nanoparticles changed from the hydrophilic oleophylic property to oleophylic hydrophobic property.

#### 4.4.2.3 Coupling Agent

When the inorganic nanoparticles with higher surface energy were mixed with the organic compound with lower surface energy, their compatibility was very poor. After the nanoparticles surfaces were disposed by the coupling agent, the dispersity in organic solvent would be greatly improved. This is often referred to as coupling agent technology. In view of its special properties, both ends of the group in the coupling agent molecules should play different roles: one organic functional group is able to react or dissolve with organics, and the other group is able to react with inorganic surface.

### 4.5 Typical Potential Rare-Earth Additives

In order to reduce friction and wear of mechanical parts to improve the mechanical efficiency, reduce energy consumption and prolong the machine life, the lubricating oil needs to add kinds of additives, like clearing agent, dispersing agent, viscosity index improver, extreme pressure and anti-wear agent, antioxidant, friction improver, pour point depressant and so on. Among them, the extreme pressure and anti-wear agent acts as a core role in the lubricating oil additives because it can generate lubricating film in machine part surface through the physical or chemical mechanism to reduce friction and wear, prevent sintering and abrade effect [27]. The research achievements of rare earth compounds in the field of tribology application shows that rare earth compounds in tribology has great application potential, especially as extreme pressure and anti-wear additives is promising.

Because rare-earth elements easily implement “in situ friction chemical processing” process, so the extreme pressure agent containing rare-earth elements possess the characteristics of extreme pressure antiwear and material surface modification. At present, it is paid more attention to the lubricating oil additives containing rare-earth including rare-earth inorganic compounds and rare-earth organic compound. The rare-earth inorganic compound mainly includes fluoride and borate. The rare-earth organic compound can be divided into rare-earth organic compound with and without phosphorus. The rare-earth organic compounds containing phosphorus mainly include rare-earth dialkyl dithiophosphate and rare-earth dialkyl phosphate. The rare-earth organic compounds without phosphorus mainly include rare-earth naphthenates, rare-earth dialkyl dithiocarbamate, rare-earth isooctate, rare-earth alkyl salicylate, etc.

## 4.5.1 Rare-Earth Fluoride

### 4.5.1.1 Antiwear Mechanisms of Rare-Earth Fluoride

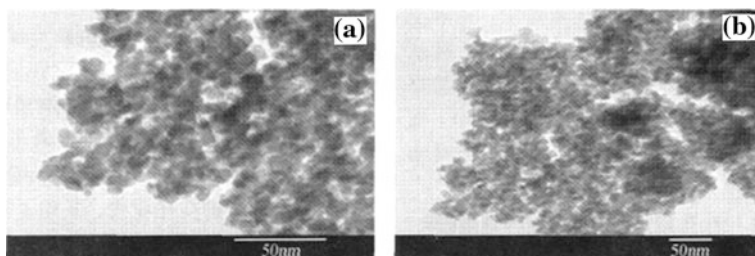
The rare-earth fluoride in the base oil has good extreme pressure, antiwear and friction performance, but its dispersion stability in the lubricating oil is not ideal as inorganic particles. The present studies mainly focus on the surface modification nanoparticles of rare-earth fluoride. Its antiwear mechanism is taking place tribochemical reaction in the friction process. The friction surface formed organic physical adsorption film and chemical reaction film which may contain rare earth oxide, ferrous fluoride, ferroferric oxide and so inorganic compounds.

### 4.5.1.2 Typical Potential Rare-Earth Fluoride Additives

The researchers from Lanzhou Chemical Physics Research Institute of Chinese Academy of Science firstly investigated on the  $\text{LaF}_3$  tribological properties in lubricating oil. They prepared succinic acid imide modified lanthanum fluoride nanocluster by using microemulsion method and found that the modified  $\text{LaF}_3$  nanocluster in liquid paraffin had good extreme pressure, antiwear and friction-reduction performance. The friction surface could take place tribochemical reaction in the friction process and formed the carbon, nitrogen organic physical adsorption film and tribochemical reaction film which might contain lanthanum oxide, ferrous fluoride, ferroferric oxide and so inorganic compounds [13]. Liang et al. [28] prepared the  $\text{LaF}_3$  nanoparticles via the coprecipitation surface modification in alcohol-water hybrid system, which was  $\text{LaF}_3$  nanocrystal for nuclear, organo-phosphorus compound for surface modification agent, good dispersion in organic solvent. It has been found that the  $\text{LaF}_3$  nanoparticles had high load-carrying capacity and good antiwear and friction-reduction properties because a firm adsorption film was formed on the friction surface during friction process.

The scholars from East China University of Science and Technology make series of researches on the nano- $\text{LaF}_3$  liquid additives. The surface modified  $\text{LaF}_3$  nano-particles were prepared in a water-ethanol system with  $\text{NH}_4\text{F}$  and  $\text{LaCl}_3$  as the raw materials, dioctyl dithiophosphoric acid di ( $\beta$ -ethoxyl) stearamine salt (DOPA) as surface modifier. A nano- $\text{LaF}_3$  powder additive and a liquid additive containing  $\text{LaF}_3$  nanoparticles were then obtained by different post-treatment methods.

In one method,  $\text{LaF}_3$  nano-powder was prepared by means of washing and drying. In another method, a liquid additive containing 10.2 %  $\text{LaF}_3$  was prepared by the phase transfer method in which  $\text{LaF}_3$  nanoparticles were moved from the aqueous phase to the oil phase by a solvent washing method with cetyl trimethyl ammonium bromide for coating agent. The dispersion stabilities of the two nano- $\text{LaF}_3$  additives in 500SN base oil were measured respectively by the centrifugal setting method and temperature-increasing method. The tribological properties of



**Fig. 4.2** TEM images of nano-LaF<sub>3</sub> powder and liquid additives. **a** powder; **b** liquid (from Zhou [29])

DOPA, nano-LaF<sub>3</sub> powder and liquid additives were investigated by four-ball tribometer. The wear scars on the steel ball surfaces were observed by scanning electron microscope (SEM). Figure 4.2 showed that the average size of LaF<sub>3</sub> nanoparticles was 10 nm; The dispersing stabilities of nano-LaF<sub>3</sub> liquid additive at room temperature and high temperature were both better than those of nano-LaF<sub>3</sub> powder additive [29]. And the dispersing stability of LaF<sub>3</sub> nanoparticles in the base oil has an important impact on their tribological properties [30]. Compared with DOPA and nano-LaF<sub>3</sub> powder additive, the nano-LaF<sub>3</sub> liquid additive had the best tribological properties. When the nano-LaF<sub>3</sub> liquid additive was added in 500SN base oil with LaF<sub>3</sub> mass fraction of 0.8 %, the extreme pressure ( $P_B$  value) of the oil sample was increased from 510 to 1069 N, and the wear scar diameter (WSD) was reduced to 0.352 mm.

Besides the amorphous rare-earth fluoride particles, the morphology controllable synthesis of nanoparticles also showed a good tribological property. The surface modified LaF<sub>3</sub> nanorods were prepared in a mixture solution of water–ethanol with dioctyldithiophosphoric acid di ( $\beta$ -ethoxyl) stearamine as a surface modifier by an ultrasonic method. The oil-soluble nano-LaF<sub>3</sub> additives were obtained by a phase transfer method. The diameter of LaF<sub>3</sub> nanorods was 4–7 nm, and its length was 20–30 nm. LaF<sub>3</sub> nanorods had a good dispersing stability and high temperature stability in 500SN base oil. Compared with 500SN base oil the maximum non-seizer load with the liquid additive was increased by 109.61 %, and the WSD was decreased by 39.13 %. According to the AES analysis, a complex film was formed by LaF<sub>3</sub> deposition and other friction elements (such as sulphur, phosphors) on the friction surface of steel balls, and LaF<sub>3</sub> was penetrated to the sub-surface, which jointly improved the tribological properties of 500SN base oil [31].

CeF<sub>3</sub> nanocrystals synthesized via ultrasound assisted route showed that the shape and size of products were strongly depended on the ultrasonic time. The morphology of the CeF<sub>3</sub> nanocrystals changed from “lantern” particles to uniform disk-like particles with the ultrasonic time. The ultraviolet absorption peak and the hand-edge of the CeF<sub>3</sub> nanocrystals exhibited a pronounced blue shift as the size of particles decreasing. The photoluminescence spectrophotometer intensity of CeF<sub>3</sub> nanocrystals was related to the crystalline extent [32]. Ma et al. [33] prepared

surface oleic acid modified hollow LaF<sub>3</sub> nanoparticles in an oleate/cetyltrimethylammonium bromide micellar emulsion. The LaF<sub>3</sub> nanoparticles with diameters of 17.5 nm exhibited excellent antiwear and friction reduction properties.

Chen et al. [34] prepared oil-soluble calcium-doped LaF<sub>3</sub> nano-particles and LaF<sub>3</sub> and CaF<sub>2</sub> mixed nano-particles in a mixture solution of water–ethanol with dioctyl dithiophosphoric acid di ( $\beta$ -ethoxyl) stearamine as a surface modifier, and calcium chloride and lanthanum chloride as reactants. The results showed that the nanoparticles were the mixture of LaF<sub>3</sub> and CaF<sub>2</sub> crystals when the molar ratio of La and Ca was 1:1, the crystalline phases of the nanoparticles were dopant structures when the molar ratio of La and Ca was 1:2. The morphology and particle diameter of the two nano-particles were the same, and the average diameter was about 10 nm. The two nano-particles had good dispersing stability and high temperature stability in 500SN base oil. The tribological properties of the dopant nano-particles were higher than that of the mixture nano-particles. Compared with 500SN base oil the maximum non-seizer load value of the oil sample with the dopant nano-particles was increased by 1.15 times, and the WSD was decreased by 42.6 %.

Yu et al. [35, 36] prepared four kinds of LaF<sub>3</sub> nanoparticles in the mixture solvent of water–ethanol by using surface modifiers with different lipophilic chain lengths, such as lauric acid diethanolamine, myristic acid diethanolamine, palmitic acid diethanolamine and stearic acid diethanolamine. The diameter of four kinds of LaF<sub>3</sub> nanoparticles was among 20–30 nm. The surface modifiers were adsorbed on the nanoparticles surface in chemistry adsorption and formed organic modified layer. The results showed that the tribological properties of LaF<sub>3</sub> nanoparticles in the base oil were improved as the lipophilic chain length of surface modifiers increased. The longer the lipophilic chain was, the better the susceptibility of the nanoparticles to liquid paraffin wax was. This was mainly because that the lipophilic chain length influenced the interface activity of the nanoparticles. Moreover, for homologous organic acid soaps, the longer the lipophilic chain was, the better its friction reducing performance was. The properties of base oils also influenced the interface activity of the nanoparticles.

Tong et al. [37] in situ synthesized the LaF<sub>3</sub> nanoparticles which the size was about 5–25 nm in Tween-80/n-C<sub>10</sub>H<sub>21</sub>OH/H<sub>2</sub>O lamellar liquid crystal and studied on the friction and wear behaviors of LaF<sub>3</sub> nanoparticles as an additive in lamellar liquid crystal. The results indicated that the LaF<sub>3</sub> nanoparticles could improve the lubrication performance of lamellar liquid crystal and the worn surfaces formed tribochemical reaction film containing fluorine during the friction process. In addition, the polarity groups from amphiprotic molecule Tween-80 and n-C<sub>10</sub>H<sub>21</sub>OH quite easily adsorbed to the steel ball surface during the friction process and played a friction reduction and antiwear role. The synergistic effect of formed inorganic reaction film and organic adsorption film made the mixed system of LaF<sub>3</sub> nanoparticles and lamellar liquid crystal possessed excellent lubrication performance.

Gao et al. [38] synthesized the surface modified rare-earth fluoride nanoparticles whose surface was modified by DDP-18(PyDDP-18) in water–ethanol solvent

by using coprecipitation surface modified method. The test results indicated that the surface modified rare-earth fluoride nanoparticles lubricating oil additive could be well dispersed in organic solvent and had better antiwear and friction reducing capacity.

## **4.5.2 Rare-Earth Borate**

### **4.5.2.1 Antiwear Mechanisms of Rare-Earth Borate**

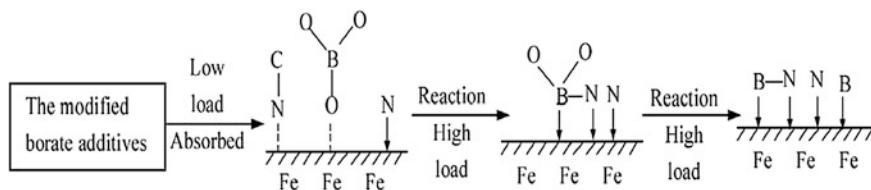
Rare-earth borate which is a typical inactive, non-toxic, odorless and non-polluting environment lubricating oil additive has excellent tribological properties. The antiwear mechanism of rare-earth borate as lubricating oil additive mainly exist two ideas: sedimentary film and boriding. It was thought that borate might form deposited film on the friction surface to generate the antiwear and friction reduction effect.

At the same time, hydration borate microspheres could melt and become sticky and sliding particles lubricating layer to generate the friction reduction effect. Adams [39] put forward the point of electrophoresis film. Qiao et al. [40] found that surface modification of borate lubricating oil additive could form mixed film which was mainly composed of adsorption film and tribochemical reaction film on the contact frictional surfaces. The elements chemistry state and content in the tribochemical reaction film were relevant to test load in a fixed speed, and changed along with the depth of the reaction film. The content of inorganic compounds which mainly formed by Fe and B, N and O gradually increased along the depth direction. The mechanism was shown in Fig. 4.3.

The key points of boriding viewpoint are following aspects: there is solid and continuous surface film on the wear surface which is formed by B interstitial compound  $Fe_xB_y$  which can dissolve the B in free state and form solid solution; Finally, on the friction surface a complex coating ( $B/Fe_xB_y/Fe_xO_y$ ) was formed. Thereby, borate lubricant additives can reduce shear stress, improve the load bearing capacity and had the effect on the antiwear and friction reducing [41]. Chen et al. [42] put forward “rare-earth friction diffusion” and “rare-earth friction speed up boriding” ideas from the essence of tribochemical treatment in situ and chemical heat treatment. These viewpoints think that the rare-earth elements can promote boron element in the rare-earth borate lubricant additive to diffuse into metal friction pair material in the friction process, increase the thickness and hardness of friction boronizing layer.

### **4.5.2.2 Typical Potential Rare-Earth Borate Additives**

The oil soluble rare-earth borate lubricating oil additive has both good load bearing capacity and antiwear property, but also it can be dissolved in feedstock oil



**Fig. 4.3** Formation of mixed film on contact surface of frictional pair under the lubrication of oil containing borate lubricant additives (from Qiao [40])

well without any disperser. It overcomes the weakness of poor dispersion stability of inorganic borate. It represents the development direction of borate lubricating additive. For example, cerium isooctoxyborate, lanthanum dodecoxyborate, cerium bodecmercapborate, lanthanum cyclohexoxyborate can form permeable layer in the friction process, carry out in situ friction chemical modification on the friction surface, produce lubrication film including  $\text{Ce}_2\text{O}_3$ ,  $\text{CeO}_2$  and  $\text{B}_2\text{O}_3$ . The rare-earth elements have sped up permeability function to boron, carbon, nitrogen and so on [43–45]. The amorphous lanthanum borate nanoparticle additive also has good antiwear and load carrying capacity [46].

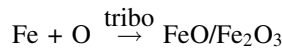
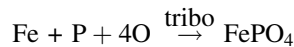
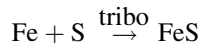
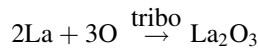
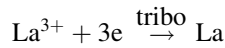
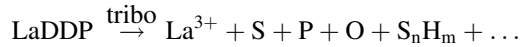
Rare-earth borate is a kind of lubricating oil additive which is of high efficient, multi-function, non-toxic and tasteless. It has special extreme pressure and antiwear property, good oxidation stability, anticorrosion performance and good sealing performance. It has more advantageous than additive with phosphorus or sulfur. It has been applied in the industrial gear oil, two-stroke lubricating oil and vehicle gear oil, and has obvious effect on the aspect of energy saving. But the borate storage stability and resistance to emulsifying ability need to improve. There are still a large number of problems to be further researched on the synthesis, tribological properties and mechanism of nanometer additive.

### 4.5.3 Rare-Earth Dialkyldithiophosphate

#### 4.5.3.1 Antiwear Mechanisms of Rare-Earth Dialkyldithiophosphate

The rare-earth dialkyldithiophosphate (REDDP) is a kind of rare-earth compound which is based on the molecular structure of commonly used extreme pressure and antiwear additive ZDDP. Its excellent antiwear and friction reducing performance mainly come from two aspects: it takes place firstly tribochemical reaction on the friction surface and forms the multiphase surface lubrication film including RE or  $\text{RE}_2\text{O}_3$ ,  $\text{Fe}_2\text{O}_3$  or FeS, organic sulfide, sulfate and phosphate etc. which may improve anti-friction performance; On the other hand, it forms rare-earth friction diffusion layer on the friction sub-surface, change the surface crack propagation resistance and surface compressive stress, increases the metal friction pair material hardness, improves the antiwear performance [47–51].

Jiang et al. [52] studied on the tribochemical characteristics and the tribological mechanism about the new oil-soluble lanthanum dialkyldithiophosphate (LaDDP) additives under boundary lubrication conditions. They simulated the tribological decomposition reactions through mass spectroscopy. The experimental results indicated that LaDDP possessed better lubrication ability than ZDDP. The reason could mainly be attributed to the formation of boundary film containing La, La<sub>2</sub>O<sub>3</sub>, FeS, sulphate, phosphate and lanthanum diffused layer. Through the analysis the main mass spectrum fragments and relative intensity (Table 4.1), they speculated that LaDDP decomposition process may be shown as follows:



#### 4.5.3.2 Typical Potential Rare-Earth Dialkyldithiophosphate Additives

The rare-earth elements (La, Pr, Sm, Eu, Gd) dialkyldithiophosphate (REDDP: RE[S(S)P(OR)<sub>2</sub>]<sub>3</sub>) as lubricating oil extreme pressure and antiwear additives, their antiwear ability are obviously superior to ZDDP. REDDP lubrication performance may be related to the species of rare-earth elements. Their change rule is as follows: Sm > La > Pr > Gd > Eu, which may be related to the rare-earth metal lattice structure. For the rare-earth metal with hexagonal crystal state, the antiwear performances of rare-earth compounds become better and better with the increasing of hexagonal crystallographic axis *c* [47–51].

Feng et al. [53–55] synthesized four kinds of oil-soluble lanthanum (neodymium) dialkyldithiophosphate, LaDDP-3, LaDDP-8, NdDDP-3, NdDDP-8, from P<sub>2</sub>S<sub>5</sub>, isopropanol or isoctyl alcohol and lanthanum chlorid or neodymium chloride. Tribological behavior was evaluated and compared with that of commercial zinc dialkyldithiophosphate. The results indicated that REDDP synthesized for this study showed more excellent tribological characteristics than ZDDP and antiwear property of LaDDP was superior to NdDDP. The rare-earth elements can diffuse into the



**Table 4.1** Main mass spectrum fragments and relative intensity

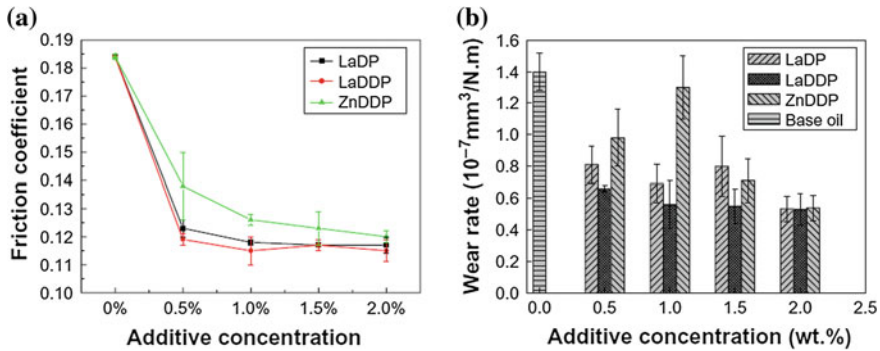
Main mass spectrum fragments	Mass number	Relative intensity
$C_8H_{16}$	112	74.29
$C_8H_{17}$	113	100
$OC_8H_{17}$	129	17.93
$O(C_8H_{17})_2$	242	17.29
$P(C_8H_{17})_2$	289	19.41
$SP(C_8H_{17})_2$	321	12.9
$S(S)P(C_8H_{17})_2$	353	63.52
$LaSP(C_8H_{17})_2$	400	17.61
$LaS_3P_2O_4C_{82}H_{61}$	806	34.81
$La[S(S)P(C_8H_{17})]_3$	1198	14.51

From Jiang [52]

wear spot surface under the tribological condition and form a rare-earth rich layer, which might be the main reason for the better extreme-pressure and antiwear properties of REDDP.

He et al. [56] prepared ten kinds of LaDDPs with different alkyl chains and studied the structure-behavior relationship of LaDDP. The properties of antiwear and extreme pressure of LaDDPs as lubricating oil additive in 500SN base oil were investigated by four-ball machine, and the effects of the length of alkyl chain and the structure of primary or secondary alkyl of LaDDP additive on the properties of friction-wear and extreme pressure were explored. The results showed that for carbon numbers in alkyl in a range of 3–8, the additives exhibit excellent properties of antiwear and extreme pressure. The antiwear abilities are improved obviously with the increase of carbon number in alkyl. Moreover, the properties of friction antiwear of secondary alkyl LaDDPs have advantages over that of primary alkyl LaDDPs. It is suggested that extreme pressure abilities decrease with the increase of carbon numbers in alkyl and the alkyl construction has little effect on the extreme pressure property.

Zhang et al. [57] tested two kinds of rare earth complexes of LaDDP and lanthanum dialkylphosphate (LaDP) as lubricant additives in liquid paraffin for the untreated 60Si<sub>2</sub>Mn steel and laser-cladding Ni<sub>35</sub>A coating on 60Si<sub>2</sub>Mn steel sliding pairs. The results showed that treated laser cladding coatings of steel can improve their hardness and strength and the coated steel possess higher load-carrying capacity than that of 60Si<sub>2</sub>Mn. Figure 4.4 indicated that the rare-earth complexes of LaDDP and LaDP possess good oil-solubility, friction-reducing and wear resistance properties. The antiwear reason is that those rare earth complexes can form a protective film containing rare-earth oxide, sulfate and sulfur compounds during the friction process when they acted as additives in liquid paraffin.



**Fig. 4.4** The evolution of friction coefficient (a) and wear rate (b) with during a load test LaDDP, LaDP and ZDDP of additives concentration and pure liquid paraffin at room temperature for the laser-cladding coatings treatment of 60Si<sub>2</sub>Mn steel (from Zhang [57]). Experimental conditions: frequency: 20 Hz, stroke: 1 mm, load: 100 N and duration: 30 mm

## 4.5.4 Rare-Earth Naphthenates

### 4.5.4.1 Antiwear Mechanisms of Rare-Earth Naphthenates

The main reason of this kind additive with excellent tribological performance is that the rubbing surfaces are of high levels of iron oxide catalytic oxidate rare-earth nanoparticles and form the rare-earth rich boundary film including iron oxide, organic acid, rare earth oxide, rare-earth compound during the friction process [58].

### 4.5.4.2 Typical Potential Rare-Earth Naphthenates Additives

Liu et al. [58] evaluated the tribological behavior of rare-earth naphthenate (REN) as lubricating oil additive in ISO VG26 white oil. The average grain diameter of as-prepared nanoparticles was less than 30 nm. The results showed that the REN exhibited good antiwear behavior and load-carrying capacity in the base stock. Moreover, REN also showed the synergistic extreme pressure and antiwear functions when it was complex with sulfurized isobutylene, while it had no such synergistic action with TCP. The excellent antiwear and extreme pressure properties of REN additive were attributed to the formation of a boundary lubricating film mainly composed of rare-earth compounds on the rubbing steel surface [59].

Besides above mentioned rare-earth compounds, both of rare-earth alkylsali-cylate [60] and rare-earth carboxylate [61] are also of excellent tribological properties and may become new lubricant additives containing rare earth.

## 4.6 Synergistic Effect of Rare-Earth Compounds with Other Additives

The synergistic effect of rare-earth compounds together with other additives is one of the important properties. The rare-earth dialkyladithiophosphate (LaDDP, SmDDP, EuDDP) combined with boric acid ester (OB) as a complex additive has excellent synergistic lubrication effect in ISO VG68 machinery oil. The complex additive of the REDDP and OB is more effective to increase the friction-reducing, antiwear, and extreme-pressure properties of the base stock, compared with the individual additive. It is attributed to the formation of an outstanding boundary lubricating and protective film composed of iron oxides, rare-earth oxides, FeS, FeSO<sub>4</sub>, FePO<sub>4</sub>, B<sub>2</sub>O<sub>3</sub>, organic boron compound and B by way of tribochemical reactions. On the other hand, it is because the tribochemical reaction products of REDDP acted to catalyze the decomposition of boric acid ester and promote the diffusion of the resulting B-containing products into the sub-worn-surface, and formed RE-B common friction permeation layer on the sub-worn-surface. As a result, the material hardness increased and wear resistance reinforced [62–64].

In ISO VG26 white oil, it has synergistic effects for the REN naphthenates and stannous naphthenate (SN) in the anti-wear and friction-reducing capabilities. The complex of REN and SN possesses better lubrication abilities than that of ZDDP. Especially this sort superiority is more obvious under high load condition. The excellent performance of the complex can be attributed to the formation of a boundary lubricating film mainly composed by rare earth and tin on the friction surface. It may be possible for the complex to be used as a novel and efficient and multifunctional lubrication additive in industries [65, 66].

It was found that the rare-earth organic compound additives had excellent friction-reducing property, and had good synergism with the additive dialkyl dithiocarbamate in case of studying on the tribological properties of rare-earth dialkyldithiocarbamate (La and Ce), REN (La and Ce) in hydrotreated lubricating oil. The lubricating mechanism of rare-earth organic compound is that the surface layer formed by the additives is composed of organic film, oxidation film, chemical reaction film, which improves the tribological properties of lubricating oil [67].

The research results on the tribological performance and tribochemical mechanisms of lanthanum cyclo-hexoxylborate (LaCHOB) and calcium iso-octoxylborate (CaIOOB) and their mixture as additives in HVI500 mineral oil indicated that the two additives showed good friction-reduction, antiwear and extreme pressure properties in the base stock. They showed a synergistic effect in improving the extreme pressure behavior but no such effect for friction-reduction and antiwear ability. The worn steel surfaces lubricated with LaCHOB-CaIOOB/base stock had higher atomic concentration of Ca as compared with the surface lubricated with CaIOOB/base stock. The reason may be that La in LaCHOB additive acts to promote the permeation of Ca in CaIOOB during the friction process [68].

In a word, the research on the synergism about the rare-earth compounds with other extreme pressure and antiwear additives is still in the stage of exploration. The mechanism of synergistic effect also needs to be further investigated.

## 4.7 Conclusions

The rare-earth compounds are expected to become new and efficient lubricating oil additives in the future, but there are many aspects that need to be further investigated, such as

- In-depth studying the change rule of tribological properties and the tribochemical nature of rare-earth materials.
- Clarifying the action mechanism of rare-earth elements.
- Developing the rare earth nanometer lubricating material combining with the development of nanotechnology.
- Solving the dispersion of nano rare-earth compounds in base oil.
- Carrying on the synthesis new rare-earth organic compounds and researching on the synergism about the rare-earth organic compounds with other additives.
- The synthesis researching on nano rare-earth inorganic compound modified by different organic compounds.
- Because the soot particles increase the wear of diesel engine, it is necessary to explore the effect of rare earth elements on the tribological performance of carbon film and friction-reduced catalytic function in order to find the ways to reduce soot-induced wear problem.

**Acknowledgments** The authors wish to express their thanks to Mr. Xiangyang Wang, Mr. Ruhong Song, Dr. Jinsi Chen and Professor Xifeng Zhu for their assistance in the summary and discussion. The financial support from the National High Technology Research and Development Program of China (863 Program) (Grant No. 2012AA051803) and National Natural Science Foundation of China (Grant No. 51275143) are gratefully acknowledged.

## References

1. Li J, Wang Y, Ren T (2006) A survey on the substitute for ZDDP. *Lubr Eng* 16:168–172
2. Olomolehin Y, Kapadia R, Spikes H (2010) Antagonistic interaction of antiwear additives and carbon black. *Tribol Lett* 37:49–58
3. Hakan K (2001) The impact of crankcase oil containing phosphorus on catalytic converters and engine exhaust emission. *Ind Lubr Tribol* 53:237–255
4. Patel M, Aswath PB (2012) Morphology, structure and chemistry of extracted diesel soot: Part II: x-ray absorption near edge structure (XANES) spectroscopy and high resolution transmission electron microscopy. *Tribol Int* 52:17–28
5. Zhang Z, Liu W, Xue Q (1996) Study on tribological properties of the complex of rare earth dialkylldithiocarbamate and phenanthroline in lubricating grease. *Wear* 192:6–10

6. Xue Q, Lian Y, Yu L (1996) The antiwear and extreme pressure properties of some oil-water double soluble rare earth complexes part 1: their tribological behavior in water. *Wear* 196:188–192
7. Yu L, Lian Y, Xue Q (1998) The tribological behaviors of some rare earth complexes as lubricating additives. I. An investigation of the friction-reducing behaviors and an XPS study of the tribochemical characteristics. *Wear* 214:47–53
8. Yu L, Nie M, Lian Y (1996) The tribological behaviour and application of rare earth lubricants. *Wear* 197:206–210
9. Zhang F, Song B, Qu J, Wang L (2006) Application of nanoadditives in lubricants. *Lubr Eng* 31:190–195
10. Li L (2011) Preparation and properties of Eu: La<sub>2</sub>CaB<sub>10</sub>O<sub>19</sub> by sol-gel method. *Guangdong Chem Eng* 38:117–118
11. Lindma B, Danielson L (1981) The definition of microemulsion. *Colloids Surf* 3:391–392
12. Liu C, Zhou X (2005) Study on the synthesis of LaF<sub>3</sub> nanoparticles in microemulsion system. *Pet Process Petrochemicals* 36:61–64
13. Zhang Z, Liu W, Xue Q (2000) Tribological properties of LaF<sub>3</sub> nanocluster modified with compound containing nitrogen. *Tribology* 20:217–219
14. Jing X, Hong G, Li Y (1990) Preparation and ultrafine powders of rare earth compounds by alcoxide method. *Chin J Appl Chem* 7(2):92–94
15. Jia Z, Xia Y (2011) Hydrothermal synthesis, characterization, and tribological behavior of oleic acid-capped lanthanum borate with different morphologies. *Tribol Lett* 41:425–434
16. Liu Y, Zhao Y, Luo H, Wu H, Zhang Z (2011) Hydrothermal synthesis of CeF<sub>3</sub> nanocrystals and characterization. *J Nanopart Res* 13(5):2041–2047
17. Ye Y, Dong J, Chen G, Chen B, Hu Z (2000) Special tribological behavior of nanometric borate. *Lubr Eng* 4:20–21
18. Ye B, Tao D (2005) Tribo-chemistry properties of nano-particles additives in environmentally friendly base fluids. *Mater Mech Eng* 29:55–58
19. Ni X, Shen J, Zhang Z (2006) Physicochemical character and application of nanometer material. Chemistry Industry Press, Beijing
20. Wang F, Yuan X, Guo H (2008) The study on the development of nanometer lubrication additives. *Intern Combust Engines* 4:1–3
21. Gu C, Zhu G, Tian X, Ji G, Li W (2009) Study on the tribological properties of the diesel engine's lubricating oils containing nanoparticles. *Adv Mater Res* 79–82:255–258
22. Li B, Wang X, Liu W, Xue Q (2006) Tribochemistry and antiwear mechanism of organic-inorganic nanoparticles as lubricant additives. *Tribol Lett* 22:79–84
23. Chen J, Li S, Jiang X, Li S, Ye P (2001) Size effect and tribological behavior of nickel oxythiomolybdate as an additive in lubricating oil. *Tribology* 21:366–370
24. Ye Y, Dong J, Chen B, Hu Z, Fan J, Ye X (2000) Effect of wear resistance on nanometer lanthanum borate in lubricating oils. *Lubr Eng* 2:23–24
25. Li N, Mai R (2009) Preparation and characterization of oil-soluble LaF<sub>3</sub> nanoparticles. *Guangdong Chem Eng* 36:56–57
26. He Y, Bian Z (2011) Synthesis and tribological properties of surface modified Sm<sub>2</sub>(CO<sub>3</sub>)<sub>3</sub> nanoparticle. *Huaxue Shiji* 33:925–928
27. Pan Y, Li J, Mei J (2005) Latest development and application of EP antiwear additive present status analysis. *Automobile Technol Mater* 10:6–10
28. Liang Q, Zhang Z, Xue Q, Liu W (1999) Tribological properties of LaF<sub>3</sub> nanoparticle. *Chin Rare Earths* 20:32–35
29. Zhou X, Li Z, Yu G, Jin Y, Li C (2008) Preparation of liquid lubricant additive containing LaF<sub>3</sub> nano-particles and its lubricating performance. *ACTA Pet Sinica (Petroleum Processing Section)* 24:325–331
30. Yi S, Yu G, Zhou X, Jin Y, Li C (2006) Effect of dispersivity and stability of LaF<sub>3</sub> nanoparticles in lubricating oils on their tribological properties. *J East China Unive Sci Technol (Natural Science Edition)* 32:1392–1395

31. Zhou X, Li Z, Yu G, Jin Y (2008) Study on preparation of oil-soluble LaF<sub>3</sub> nanorods and its tribological properties. *Lubr Eng* 33:15–17
32. Liu Y, Zhao Q, Zhao Y (2009) Synthesis and characterization of CeF<sub>3</sub> via ultrasound assisted route. *J Inorg Mater* 24:549–552
33. Ma X, Zhao Y, Wu Z (2008) Preparation of surface-modified hollow LaF<sub>3</sub> nanoparticles and their tribological performances. *Acta Phys Chim Sin* 24:2037–2041
34. Chen X, Yu G, Zhou X, Li Z (2008) Study on preparation of oil-soluble calcium-doped LaF<sub>3</sub> nano-particles and its tribological properties. *Lubr Eng* 33:25–28
35. Yu G, Zhou X, Li Z, Wu W, Zhu M (2007) Effect of length of surface modifier lipophilic chain on tribological properties of LaF<sub>3</sub> nanoparticles. *Lubr Eng* 32:50–53
36. Yu G, Jin M, Wan K, Lu P (2011) Effect of surface modifier on their tribological properties of LaF<sub>3</sub> nanoparticles in base oil. *Adv Mater Res* 230–232:774–781
37. Tong F, Yuan F, Wu Y, Wu X, Tang D, Yang H (2009) In-situ synthesis and lubricating mechanism of LaF<sub>3</sub> nanoparticles. *Chem Bioeng* 26:43–47
38. Gao L, Li J, Fu L, Zhang Y, Zhang L (2006) Synthesis and application of nanore-earth fluoride lubricating oil additives. *J Xi'an Shiyu Univ (Natural Science Edition)* 21:83–86
39. Adams JH, Godfrey D (1981) Borate gear lubricant-EP film analysis and performance. *Lubr Eng* 37(1):16–21
40. Qiao Y, Xu B, Ma S, Liu W, Xue Q (2001) Study on tribochemical mechanisms of the modified borate additives. *Tribology* 21:416–420
41. Dong J, Chen G (1994) A new concept-formation of permeating layers from nonactive antiwear additive. *Lubr Eng* 50:17–22
42. Chen B, Dong J, Chen G, Ye Y (1996) Synergistic lubrication mechanism of a lanthanum-boron composite lube additive. *J Chin Rare Earth Soc* 14:306–310
43. Li F, Chen B, Huang W, Dong J (2000) Preparation and tribological behaviors of cerium isooctoxyborate. *Tribology* 20:220–222
44. Li F, Liang Y, Huang W, Huang K, Wang Z (2005) Synthesis and tribological properties of lanthanum dodecoxy borate as antiwear and friction-reducing additive. *Lubr Eng* 1:52–54
45. Li F, Liang Y, Huang W, Wang Z (2003) Synthesis and tribological behavior of cerium bodecmercaptopborate. *Hunan Nonferrous Met* 19:30–32
46. Hu Z, Dong J, Chen G (2000) Preparation and tribological properties of nanoparticle lanthanum borate. *Wear* 243:43–47
47. Chen B, Dong J, Chen G (1995) Studies on tribochemical characteristics of gadolinium dialkyldithiophosphate. *ACTA Pet Sinica (Petroleum Processing Section)* 11:68–74
48. Chen B, Dong J, Chen G (1995) Study on antiwear and antifricition mechanism of dialkyldithiophosphate praseodymium. *J Rare Metal* 19:32–36
49. Chen B, Dong J, Chen G (1994) Study on the tribological behaviors of europium dialkyldithiophosphate. *Tribology* 14:153–161
50. Chen B, Dong J, Chen G (1995) Study on the preparation and characteristics of lanthanum dialkyldithiophosphate. *Synth Lubricating Mater* 23:1–3
51. Chen B, Chen G, Ye Y (1996) Preparation and anti-wear characteristics of lanthanum alkyl phosphonic mono-ester. *Petrochemical Eng* 25:821–874
52. Jiang S, Chen G, Cheng X, Cheng B, Yu H (2000) Study on the tribochemical mechanism of lanthanum dialkyldithiophosphate. *J Chongqing Univ (Natural Science Edition)* 23:15–18
53. Feng Y, Tian Y, Sun X, Sun L, Cai W (2002) Investigation on tribological behaviors of neodymium dialkyldithiophosphate. *J Chin Rare Earth Soc* 20:207–210
54. Feng Y, Sun L, Sun X, Hu Y, Cai W (2002) Study on tribology of oil-soluble lanthanum (neodymium) dialkyldithiophosphates. *J Harbin Inst Technol* 34:324–328
55. Tian Y, Ren N, Feng Y, Sun L, Cai W (2004) Investigation on tribological behaviors of neodymium dialkyldithiophosphate. *J Harbin Inst Technol* 11:34–37
56. He M, Li W, Guo L, Cheng X (2007) Structure-behavior relationship of lanthanum dialkyldithiophosphate as lubrication additive. *J Jiangsu Univ (Natural Science Edition)* 29:445–452

57. Zhang H, Xia Y, Liu Z (2012) Tribological properties of two kinds of rare earth complexes as lubricant additives for laser cladding coatings. *Ind Lubr Tribol* 64:23–32
58. Liu R, Wei X, Tao D (2010) Study of preparation and tribological properties of rare earth nanoparticles in lubricating oil. *Tribol Int* 43:1082–1086
59. Liu R, Tao D, Zhang J, Cao W, Sun J (2003) Tribological behavior of rare-earth naphthenate as lubricating oil additive. *Tribology* 23:394–397
60. Liu R, Tao D, Zhao Y (2008) Tribological performance of rare-earth alkylsalicylate as lubrication additive. *Ind Lubr Tribol* 60(1):9–13
61. Wang Z, Hu L (2010) Tribological study of a novel oil-soluble lanthanum carboxylate as lubricant additive. *Tribology* 30:140–144
62. Chen B, Dong J, Ye Y, Chen G (1996) Tribological characteristics of gadolinium thiophosphate and borate composite lube additive. *Mater Mech Eng* 20:14–16
63. Chen B, Dong J, Chen G, Ye Y (1996) Tribology characteristics of a samarium-boron composite lube additive. *Lubr Eng* 20:14–17
64. Jiang S, Cheng X, Chen B, Wei Y (2004) Synergistic friction-reducing and antiwear effect of lanthanum dialkyldithiophosphate and boric acid ester. *Tribology* 24:38–41
65. Liu R, Tao D (2004) Study on the tribological properties of a novel complex containing rare-earth and tin as lubricant additive. *Mech Sci Technol* 23:30–32
66. Liu R, Tao D, Zhang J, Fu S, Jiang H (2004) Synergistic lubrication effects of organorare-earth and organotin compounds. *J Chin Rare Earth Soc* 22:225–228
67. Ye B, Tao D (2009) Study on tribological properties of rare-earth organic compound as additives in hydrotreated lubricating oil. *Lubr Eng* 34:81–83
68. Li F, Huang W, Chen B, Dong J (2000) Effect of lanthanum cyclohexoxylborate and calcium iso-octoxylborate as additives on the tribological behavior of HVI500 mineral oil. *Tribology* 20:392–394

# Chapter 5

## Structural Dynamics and Viscoelastic Passive Damping Treatments

R. A. S. Moreira

**Abstract** Viscoelastic damping treatments are an interesting solution for the vibration control of light and large structures. Despite the simplicity of the damping mechanism that characterizes this vibration control system, the design of these treatments is complex and requires some specific strategies to handle properly the description of the material behavior, to represent accurately the kinematics of all the layers of the damped structure and to apply optimization procedures to improve the damping efficiency. This Chapter presents a review on viscoelastic damping treatments, and those important issues related to the material characterization, numerical modeling and optimization are thoroughly analyzed and explored.

### 5.1 Introduction

Structural engineering is a research field where a continuous evolution of the applied materials and manufacturing methods has provided the instruments for the development of lighter, efficient and optimized structures that fill our landscape. From airplane fuselages and aerospace structures to the tallest skyscrapers, all these remarkable engineering productions were only possible due the continuous evolution of the research and experience-based knowledge of the structural engineering field.

The trends of the structural engineering are focused onto the ability to construct the most efficient structure using light and tailored materials, by employing cost-efficient manufacturing methods. The use of aided design and optimization

---

R. A. S. Moreira (✉)  
Department of Mechanical Engineering, University of Aveiro,  
Campus Santiago, Aveiro 3810-193, Portugal  
e-mail: rmoreira@ua.pt



methods based on numerical simulation tools provide most of the required resources to aim this goal; but the development of more efficient materials plays also an important role in this structural progress.

## 5.2 Structural and Material Damping

When dealing with light and stiff materials, currently considered as the first choice for the structural engineer, an immediate outcome comes towards the designer attention: these efficient materials do not provide an evident and valuable energy dissipation mechanism. These dissipation mechanisms, or structural damping sources, are easily found in old structures made of timber and cast iron using riveted or bolted joints. In fact, these materials possess an interesting material damping capability able to mitigate the undesired levels of vibration caused by internal or external sources. Likewise, the presence of joints using rivets and bolt connections provides an effective dissipation mechanism within the vibration energy path.

Nowadays, more efficient materials are used—like aluminium and titanium alloys, carbon and aramid composites, polymeric and metal foams—providing a valuable structural strength in combination with a very low mass. However, the material damping characteristic of these materials is quite reduced and unable to provide a valuable dissipation mechanism. Therefore, a workaround for this missing feature has been driving the attention of the structural engineer, in order to design and apply in those critical structures the most effective way to recover the lost damping capacity. Add-on devices and mechanisms, often named as damping treatments, have been designed, tested and implemented for this purpose. These damping mechanisms are designed as additions to the base structure and have been strategically tailored to promote a simple and cost effective vibration control solution, while maintaining the high stiffness-weight ratio that characterizes the trends of the contemporary structural engineering.

## 5.3 Damping Treatments: Active and Passive Devices

Damping treatments can be implemented using passive or active-based mechanisms. Passive devices rely on the addition of tailored mechanisms based on materials, fluids or electrical systems able to dissipate a significant amount of the deformation energy that is absorbed by the damping treatment when the host structure vibrates. Shunted-piezoelectric ceramics, shape-memory alloys and viscoelastic material (VEM) layers represent interesting forms of passive damping implementation and signify an important on-going research effort.

Active damping, often called active control, classifies a set of devices that actively react against the vibration motion to mitigate it. These devices include

those mechanisms based on piezoelectric patches, electro-rheologic and magneto-rheologic fluids, and shape-memory alloys and polymers.

While the active devices provide an efficient control mechanism able to reduce the undesired levels of vibration within a low frequency range, it requires power, vibration transducers and control hardware. On the contrary, passive devices are effective when working within a medium–high frequency range, can operate with a minimum set of external hardware or even in an autonomous fashion, and do not rely on external power sources and control circuit with the associated potential failures.

## 5.4 Viscoelastic Damping Treatments

Damping treatments using the viscoelastic effect apply soft polymers having a significant material damping capability. These materials are formed by long reticulated molecular chains, which are responsible for the high-energy dissipation effect when cyclic deformed. This molecular arrangement is also responsible for the important dependence of the viscoelastic materials upon frequency and temperature.

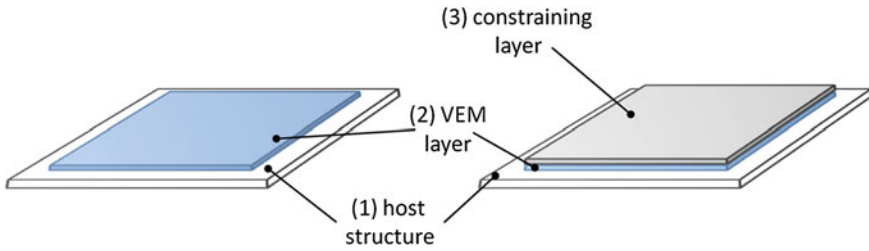
Viscoelastic materials are applied as layers deposited onto the surface of the host structure following two different treatment configurations: free-layer damping (FLD) and constrained layer damping (CLD) treatments.

The FLD configuration (Fig. 5.1a) is formed by a layer of viscoelastic material laid directly onto the surface of the host structure. The dissipation effect provided by this configuration relies on the continuous and cyclic extensional deformation of the viscoelastic layer as a result from the flexural motion that the adjacent structure undergoes during vibration.

The CLD treatment (Fig. 5.1b) uses a viscoelastic layer, also laid onto the host structure surface, covered by a stiff and thin constraining layer. The presence of the constraining layer promotes the shear deformation of the viscoelastic layer during the flexural deformation of the host structure. This configuration is effective even when using very thin viscoelastic layers, whereas the FLD treatments require the use of relatively thick layers to attain valuable levels of structural damping.

## 5.5 Viscoelastic Materials: Properties and Characterization

Viscoelastic materials present a viscous-elastic behaviour evidencing a high dependence upon the room temperature and frequency of the applied cyclic load. Therefore, it is important to correctly characterize these materials in order to select and tailor the material according to the conditions of the treatment application.



**Fig. 5.1** Viscoelastic damping treatment configurations **a** free layer damping (FLD) **b** constrained layer damping (CLD)

The viscoelastic nature of these materials is evidenced by a phase lag between a cyclic harmonic load imposed to a VEM specimen and the resulting strain response. This relation can be described by the complex modulus approach, where the extensional or shear modulus, relating the stress and strain fields during uniaxial harmonic deformation of a VEM sample, is a complex entity as:

$$E^*(\omega, T) = E'(\omega, T) + jE''(\omega, T) \quad (5.1)$$

where  $E'(\omega, T)$  and  $E''(\omega, T)$  represent, respectively, the real component or storage modulus and the imaginary part or loss modulus. The ratio between the imaginary and the real parts represents the damping capability of the VEM and is designated as loss factor ( $\eta$ ). As indicated, both components of the complex modulus are frequency and temperature dependent.

The practical characterization procedure for VEMs uses an experimental setup where a sample of material is subjected to a harmonic excitation and both load and response are measured and correlated to obtain the complex modulus data. These experimental characterization procedures can be performed on beams with special FLD or CLD configurations, as indicated by the related standard ASTM E756-98 [1]. Other experimental setups, excitation methods and response measurement techniques, as well as sample configurations, can also be used. Direct methods, such as Dynamic Mechanical Thermal Analysis (DMTA), impose directly a harmonic load to a small sample of material using a tension, compression, bending or shear deformation configuration. DMTA is usually performed in special testing devices equipped with a small thermal chamber with controlled temperature, and complex modulus data is obtained for predefined pairs of temperatures and excitation frequencies.

The characterization of viscoelastic materials over a large range of temperatures and frequencies results in a large set of data. Fortunately, simple rheologic VEMs evidence a frequency–temperature correlation called frequency temperature superposition principle—a principle stating that the effect of a frequency variation onto the modulus of the VEM is inversely proportional to the effect promoted by a well-defined variation in temperature. This superposition principle is often applied to reduce, extrapolate and represent the complex modulus data. The nomogram representation [2] takes advantage of this superposition effect in order to provide a

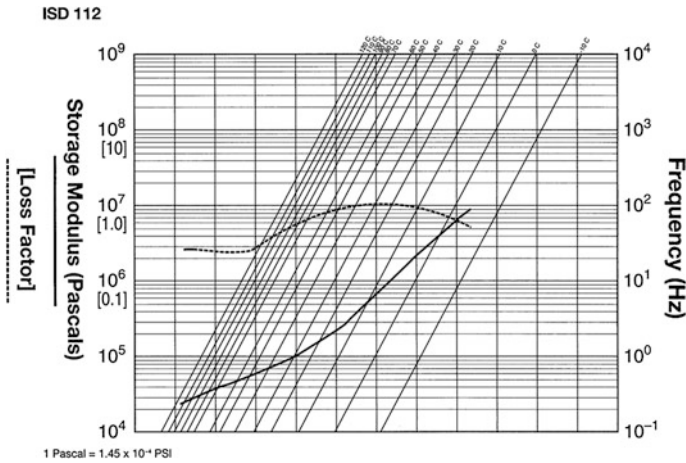


Fig. 5.2 Nomogram of 3 M ISD112 [3]

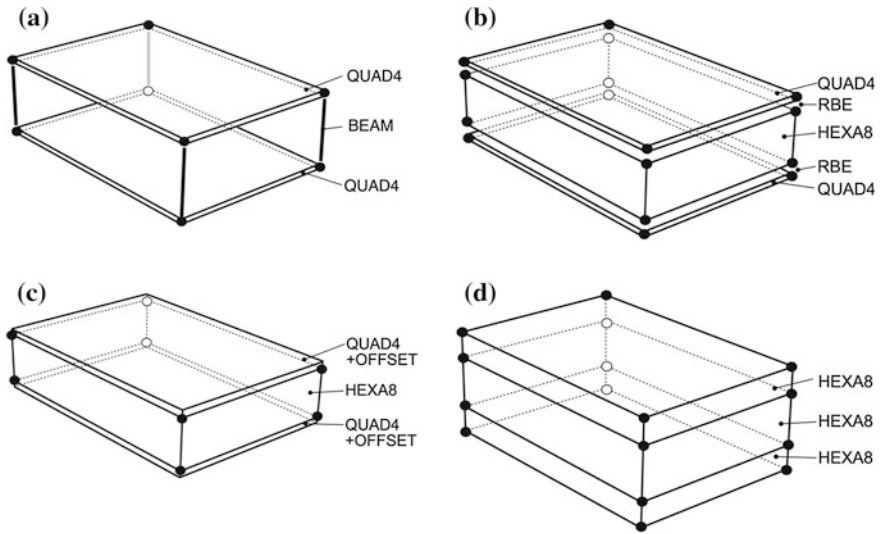
simple and expeditious representation of the complex modulus data for several isothermal conditions over a large range of frequencies. Figure 5.2 depicts the nomogram for a commercially available VEM, the 3 M ISD 112 [3].

## 5.6 Numerical Simulation of Viscoelastic Damping Treatments

Despite the simplicity associated to the damping mechanism and application of VEM treatments, the correct design and simulation of these damping treatments is not simple and requires special workarounds. The prime issues related to this design difficulty are: the correct representation of the strain field imposed to the VEM layer, the proper modelling of the VEM properties and the use of the adequate solution method able to deal with the selected viscoelastic constitutive model and to provide the solution in the desired domain.

### 5.6.1 Spatial Model of a VEM Damping Treatment

Contrary to the FLD configuration, which can be easily represented by an Equivalent Single Layer (ESL) approach or a model based on the Classic Laminate Plate Theory (CLPT), the CLD treatment requires a good representation of the shear strain pattern induced inside the VEM layer. In fact, the damping mechanism developed inside the VEM layer is proportional to its shear strain field, and therefore it is important to select a spatial model able to represent it accurately.



**Fig. 5.3** Layered models: **a** beam-plate, **b** brick-plate with rigid link, **c** brick-plate with offset, **d** layered brick

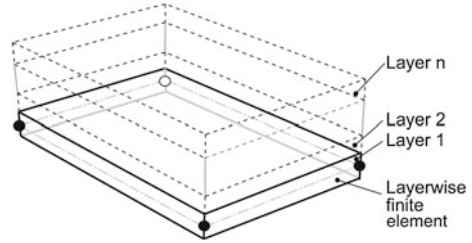
Literature review indicates several approaches that can be divided into: layered models and discrete layer models.

Layered (or combined) models are commonly applied when commercial finite element software is used. This approach uses a spatial model obtained as a layered scheme of standard finite elements, such as beams, plates and hexahedral elements. Figure 5.3 illustrates four modelling approaches using a layered combination of standard finite elements.

The beam-plate model, initially proposed by Killian and Lu [4], uses four beams to represent the shear and compression behaviour of the VEM layer, while two plates (QUAD4) sharing the nodal positions of the beams represent the host structure and the constraining layer. The second (Fig. 5.3b) and third (Fig. 5.3c) models also apply two plates to represent the outer stiff skins, the host structure and the constraining layer. However, they use a hexahedral/brick finite element (HEXA8) to represent the VEM layer. The difference between the two models relies on the mechanism use to connect the degrees-of-freedom of the brick of the core to the adjacent external plates. While the first uses rigid link connections (RBE) to interconnect the degrees-of-freedom of the adjacent nodes, the former modifies the nodal location of the outer plates by introducing an offset constrain into the plate formulation.

The last model uses three layers of bricks sharing nodal locations to represent the entire set of physical layers of the CLD treatment. It is an expeditious approach able to be used in any finite element package but requires spatial attention to avoid numerical locking duo to the aspect ratio of bricks of the stiff outer layers [5].

**Fig. 5.4** Discrete layer model



Despite the straightforward use of these layered models, it requires a time-consuming and laborious modelling task, especially when the structure has a complex and three-dimensional shape. Furthermore, this modelling approach also presents significant limitations when used inside an optimization procedure requiring continuous update of the spatial model of the damped structure. To overcome this difficulty, discrete layer models have been proposed and successfully applied. These discrete layer models rely on a partial [6] or full layerwise theory [7], and have the benefit of a simple and general spatial model with a single set of four nodes (Fig. 5.4), while the entire definition of the out-plane geometry is externally defined by a tabular data file.

A simple, yet efficient, discrete layer model can be constructed using a partial layerwise formulation where each layer is represented by a Mindlin or first-order plate formulation [6, 8]. The displacement field is defined ensuring the continuity between the layers and mixed formulations can also be employed to impose additionally the continuity of stresses at the layers' interfaces [7, 9].

### 5.6.2 Viscoelastic Constitutive Model

The definition of a reliable, accurate and cost-efficient constitutive model for VEMs is still an open challenge for both experimental and numerical researchers. While the experimental characterization procedure requires a simple, yet representative, constitutive model able to describe the material properties with a limited set of data parameters, a numerical simulation calls for a simple model able to be used in a standard manner inside a numerical algorithm.

VEMs present a significant dependency upon temperature and frequency. The temperature dependency is usually disregarded from the usual VEM constitutive models since isothermal conditions are assumed. Nevertheless, even for structures working at specific and constant temperature conditions, the temperature parameter can interfere in the material properties. In fact, the damping obtained from VEM treatments is a result of a thermodynamic mechanism related to an energy transfer process—vibration energy is transferred to deformation energy and subsequently to thermal energy radiated to the surrounding medium. Therefore, the efficiency of the damping mechanism and the maintenance of isothermal conditions inside the VEM

layer are strongly conditioned by the way that generated heat is rapidly dissipated; otherwise, temperature may change modifying significantly the material properties. This temperature dependency is often included into frequency dependent models using the frequency–temperature superposition principle [10–12].

### 5.6.2.1 Complex Modulus

The complex modulus approach is a simple and accurate model, but requires the storage and direct use of a long tabular data file comprising the storage and loss modulus values defined or measure for a long set of frequency–temperature pairs. The superposition principle is often used to extrapolate and interpolate data.

Considering a generalized damped system, its equation of motion in time domain can be written as:

$$[M]\{\ddot{x}(t)\} + [C]\{\dot{x}(t)\} + [\bar{K}]\{x(t)\} = \{f(t)\} \quad (5.2)$$

where matrices  $[M]$ ,  $[C]$  and  $[\bar{K}]$  represent, respectively the mass, viscous damping and complex stiffness matrices of the dynamic system.

Complex stiffness matrix  $[\bar{K}]$  is subdivided into a real matrix characterizing the stiffness of the elastic elements of the sandwich, i.e. the stiffness of host structure and constraining layer, and a complex matrix  $[\bar{K}_v(\omega)]$  representing the stiffness of the VEM layer. This complex stiffness matrix includes terms related to extensional stiffness and terms associated to the transverse shear component, as:

$$[\bar{K}_v(\omega)] = \bar{E}_v(\omega) \frac{K_e(E_0)}{E_0} + \bar{G}_v(\omega) \frac{K_s(G_0)}{G_0} \quad (5.3)$$

where  $K_e(E_0)$  and  $K_s(G_0)$  represent, respectively, the extensional stiffness matrix calculated for a predefined real extensional modulus  $E_0$  and the transverse shear stiffness matrix calculated for a predefined real shear modulus  $G_0$ .

Assuming stationary harmonic motion, where both excitation and response are harmonic functions, the equations of motion can be rewritten as:

$$(-\omega^2[M] + j\omega[C] + \bar{K}(\omega))\{\bar{X}(\omega)\} = \{F(\omega)\} \quad (5.4)$$

### 5.6.2.2 Golla-Hughes-McTavish Model

The Golla-Hughes-McTavish (GHM) [13] model introduces the frequency dependency through a series of mini-oscillators. The shear modulus of a VEM is thus defined, in Laplace domain, as:

$$\bar{G}(s) = G^0 \left( 1 + \sum_i \alpha_i \frac{s^2 + 2\zeta_i \omega_i s}{s^2 + 2\zeta_i \omega_i s + \omega_i^2} \right) \quad (5.5)$$

where  $G^0$  is the low frequency shear storage modulus and  $\alpha_i$ ,  $\zeta_i$  and  $\omega_i$  are model parameters—real positive constants—that define each mini-oscillator  $i$ . The same model can be applied to the extensional modulus  $E(s)$ , which approximately shares the same model parameters, i.e. the two low frequency shear and extensional storage moduli can be related using a constant and frequency independent Poisson's ratio. This equivalence between both moduli is not true [14, 15] but can be used as a simplification without significant error.

The model is introduced into the equations of motion written in Laplace domain as:

$$(s^2[M] + s[C] + \bar{K}(s))\{\bar{X}(s)\} = \{F(s)\} \quad (5.6)$$

where  $\bar{X}(s)$  and  $F(s)$  represent the Laplace transform of the response and load vectors, respectively. Complex stiffness matrix  $\bar{K}(s)$  can be divided into the elastic stiffness matrix of the outer layers and the viscoelastic stiffness matrix  $[\bar{K}_v(s)]$  that represents the stiffness of the VEM core. This last matrix can be decomposed as described in Eq. (5.3) where  $E_v(s)$  and  $G_v(s)$  are replaced by the GHM model described in Eq. (5.5).

### 5.6.2.3 Anelastic Displacement Fields Model

The Anelastic Displacement Fields (ADF) model, initially proposed by Lesieutre and co-workers [10, 16], represents the shear modulus of a VEM as:

$$\bar{G}(s) = G^0 \left( 1 + \sum_i \Delta_i \frac{\omega^2 + j\omega\Omega_i}{\omega^2 + \Omega_i^2} \right) \quad (5.7)$$

where  $\Delta_i$  and  $\Omega_i$  represent the model parameters.

Following the same procedure presented for the GHM model, the ADF model can be introduced in the equations of motion.

### 5.6.2.4 Fractional Derivative Model

Initially proposed by Bagley and Torvik [17, 18], the fractional derivative model relates the stress and strain fields through the relation:

$$\sigma(t) + \sum b_i \mathcal{D}^{\beta_i} \sigma(t) = E^0 \varepsilon(t) + \sum a_j \mathcal{D}^{\alpha_j} \varepsilon(t) \quad (5.8)$$

where  $a_j$ ,  $b_i$ ,  $\alpha_j$  and  $\beta_i$  are the fractional derivative model parameters.

When using a single series of parameters it is possible to define several simple yet accurate fractional derivative models, such as the 5-parameter model proposed by Bagley and Torvik [17]:



$$\bar{G}(s) = G^0 \left[ \frac{1 + as^\alpha}{1 + bs^\beta} \right] \quad (5.9)$$

written in Laplace domain, and the four and seven parameters presented by Schmidt and Gaul [19–21].

### 5.6.3 Analysis Method

The analysis method not only depends on the desired solution domain but the selected constitutive model also dictates it. In fact, this last relation is so imperative that often literature does not dissociate the analysis method from the VEM model. Several methods are available and the commonly used ones are presented in what follows.

#### 5.6.3.1 Direct Frequency Analysis

The complex modulus approach is a frequency domain model and therefore it can be straightforwardly used by a frequency domain analysis method such as the Direct Frequency Analysis (DFA). The DFA can be regarded as a discrete frequency analysis method where the equation of motion written in the frequency domain is solved for each frequency step. If isothermal conditions are assumed, then frequency dependency of the VEM can be considered since the stiffness matrix of the VEM layer is updated for every frequency step. This solution method can be directly performed using several commercial software packages available in the market, where the complex modulus is simply introduced into the code as a tabular data file comprising the real and imaginary parts of the frequency dependent complex modulus of the selected VEM.

#### 5.6.3.2 Numerical Integration in Time Domain

Parametric models described in time domain, like the GHM and ADF, can be introduced in the equation of motion written in the time domain. The resulting equation can be reorganized to permit a straightforward resolution by using a numerical integration scheme, but this reorganization increases the dimension of the problem; at least it doubles the system of equations dimension if a single series of parameters is used in the model.

The fractional derivative model can also be introduced in a time domain analysis applying directly a numerical integration procedure when using the Grünwald definition [22], as suggested by Schmidt and Gaul [19–21] and Enelund [23].

### 5.6.3.3 Complex Mode Analysis

Assuming a free response model of a dynamic system with viscoelastic damping, its complex eigenvalue problem is described by:

$$[\bar{K}(\omega)]\{\bar{\phi}\}_r = \bar{\lambda}_r^2[M]\{\bar{\phi}\}_r \quad (5.10)$$

The real part of the complex eigenvalue  $\bar{\lambda}_r$  corresponds to the natural frequency of the damped system  $\omega_r$  and the ratio between the imaginary and real parts determine the modal loss factor  $\eta_r$ .

This analysis method can be applied directly onto the equation of motion written in frequency domain, allowing the use of the complex modulus approach to introduce the VEM modulus. However, this procedure assumes a constant storage modulus and loss factor within the frequency band used in the analysis [24]. Some solutions to this limitation have been proposed, like the procedure suggested by Kung and Singh [25] where each mode is analyzed separately allowing the use of a constant but more realistic modulus, which is defined for the narrow frequency band around the selected mode. This solution can also use an iterative process where the stiffness matrix of the VEM layer is updated continuously following the real complex modulus value for the calculated natural frequency.

### 5.6.3.4 Modal Strain Energy Method

The Modal Strain Energy (MSE) method, as originally proposed by Johnson and Kienholz [26], determines the loss factor of a damped structure based on a strain energy contribution analysis. The method assumes that the real modes of the undamped structure can represent the modes of the damped structure. This assumption can only be considered for low damping structures, otherwise, the modes for the damped structure may differ significantly from the modes calculated for the real part of the complex stiffness matrix.

If the eigenvalue problem defined in Eq. (5.10) is written as:

$$[K_R(\omega_r)]\{\phi\}_r = \omega_r^2[M]\phi_r \quad (5.11)$$

where  $K_R$  is the real component of the complex stiffness matrix of the entire dynamic system and  $\phi_r$  is the real mode shape for natural frequency  $\omega_r$ .

Equation 5.10 can be algebraically arranged as:

$$\eta_r = \frac{\{\phi_r^T\}[K_I]\{\phi_r\}}{\{\phi_r^T\}[K_R]\{\phi_r\}} \quad (5.12)$$

where  $K_I$  is the imaginary part of the complex stiffness matrix, and real modes  $\phi_r$  are used replacing the complex ones.

Numerically, the MSE method is introduced in a finite element analysis as a strain energy analysis defined by:

$$\eta_r = \frac{\sum_k \eta_{kr} \Pi_{kr}}{\sum_k \Pi_{kr}} \quad (5.13)$$

where the sum is performed over the entire set of finite elements  $k$ , and  $\Pi_{kr}$  is the strain energy calculated at an individual finite element  $k$  for a selected mode  $r$ .

The method is easily used in a commercial finite element package but is limited to lightly damped structures. Another drawback of the MSE method is related to how the representative mode shapes are calculated. For CLD damping treatments, the real eigenvalue problem is usually solved considering merely the host structure. If the damping treatment does not introduce an important modification onto the mass and stiffness of the complete structure, the approximation provide by the analysis is satisfactory. However, when using this method for the analysis of symmetric CLDs with high damping levels, the VEM layer alters significantly the stiffness of the entire composite due to the decoupling effect promoted by the soft core. An iterative procedure can then be used to solve this issue, where the VEM properties are iteratively updated according to the calculated natural frequency and the real eigenvalue problem is performed considering the real part of the updated stiffness matrix for the entire sandwich structure.

Other analysis methods were proposed and used with success, like the perturbation method [27, 28], where the complex eigenvalue problem solution is estimated by a stiffness and mass perturbation to the initial set of undamped natural modes of the pristine structure, and the modal projection method [29, 30], where the equations of motion of the system are projected onto a truncated and modified modal base to reduce the problem dimension.

## 5.7 Optimization of Viscoelastic Damping Treatments

The design of structures is often ruled by a set of constraints following an intended goal. Low cost and high efficiency lead the list of desired goals of an optimized design. Optimization procedures are often applied during the design of structures to attain the required qualities while reducing the cost, weight or maintenance needs. The design of viscoelastic damping treatments is also a task where the use of optimization techniques may provide a significant benefit, enabling the definition of an efficient configuration of the treatment with lower weight and cost. Low mass and treatment cost are usually regarded as two important characteristics for lightweight structures where such damping treatments are the primary choice among the available vibration control strategies.

The optimization of FLD treatments is mainly established by a redesign of the coverage area and treatment location. This treatment configuration requires the use of thick VEM layers and thus the reduction of the coverage area along with a proper location of the treatment patches, while maintaining its effectiveness for the selected frequency range or target natural modes, represents a significant and valuable effort to reduce the cost of the damping treatment and specially the additional mass.

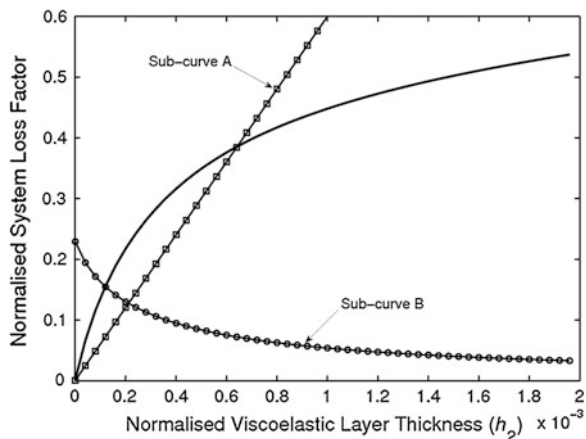
The constrained configurations, CLD treatments, can also be optimized seeking for optimum locations and area coverage of the damping patches. Additionally, since the damping mechanism of this treatment configuration relies on the shear strain that the host structure imposes to the thin VEM layers, some exploration can be performed towards the definition of a set of geometric parameters that maximize such shear deformation energy. Experimental and numerical studies [5] reveal that the symmetric configuration of a CLD treatment, also designated as integrated layer damping (ILD), represents the most effective configuration when the total thickness of the whole structure is maintained constant. In this configuration both the constraining layer and host structure have the same thickness and the VEM layer is thus located on the neutral plate of the structure where the shear strain presents its maximum value. Further, multiple layer configurations [31] can also be investigated in order to promote the shear deformation inside the VEM layers. In these multiple layer configurations the VEM layer is subdivided into several thinner layers separated by thin constraining layers. This configuration provides the ability to introduce VEM materials with different transition temperatures, thus enlarging the effective temperature range of the treatment. It also improves the treatment efficiency since thinner VEM layers can offer higher shear deformations levels.

Contrary to the FLD configuration, where a direct relation between the VEM layer thickness and the damping efficiency of the treatment exists, the CLD configuration does not present such direct relation. It is true that the damping efficiency of the CLD treatment is proportional to the shear deformation energy stored inside the VEM layer during the vibration of the host structure. However, this energy depends on the volume of material contributing for this energy storage and also on the shear strain level imposed to this layer of material. Unfortunately, while the relation between the VEM layer thickness and the shear strain level is opposite, i.e. thinner VEM layers are subjected to higher shear strain for the same deformation of the host structure, the volume of VEM material contributing for the total strain energy is directly proportional to the thickness of the layer. This opposing relation can also be explored for optimization purposes.

A recent study [32] demonstrates that this opposing relation may be used to design very thin and efficient damping treatments for relatively thick host beams, since for these conditions the relation between the treatment efficiency and the VEM layer thickness is not described by the conventional monotonic and asymptotic curve given by the books on this subject, but rather presents a peak. This interesting feature is explained by the graphical representation of the two counterparts of the shear energy stored inside the VEM layer—the layer volume (Fig. 5.5—subcurve A) and the shear strain (Fig. 5.5—subcurve B)—whose combination in the form of strain energy may evidence a peak (Fig. 5.4). The main conclusions gathered from this study are enumerated as:

- this feature only occurs for relatively thick host beams in comparison to its length;
- contrary to the common sense, thicker beams correspond to thinner optimum VEM layers—peaks move to lower VEM layer thicknesses;

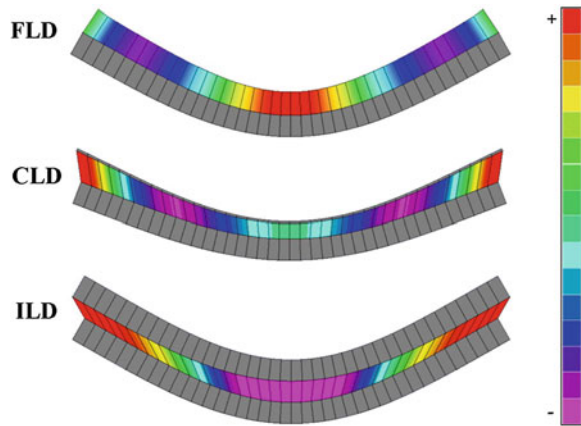
**Fig. 5.5** Normalized efficiency curve for relatively thick beams



- thicker constraining layers evidence pronounced peaks of the efficiency curve and correspond to thinner optimum VEM layers;
- higher modulus ratios, i.e. the ratio between the host structure modulus and the VEM storage modulus, lead to thinner optimum VEM layers;
- mode shapes and boundary conditions leading to more complex and border restricted displacement fields of the host structure, which correspond to more complex shear strain fields of the VEM layer, move the efficiency peaks to lower values of the VEM layer thickness.

Another controversial issue related to the optimization of CLD treatments is related to the best location of the damping patches. While some authors state that the best location for the damping treatment is near the antinodes of the vibrating structure, corresponding to the location with highest out-plane displacement, others indicate the nodes as the location where shear strain reaches its maximum value and thus correspond to the best treatment spots. Experimental observations and numerical results reveal that both situations occur for different configurations of the CLD treatment. As revealed by the images depicted in Fig. 5.6, while the FLD configuration presents the maximum strain deformation energy near the antinodes of the first flexural mode, this maximum location is located near the nodes for the symmetric CLD treatment (or ILD configuration). The non-symmetric CLD configuration can be regarded as a damping treatment in between the FLD and ILD configurations, i.e. despite being a damping treatment mostly ruled by the shear strain imposed to the VEM core of the laminate, the best locations for the damping patches depend on the relative thickness of the constraining layer in respect to the host structure. For thick constraining layers, close to the symmetric configuration, the best treatment spots are located near the nodes; but for thinner constraining layers these locations move towards the antinodes regions. Additionally, it is also observed that the treatment borders, as well as cross sectional cuts deliberated made to the VEM layer, promote high shear strain locations and, when correctly located, can provide significant improvements to the treatment efficiency.

**Fig. 5.6** Strain energy distribution inside the VEM layer for FLD, CLD and ILD (free–free beam—first mode)



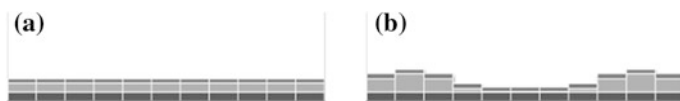
The use of spacers intercalated between the host structure and the VEM layer is another optimization strategy to improve the damping efficiency of FLD and CLD treatments [33]. The spacer, usually a light and stiff substructure, amplifies the bending deformation of the host structure, increasing thus the shear strain imposed to the adjacent VEM layer.

As evidenced in the above lines, the optimization of VEM damping treatments can explore several approaches, including materials, relative thickness ratios, coverage area, treatment location and even additional elements, such as the spacers, or special enhancement strategies, like the cuts in selected locations of CLD.

The manipulation and selection of the best combination of these intervening variables is not straightforward and requires the use of optimization strategies. The Genetic Algorithm (GA) method and Topological Optimization are two available optimization strategies with potential use for the design of highly effective damping treatments based on VEMs.

GA methods are based on an evolutionary scheme where a family of initial and random treatment configurations is evaluated and combined according to the fitness of each individual of the family to create a new and improved set of offspring. The optimization procedure ends with a set of offspring with the best performance within those evaluated inside the selected search domain.

Topological optimization procedures initiate with a predefined treatment configuration, which is evaluated in terms of its damping efficiency. The contribution of each subdomain—if the finite element method is used the subdomain corresponds to the domain of each finite element—is determined, providing direct information on the importance of the subdomain for the damping efficiency of the entire structure. According to this contribution, the VEM is redistributed, removed or introduced depending on the topologic strategy. Figure 5.7 illustrates the initial and final configurations of a Topological optimization analysis of a CLD treatment applied to a host beam with free–free boundary conditions; this optimized VEM layer configuration was obtained considering the first natural mode.



**Fig. 5.7** Topological optimization of a free-free beam for the first mode **a** initial configuration **b** final configuration

## 5.8 Recent Trends and Future Directions

Active control mechanisms domain the majority of recently publish studies on the vibration control theme. This huge interest results from the recent advances in materials, especially those with piezoelectric effect or shape-memory characteristics, compact and light sensors and controllers using MEMS technology, and new and efficient control strategies. Nevertheless, passive-damping treatments, especially those based on VEMs, maintain a strong and irreducible position among the vibration control strategies for light and large structures, from aerospace structures to common household appliances.

Viscoelastic damping treatments provide a valuable and interesting solution for vibration control of light structures simply because they are identically light, do not introduce a significant modification to the host structure and can perform well within a narrow frequency and temperature ranges. Furthermore, its efficiency does not depend on complex control networks and power sources, and therefore is immune to the potential fails that may affect the active devices.

Since the efficiency of VEM-based damping treatments depends on the amount of vibration energy transferred to the VEM layer as stored strain energy, this vibration control mechanism can only perform satisfactory when the strain field imposed to the VEM layer is effectively able to transfer an important portion of the vibration energy. This energy storage mechanism does not depend only on the geometric parameters of the damping treatment but is also dictated by the shape of the vibration mode and how it deforms the VEM layer, i.e. the more complex is the mode shape higher will be the damping efficiency. This is the main reason why passive damping treatments based on VEMs are usually regarded as a good solution for medium/high frequency ranges while, for the low frequency range, active control devices perform better. The contrary also applies—active control in the high frequency range requires fast control devices and reliable control strategies, which may limit the efficient use of this damping solution for high frequency applications.

Due to simplicity of the treatment application and the increasing level of know-how on these solutions, the use of viscoelastic damping treatments is gaining more interest in the structural engineering, from high-tech applications to common day-to-day appliances.

One of the key issues impelling the use of these passive damping solutions is the current development of effective design tools based on numerical methods. Tailored finite elements along with simple and representative constitutive models

able to describe accurately the behavior of VEMs is a constant contribution of authors from the four corners of the world in recent technical journal issues. Generalized models able to describe accurately the displacement, strain and stress fields of single or multiple viscoelastic damping treatments is currently one of the most interesting developments and the scope of the papers being published in this particular subject. Recent contributions are divided between those models based on full layerwise theories, able to describe completely the strain field of the sandwich, and those based on equivalent layer theories for general use in the design of large damped structures. Further, mixed formulations working directly onto the stress and strain fields are able to impose directly the required stress and strain interface conditions.

The research for effective constitutive models able to easily represent the real behavior of VEMs is still an open interest. These models are required to represent accurately the properties of the VEM while providing a straight and cost-effective numerical implementation. The ability to introduce the temperature dependency is also a required feature, especially when analyzing critical structures subjected to important and variable thermal gradients (like satellite panels or aeronautic fuselages) or when the energy dissipation mechanism is not efficient enough to avoid the temperature modification inside the damping treatment (an issue for damping treatments applied onto polymer composite or ceramic host structures).

Recently, the huge interest on micro-size and nano-size reinforcements for composites and polymers has been driving the attention of the researchers for the development of more efficient damping solutions, taking advantage of the associated potential benefits. The application of fibers inside the VEM layer [34] provides a way to improve the damping efficiency of FLDs, since the fibers promote the localized shear deformation of the VEM. Following this idea, the inclusion of nano-size reinforcements such as carbon nanotubes (CNTs) inside a soft VEM material seems to be a promising solution for the development of oriented and efficient damping treatments. These modified VEMs would be applied as thin FLD solutions providing efficiency levels similar or even higher than those obtained from CLD configurations, while taking advantage of all the design and application benefits of the FLD configuration.

The viscoelastic-based damping treatments have been successfully applied since the 1960s. Currently, this vibration control strategy is just around the corner and fits inside almost every critical and light structure, from airplanes and cars to simple washing machines. Despite the constant research done to this damping solution, there is still a lot of work to be done, specially on efficient simulation and optimization methods, material characterization and constitutive models, and material formulations.



## References

1. ASTM (1998) E756-98 standard test method for measuring vibration-damping properties of materials. In: Annual book of ASTM standards, vol 04.06. ASTM, New York
2. Jones DIG (2001) Handbook of viscoelastic vibration damping, 1st edn. Wiley, New York
3. 3M (1993) ScotchDamp TM vibration control systems. 3M Industrial Specialties Division, St Paul, Tarsus
4. Killian JW, Lu YP (1984) A finite element modeling approximation for damping material used in constrained damped structures. *J Sound Vib* 97(2):352–354
5. Moreira R, Rodrigues JD (2004) Constrained damping layer treatments: finite element modeling. *J Vib Control* 10(4):575–595
6. Moreira RAS, Rodrigues JD, Ferreira AJM (2006) A generalized layerwise finite element for multi-layer damping treatments. *Comp Mech* 37(5):426–444
7. Reddy JN (1997) Mechanics of laminated composite plates: theory and analysis. CRC Press, Boca Raton
8. Moreira RAS, Rodrigues JD (2006) A layerwise model for thin soft core sandwich plates. *Comp Struct* 84(19–20):1256–1263
9. Moreira RAS, Melo FJ, Rodrigues JD (2010) Static and dynamic characterization of cork compounds for sandwich beam cores. *J Mater Sci* 45(12):3350–3366
10. Lesieutre GA, Govindswamy K (1996) Finite element modeling of frequency-dependent and temperature-dependent dynamic behavior of viscoelastic materials in simple shear. *Int J Solids Struct* 33(3):419–432
11. Gaul L (1989) Structural damping in frequency and time domain, pp 177–185. In: 7th international modal analysis conference (IMAC VII), Las Vegas
12. Moreira RAS, Corte-Real JD, Rodrigues JD (2010) A generalized frequency–temperature viscoelastic model. *Shock Vib* 17(4–5):407–418
13. McTavish DJ, Hughes PC (1993) Modeling of linear viscoelastic space structures. *J Vib Acoustics* 115:103–110
14. Pritz T (1998) Frequency dependences of complex moduli and complex Poisson's ratio of real solid materials. *J Sound Vib* 214(1):83–104
15. Muravyov A (1998) Forced vibration responses of a viscoelastic structure. *J Sound Vib* 218(5):892–907
16. Lesieutre GA, Bianchini E (1995) Time domain modeling of linear viscoelastic using anelastic displacement fields. *J Vib Acoustics* 117:424–430
17. Bagley RL, Torvik PJ (1983) Fractional calculus—a different approach to the analysis of viscoelastically damped structures. *AIAA J* 21(5):741–748
18. Bagley RL, Torvik PJ (1985) Fractional calculus in the transient analysis of viscoelastically damped structures. *AIAA J* 23(6):918–925
19. Schmidt A, Gaul L (2001) FE implementation of viscoelastic constitutive stress–strain relations involving fractional time derivatives, pp 79–89. In: 2nd European conference on constitutive models for rubber, Hannover, 10–12, Sept 2001
20. Schmidt A, Gaul L (2002) Application of fractional calculus to viscoelastically damped structures in the finite element method, pp 297–306. In: International conference on structural dynamics modelling, Madeira, 3–5, June 2002
21. Schmidt A, Gaul L (2002) Parameter identification and FE implementation of a viscoelastic constitutive equation using fractional derivatives. *Applied Mathematics and Mechanics* 1(1):153–154
22. Oldham KB, Spanier J (1974) The fractional Calculus. Academic Press, NY
23. Enelund M (1996) Fractional calculus and linear viscoelasticity in structural dynamics. PhD thesis, Chalmers University of Technology. Goteborg
24. Cupial P, Niziol J (1995) Vibration and damping analysis of a three-layered composite plate with a viscoelastic mid-layer. *J Sound Vib* 183(1):99–114

25. Kung SW, Singh R (1998) Complex eigensolutions of rectangular plates with damping patches. *J Sound Vib* 216(1):1–28
26. Johnson CD, Kienholz DA (1982) Finite element prediction of damping structures with constrained viscoelastic layers. *AIAA J* 20(9):1284–1290
27. Kelly WJ, Stevens KK (1989) Application of perturbation techniques to the modal analysis of a shaft with added viscoelastic damping, 1516–1520. In: 7th international modal analysis conference (IMAC VII), Las Vegas
28. Lin RM, Lim MK (1996) Complex eigensensitivity-based characterization of structures with viscoelastic damping. *JASME* 100(5):3182–3191
29. Balmès E (1996) Super-element representation of a model with frequency dependent properties. *International Seminar Modal Analysis, Leuven, Belgium*, pp 1767–1778
30. Balmès E (1997) Model reduction for systems with frequency dependent damping properties, pp 223–229. In: 15th International modal analysis conference (IMAC XV), Orlando
31. Moreira RAS, Dias Rodrigues J (2010) Multilayer damping treatments: Modeling and experimental assessment. *J Sandwich Struct Mater* 12(2):181–198
32. Sher BR, Moreira RAS (2013) Dimensionless analysis of constrained damping treatments. *Compos Struct* 99:241–254
33. Van Vuure AW, Verpoest I, Ko FK (2001) Sandwich-fabric panels as spacers in a constrained layer structural damping application. *Compos B* 32:11–19
34. Alberts TE, Xia H (1995) Design and analysis of fiber enhanced viscoelastic damping polymers. *J Vib Acoust* 117:398–404

# Chapter 6

## Thermomechanical Modeling of Multiphase Steels: Classical and Modern Engineering Analyses

A. Andrade-Campos, P. Vasconcelos, J. F. Caseiro and J. A. Oliveira

**Abstract** This chapter presents the thermodynamic foundations for the analysis of multiphase steels and some conventional approaches used in the analysis of these materials, including classical constitutive equations. Afterwards, the standard strategies for modeling multiphase metals are introduced as well as some examples. The new assumptions and today's modeling strategies, that include complex numerical simulation methods such as the finite element method, finite volume, finite differences and phase-field methods, are also presented and compared with the previous approaches, with the aid of examples and research results.

### 6.1 Introduction

Nowadays, steel plays a major role in our lives. It is one of the most important, versatile and adaptable material available. Although other materials replaced steel in many applications, it still remains very successful and cost-effective, with millions of tons consumed throughout the world. The research associated with these metal alloys still provides very interesting and stimulating discoveries such as, for example, multiphase steels (TRIP, dual-phase steels, etc.), which have gained a relevant role due to their complex microstructure and properties. Multiphase steels have excellent mechanical properties, combining good formability properties with high strength and have become an important construction material, mainly in the automotive industry. Additionally, with the improvements in the heat treatment process during production, phase distribution in these alloys can be adapted resulting in different mechanical behavior. In multiphase steels, heat treatments are usually applied to achieve a desired metallurgical composition in order to obtain the expected

---

A. Andrade-Campos (✉) · P. Vasconcelos · J. F. Caseiro · J. A. Oliveira  
Department of Mechanical Engineering, Centre for Mechanical Technology and  
Automation, GRIDS Research Group University of Aveiro, Campus Universitário de  
Santiago 3810-193 Aveiro, Portugal  
e-mail: gilac@ua.pt

mechanical properties. The metallurgical composition of these materials, including the quantity and distribution of each phase, is achieved during the manufacturing process and depends directly on the thermal path undergone by the material.

A special case of multiphase steels are the dual-phase (DP) steels which contain low carbon phase and exhibit increased strength but inferior ductility and formability when compared with plain carbon steels. The microstructure, consisting in hard martensite particles in a softer ferrite matrix, is obtained by the intercritical heat treatment of low carbon steels. The rule of mixtures and other classical analytical techniques have been employed in the past to determine the behavior of this kind of materials. However, due to the complexity of the deformation process, such techniques could not realistically represent the material behavior in terms of stress–strain tendency and mechanical deformation. In order to predict the multiphase steel material behavior, different multiphase model approaches can be used. These approaches can be classified as classical or modern engineering analyses. The classical approaches consider the material as homogenous and use empirical knowledge to reproduce the behavior of the material. The modern engineering analyses use computational resources to calculate the thermomechanical and metallurgical behavior of the material, not only at the macroscopic scale but also at the micro- and nano-scales. These last approaches rely on multiscale numerical material models to obtain accurate results. These models must contain information concerning the evolution of the metallurgical, thermal and mechanical state of the material. In recent years, the developments in numerical modeling in material science and engineering also enabled the simulation of phase changes using, for example, the phase-field method which allows to have a quantitative perspective of the phase transformation.

In this chapter, initially, a brief review on phase transformation in metals and alloys including metallurgy thermodynamics (as well as solidification, nucleation and growth, diffusion and other solid-state transformations) is presented. Then, a brief study of phase transformation in steels is conducted. Afterwards, classical and modern engineering approaches, based on numerical analyses, are presented and compared. This chapter ends with some considerations regarding the future of thermomechanical models for steels.

## **6.2 Phase Transformation and Microstructure in Metals and Alloys**

### ***6.2.1 Thermodynamics of Phase Transformations***

Phase transformation is a key subject to understand microstructure evolution and, therefore, properties in metals. Phase transformations<sup>1</sup> can be divided into three categories [1]:

---

<sup>1</sup> The changes of the microstructure in metals and alloys.

- diffusion-dependent transformations, in which there is no change in the phase composition or in the number of phases present. These include solidification of a pure metal, allotropic transformations, and recrystallization and grain growth;
- diffusion-dependent transformation, with changes in the phase compositions and in the number of phases presented; the final microstructure ordinarily consists of two phases;
- diffusionless solid-state phase transformations do not require long-range diffusion during the phase change; only small atomic movements (over usually less than the interatomic distances) are needed.

Phase transformation in metals is due to the instability of the initial state of the alloy. The stability of a system can be evaluated by calculating its Gibbs free energy,  $G$ . The Gibbs energy of a system is defined in terms of its enthalpy,  $H$ , entropy,  $S$ , and temperature,  $T$ , according to

$$G = H - TS. \quad (6.1)$$

At constant temperature and pressure, the system moves toward the equilibrium state which minimizes  $G$ . When a system reaches the steady state, it is said to be in equilibrium. A key consequence of the thermodynamic laws is that, in a closed system, at constant temperature and pressure, the equilibrium phase  $i$  is reached if the lowest value of the Gibbs free energy is achieved, *i.e.*  $dG = 0$ . When the temperature and pressure varies, the change in the Gibbs free energy is obtained by a system of fixed mass and composition (closed system) by

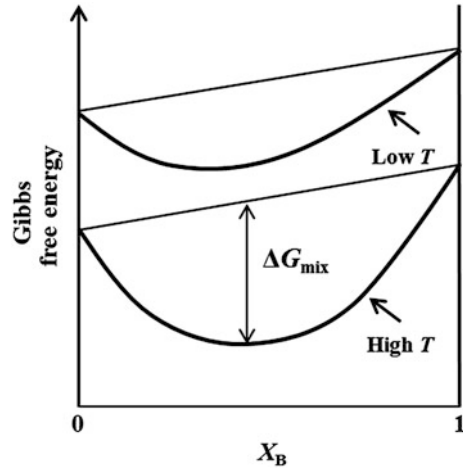
$$dG = -SdT + VdP + \sum \mu_i dn_i, \quad (6.2)$$

where  $\mu_i$  is the chemical potential for the phase  $i$  composed by  $n$  mols of  $i$ . For the case of constant pressure, it can be showed that  $G$  decreases with increasing  $T$  at a rate given by  $-T$ . A single-component system contains a pure element that does not dissociate the temperature range. Considering only liquid and solid states (where numerous allotropic and polymorphic forms of the solid state are considered), the liquid and solid states co-exist in equilibrium when

$$\Delta G_m = 0 \rightarrow G_l = G_s. \quad (6.3)$$

At pressure  $P$ , this equilibrium occurs at the temperature  $T_m$  which is the equilibrium melting temperature, defined as  $T_m = \Delta H_m / \Delta S_m$ . It is known that the liquid phase has a higher enthalpy and entropy than the solid. Consequently, the Gibbs free energy of the liquid decreases more quickly with the increasing temperature than the solid. For temperatures up to  $T_m$ , the solid phase has lower energy, leading to equilibrium. For temperatures above  $T_m$ , the equilibrium of the system is reached in the liquid phase. At  $T_m$ , both phases have equal value of  $G$  and both coexist in equilibrium. The phenomenon of phase transformation is frequently related with the difference in free energy between two phases at a non-equilibrium temperature. For a pure metal, if the liquid temperature is decreased

**Fig. 6.1** Gibbs free energy as a function of composition and temperature for a binary alloy [4]



(e.g.  $T = T_m - \Delta T$ ), the liquid will solidify and the Gibbs free energy will also decrease, providing the driving force for solidification. The driving force is proportional to the undercooling  $\Delta T$ , considering that the latent heat and the entropy of fusion are almost constant with the temperature.

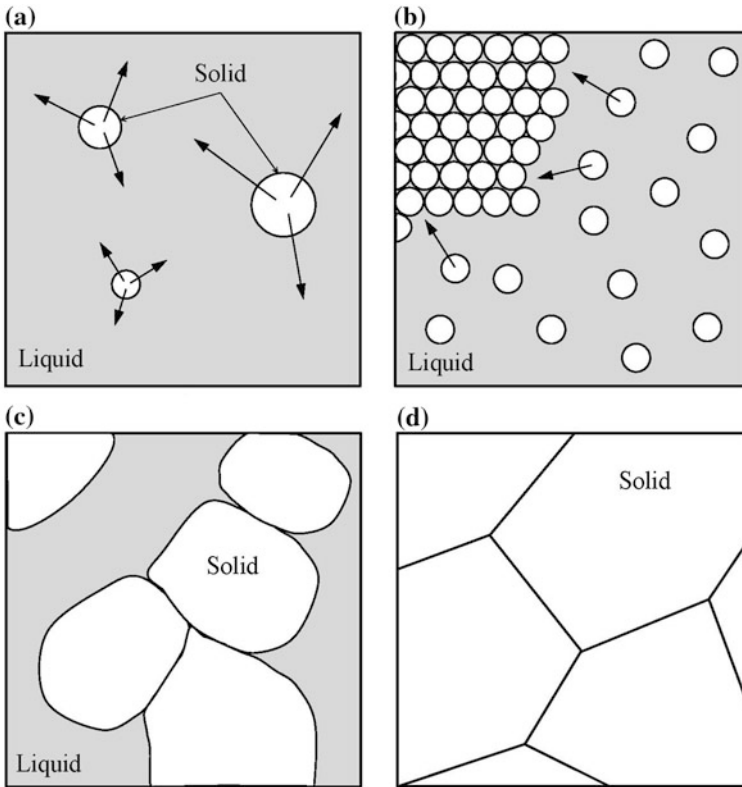
In systems containing two or more components, the Gibbs free-energy for a given phase is  $P$ ,  $T$  and composition dependent. The development of a solid solution phase or a compound phase involves atomic level mixing of several component elements. The Gibbs free energy of each phase is related with the elements involved before mixing and the change in free energy related with the mixing process. Considering an alloy with two constituents A and B, the free-energy of a binary solution is a function of the mass or mole fractions  $X_A$  and  $X_B$  and can be written as

$$G = X_A G_A + X_B G_B + RT(X_A \ln X_A + X_B \ln X_B), \quad (6.4)$$

where the last term is related with the mixing process. Figure 6.1 shows the Gibbs free energy as a function of composition and temperature.

## 6.2.2 Liquid–Solid Phase Transformation

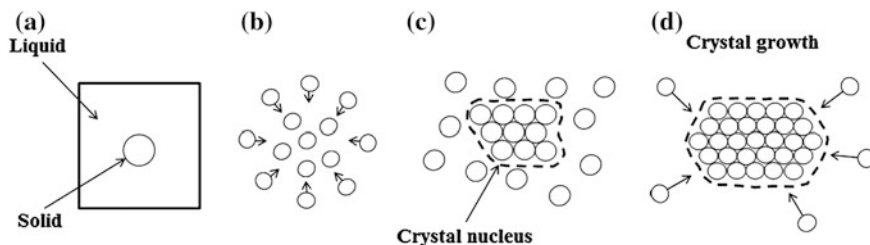
For the liquid-to-solid phase transformation, the temperature at which both phases have equal free energy is the melting temperature  $T_m$ . As temperature decreases toward  $T_m$ , the liquid phase becomes more ordered and the free energy increases. Below  $T_m$ , the free energy becomes negative and the metal solidifies. A schematic representation of the solidification process for a pure metal is illustrated in Fig. 6.2. It shows the formation of stable nuclei (nucleation), crystal growth and the grain structure.



**Fig. 6.2** Nucleation and growth of crystalline structure: **a** nucleation, **b** solid-liquid interface, **c** impingement and **d** solid grains [35]

As the metal cools, solidification begins on a small scale (Fig. 6.3a) with several moving atoms bonding each other to form clusters (Fig. 6.3b). As the temperature decreases, the thermal kinetics of the atoms is lower in the liquid, allowing random aggregations of atoms to form small crystalline regions called embryos. An embryo is a cluster of atoms that has not yet reached a critical size to become stable and grow. Therefore, embryos are continuously developing and remelting [2, 3]. Eventually, as the temperature decreases, some of the embryos will reach a critical size and become stable nuclei capable of growing into crystals. These crystals then continue to grow until they impose on each other and eventually become grains in the final solidified structure. The crystalline structure within each grain is uniform. However it changes abruptly at the interfaces (grain boundaries) with adjacent crystals. This process of forming nuclei in the freezing melt and their subsequent growth is known as a nucleation and growth process.

Two types of nucleation can be found: homogenous and heterogeneous. When the solid particle forms within its own melt without the interaction with the mold



**Fig. 6.3** Solidification steps: **a** Solid precipitate in a liquid matrix, on a microscopic scale; **b** Clustering, **c** Crystalline nucleus and **d** Growth of the crystalline phase [5]

walls or with the help of external materials, it's called homogeneous nucleation. This form of nucleation requires a large driving force due to the important contribution of surface energy. The heterogeneous nucleation, the most general type, occurs when nuclei form on the mold surface wall.

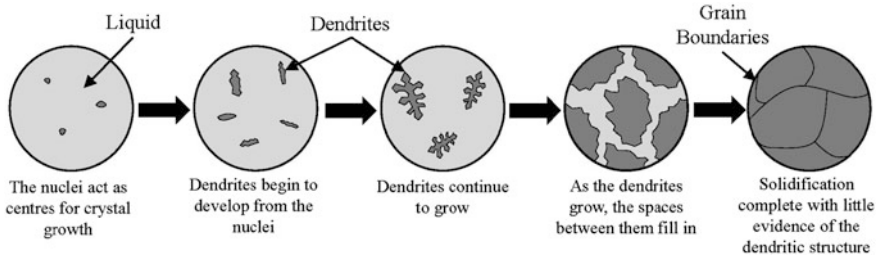
The next step after nucleation is crystal growth, defined as the increase in size of the particles after nucleation. The significant physical processes dictating growth are solute diffusion (in alloys), capillarity (the Gibbs–Thomson effect on interface curvature on phase equilibrium) and thermal diffusion. In the growth of a pure solid, there are two categories of solid–liquid interface: a rough or diffuse interface related to metals and a flat interface associated with non-metals. The growth of a rough interface is called continuous or normal growth since the interface can spread normal to itself in a continuous manner due to the large number of sites for easy atom attachment. The solidification on metals is a diffusion controlled process.

A certain supercooling is previously necessary for nucleation to occur. The primary solid particles grow into the supercooled liquid and the latent heat of solidification is conducted away into the liquid. The initial spherical particle develops tips in numerous directions. As the tips extend their surfaces, these become unstable and break into secondary and tertiary tips. This grown solid, which can be seen in Fig. 6.4, is called a dendrite<sup>2</sup> [4]. In pure metals, dendrites are named thermal dendrites to differentiate from those in alloys. In alloys, specifically in single-phase alloys, the way in which the material solidifies depends on temperature gradients, cooling rates and growth rates. This is also the reason why the mechanical properties of a material depend strongly on the solidification.

Several processes and reactions which are significant in heat treatment of metals and alloys are dependent of the mass transfer within a solid. This is achieved with a phenomenon called diffusion. The diffusion that occurs in pure metals is called self-diffusion, considering that all atoms exchanging position are of the same sort. In alloys, when atoms from one metal alloy diffuse into another, the process is named interdiffusion or impurity diffusion [1]. The two most usual diffusion mechanisms in solid metals are interstitial diffusion, where the interstitial

<sup>2</sup> From the Greek, *Dendros* means tree.





**Fig. 6.4** Dendritic solidification in a pure metal

small atoms migrate by forcing their way between the large atoms, and substitutional or vacancy diffusion, where atoms diffuse by vacancy mechanisms. Interstitial diffusion is generally faster than vacancy diffusion due to the fact that bonding of interstitials to the neighboring atoms is normally weaker and there are many more interstitial sites than vacancy sites to jump to. The classical equations governing diffusion processes are the Fick's laws, also widely employed in classical modeling approaches. The mathematical theory of diffusion in isotropic materials is hence established on the hypothesis that the rate of transfer or the flux of diffusing atom per unit area is proportional to the concentration gradient, i.e.,

$$J = -D \frac{\partial C}{\partial x}, \quad (6.5)$$

where  $J$  is the diffusion flux and  $D$  is the diffusion coefficient. In the previous equation, the concentration varies with distance in a straight line. However, the gradient of concentration is constant. If the diffusion flux does not change with time, a steady-state condition exists. The generalization of the Fick's equation can be written as

$$\frac{\partial C}{\partial t} = \nabla \cdot \mathbf{D} \nabla C = \text{div}(\mathbf{D} \text{ grad} C). \quad (6.6)$$

The solution for the diffusion equation can be obtained for a number of initial boundary conditions if the diffusion coefficient is considered constant. There are several factors which influence the diffusion phenomenon, such as temperature, diffusing species and pressure. Temperature has a significant influence in the diffusion rate and diffusion coefficient. As temperature increases, the coefficient of diffusion increases exponentially. An expression, of the Arrhenius type, can mathematically quantify the diffusion coefficient as

$$D = D_0 \exp\left(\frac{-Q_d}{kT}\right), \quad (6.7)$$

where  $D_0$  is the frequency factor and  $Q_d$  is the activation energy.<sup>3</sup>

<sup>3</sup> The activation energy is the necessary energy to move an atom over a barrier from one lattice site to another. The barrier is related with the assumption that the atom must vibrate with adequate amplitude to break the adjacent neighboring bonds in order to move to a new position.

### 6.2.3 Phase Transformation in Multiphase Metals

Specifically for solid-state steels, phase changes take place when one or more initial phases produce a product phase (or phases) during a cooling process. Nucleation, growth and diffusion are the main processes in solid state transformations. Although diffusion is an important mechanism in this type of transformations, not all are diffusional. As example, the martensitic transformation is diffusionless.

The general characteristics of nucleation and growth transformations are time and temperature dependence, irreversibility of the transformation, effect of the plastic deformation, atomic volume, chemical composition and shape of the product/new phase and orientation relationship [2]. Most of the solid state phase transformations are thermally activated atomic movements. Diffusional phase transformations can be divided in eutectoid reactions, precipitation reactions, ordering reactions, massive transformations and polymorphic changes [4]. The eutectoid reaction is described by the transformation of the solid phase  $\gamma$  in two new solid phases,  $\alpha$  and  $\beta$ , i.e.  $\gamma \rightarrow \alpha + \beta$ . An example of eutectoid reaction is the decomposition of austenite into ferrite plus cementite in (metastable) iron-carbon alloys. Polymorphic transformations arise when there is a change of crystal structure of the metal, affecting all the atoms in the alloy and having an important capacity for changing the alloys microstructure [2, 5]. The most familiar of these in metallurgy are the transformations between  $\gamma$ -Fe and  $\alpha$ -Fe. In precipitation reactions, to change the microstructure it is necessary to add elements to the alloy that are soluble in the base metal at high temperature but that come out of solution at lower temperatures. The particles of the new phase are called precipitates [1]. In massive transformations, the parent phase transforms into the new phase rapidly, maintaining the composition as the original phase [2, 6].

The nucleation theory in solid-state transformation (including steels) is based on the formation of a new phase within a pre-existing (parent) phase. The separation between these two phases is known as interface. In phase transformation controlled by diffusion, the new phase is different from the parent phase in crystal structure and/or composition.

In solid-state phase transformation, a variation in volume accompanies the general phase changes. This volume change has to be adjusted elastically leading to a strain energy effect, which is an important parameter in the phase transformation and in the development of models. When a cluster forms in a solid, a misfit strain energy appears due to the volume difference between the solid and the new phase. The contribution of this energy must be added positively to determine the expression of the total free-energy. Along with the interfacial energy, it is a barrier to nucleation, despite of the type of the crystal/matrix interface. The misfit strain energy of the  $\beta$  phase in the  $\alpha$  matrix can be classically determined assuming the following steps [4, 7]:

1. The  $\alpha$  and  $\beta$  phase are linearly elastic continua, therefore, removing the cluster (modeled as an elastic inclusion) from the  $\alpha$  matrix, leaving a void

and leading to a relaxation of the stresses in the matrix and in the inclusion. The inclusion  $\beta$  will have a different shape than the hole. The transform strain is the homogeneous strain necessary to convert the hole shape to the inclusion shape;

2. Apply surface tractions to the inclusion to return it to its initial shape and insert it into the hole in the matrix again;
3. Join together the inclusion and the matrix along the  $\alpha/\beta$  interface in a way that reflects the interface type that previously existed between them;
4. Allow the  $\beta$  cluster and  $\alpha$  matrix to relax: remove the tractions by applying equal and opposite tractions. This step reestablishes the initial state. The final  $\beta$  crystal and  $\alpha$  matrix are in a self-stress state. The tractions that act on the system at the contact surface will create “constrained” displacements and therefore strains in the  $\alpha$  and  $\beta$  which can be computed employing strain-displacements relationships of elasticity. The misfit strain energy will be calculated from these stresses and strains.

Nucleation, in steels (in solid phase), is heterogeneous. The preferential nucleation sites are non-equilibrium imperfections, dislocations, grain boundaries, stacking faults, free surfaces and inclusions. Nucleation in grain boundaries is analogous to nucleation in solidification on a mold wall. If the misfit strain energy effects are neglected, the nucleation on the grain boundary may be considered as an extension of the heterogeneous nucleation. The expression for the heterogeneous nucleation rate is given by

$$N_{\text{het}} = \omega C_1 \exp\left(\frac{-\Delta G_m}{kT}\right) \exp\left(\frac{-\Delta G^*}{kT}\right), \quad (6.8)$$

where  $C_1$  is the concentration of nucleation site per unit volume. The preferential sites which provide major nucleation rates will be influenced by the driving force ( $\Delta G_V$ ). When the activation barrier has a higher value, the higher nucleation rates will be provided by the grain corner nucleation. As  $\Delta G_V$  increases, grain edges and then grain boundaries will determine the transformation process.

The described process is related to nucleation during isothermal transformations. However, if the nucleation process develops during a continuous cooling, the driving force will rise with time. Assuming these conditions, the early transformation steps will be governed by the nucleation sites which give rise to a measurable volume nucleation rate [4].

In the growth process of a solid state metal where the two phases have distinct crystal structures, an important difference arises from the existence or absence of a good atomic fit through the interface. If the atomic fitting is inadequate as predictable, then simple atom transfer can be expected through the incoherent interface by a thermally active jump process. This form of growth can be categorized as civilian or reconstructive process. Examples of this type of growth, for solid-state transformations, are the growth of two-phase eutectoid structures, such as pearlite in a steel alloy, and discontinuous precipitation. All of the processes nucleate, or appear to nucleate, at the pre-existing grain boundaries. In a coherent interface,

where there is a good atomic matching through the interface, growth is stimulated by the necessity of decreasing the interfacial energy barrier to nucleation [8].

During the growth of the precipitate, the critical nuclei which prevail are those with the small critical volume, meaning they have the minimal nucleation barrier. The nuclei will be surrounded by coherent or semi-coherent facets and smooth curved incoherent interfaces. For the precipitate to grow these interfaces must migrate. Throughout the growth process, the developing shape will be controlled by the migration rates.

Diffusion-controlled growth occurs when most of the driving force is dissipated in diffusion and the interface moves at a rate controlled by diffusion. In precipitation from a supersaturated solution, precipitate growth requires long-range transport of solute to the growing particle and the particle growth kinetics can be modeled as a diffusion problem. In processes like grain growth that do not involve composition changes but only interface mobility, the boundary migration kinetics involve local atomic rearrangements as atoms jump from one grain to its neighbor. Such process is said to be interface controlled growth.

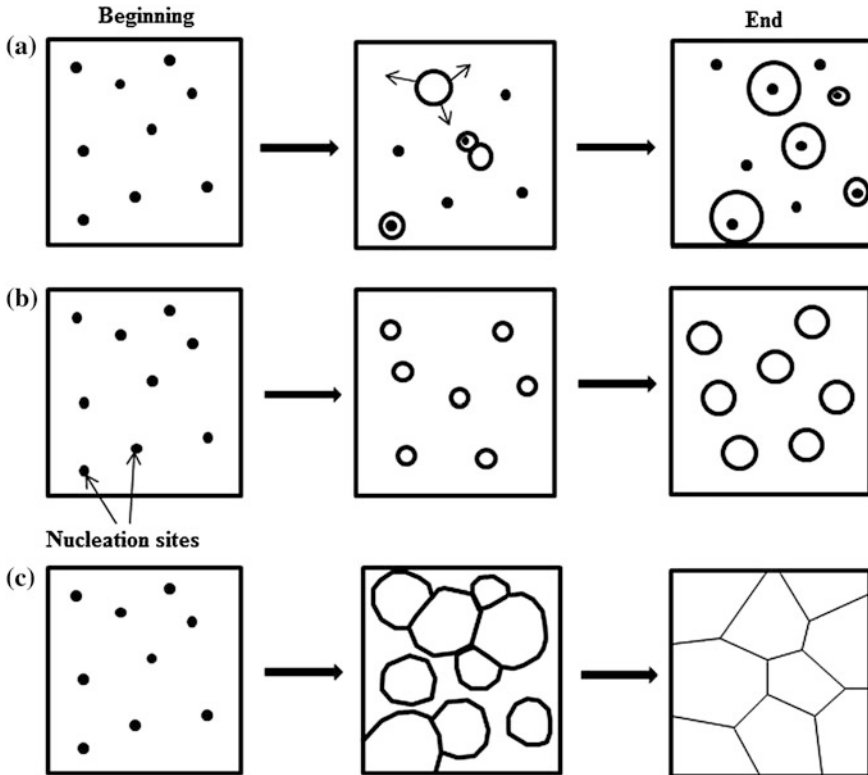
It is also worth mentioning the types of phase transformation in metals (and specifically in steels) considering that these types affect the subsequent modeling formulation. In the quenching process, at the transformation temperature, the  $\alpha$  phase will have numerous heterogeneous nucleation sites. Figure 6.5 shows the possible events. For example, when the nuclei forms throughout the transformation so that a wide range of particles sizes exists at any time (Fig. 6.5a) and when all the nuclei form right at the beginning of the transformation (Fig. 6.5b). If all potential nucleation sites are depleted in the process this is known as site saturation. Figure 6.5c shows the cellular transformation, where all the parent phase is used by the transformation process. In these cases, transformations finish with impingement<sup>4</sup> of adjacent precipitates growing with constant velocity and not because of the gradual reduction of growth rate [1, 4].

#### ***6.2.4 The Classical Model of Johnson, Mehl, Avrami and Kolmogorov***

The kinetics of nucleation and growth in solid state transformations is generally described by the classical theory of Johnson, Mehl, Avrami and Kolmogorov (JMAK) [9–11]. The JMAK theory considers isothermal transformation kinetics of the nucleation, growth and impingement of a product phase in the matrix of the parent phase until the entire parent phase is transformed. In isothermal transformations, in order to model this type of change it is necessary to determine the nucleation and growth rate. However, to estimate the volume fraction, the

---

<sup>4</sup> The restriction of the transformed region growth by other transformed region.

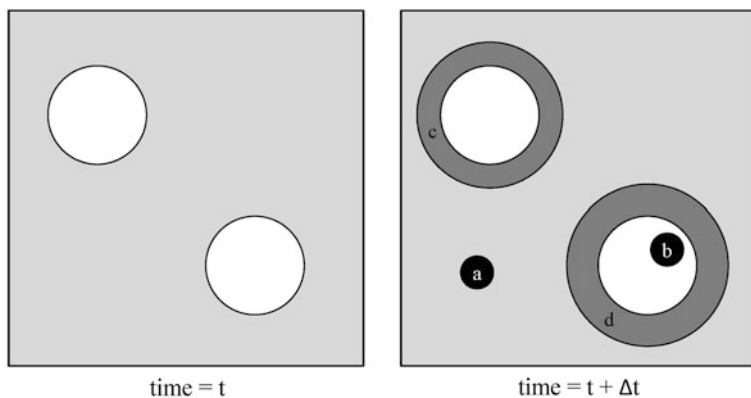


**Fig. 6.5** **a** Nucleation at a constant rate during the transformation process; **b** site saturation and **c** a cellular transformation [4]

impingement between precipitates must be considered. This is achieved by the JMAK extended volume concept.

The JMAK theory is based on fundamental assumptions such as (i) nucleation occurs randomly and homogeneously and (ii) with constant isotropic growth rate with an infinite volume available for transformation. Other simplifying assumptions have been taken in consideration about geometry and kinetics of nucleation and growth, in order to derive analytical solutions for particular cases, such as zero-nucleation rate (pre-existing nuclei), linear growth velocity, diffusion-limited growth, and growth of crystals in needle- or plate-like configurations [6, 12]. There are several main concepts in the JMAK classical model: (i) the model does not consider impingement and (ii) it considers a “phantom” nucleus which is formed in the volume fraction that has already been transformed.

Consider two particles at a given time  $t$ . After a short period of time  $\Delta t$ , new regions (a, b, c and d) are formed assuming that they are able to grow unrestricted in extended space whether or not the region into which they grow is already transformed, as illustrated in Fig. 6.6 [12]. However only those elements of a, b, c



**Fig. 6.6** Extended volume definition: nucleation and growth of two particles. The new regions c and d are created as the initial particles grow. The new precipitates a and b, in which particle b has been created in a previously transformed region

and d which lie in formerly untransformed parent phase can contribute to a change in the real volume of the product phase, here identified by  $\alpha$ . Therefore, the Avrami equation is written as

$$dV^\alpha = \left(1 - \frac{V^\alpha}{V}\right) dV_e^\alpha, \quad (6.9)$$

where it is assumed that the particles are located randomly in the space. The  $dV_e^\alpha$  term is the extended,  $V^\alpha$  is the volume of  $\alpha$  and  $V$  is the total volume. The probability that the extended regions in Fig. 6.6 will fall into untransformed regions, is given by  $(1 - V^\alpha/V)$ .

Considering that the growth rate is constant, the particle has an isotropic growth, and an interval of time  $\tau$  and  $\tau + \Delta\tau$ , the extended volume change can be written as

$$dV_\alpha = \left(1 - \frac{V^\alpha}{V}\right) = \frac{4}{3}\pi G^3(t - \tau)^3 I_V V d\tau, \quad (6.10)$$

where the particle has a sphere shape, the nucleation rate per unit volume is  $I_V$  and the time  $t = \tau$ . Considering  $\xi = \frac{V^\alpha}{V}$ , assuming a constant nucleation and growth rates and random nucleation, Eq. 6.10 can be integrated leading to the following JMAK equation:

$$\xi = 1 - \exp(-k_A t^n), \quad (6.11)$$

where  $k_A$  and  $n$  describe the reaction as a function of time, temperature and other parameters. The Avrami exponent  $n$  is a numerical value obtained from experimental data. If there is no change in the nucleation process,  $n$  can be independent

of the temperature. However  $k_A$ , which can be also obtained by experimental data, is sensitive to temperature because of nucleation and growth rates dependence.

Typically, and depending on the assumptions made concerning the nucleation and growth process, a variety of analogous expressions can be obtained with this form [12].

Recently, numerical and analytical kinetic models have been proposed. These identify the three following processes: nucleation, growth and impingement of particles as entities that can be modeled separately. Also, the JMAK parameters can be expressed concerning individual nucleation and growth parameters. The Avrami exponent can be given by  $n = a + bc$ , where the empirical nucleation index  $a$  controls the time dependence of the number of particles per unit volume of untransformed material as a function of time. The parameter  $b$  is the dimensionality of growth:  $b = 1, 2$  or  $3$  for one-two or three dimensional growth, namely when the particles of the new phase grow like needles, disks or spheres, respectively. Finally,  $c$  is the growth index:  $c = 1$  for interface growth and  $c = 0.5$  for diffusion-control growth [2].

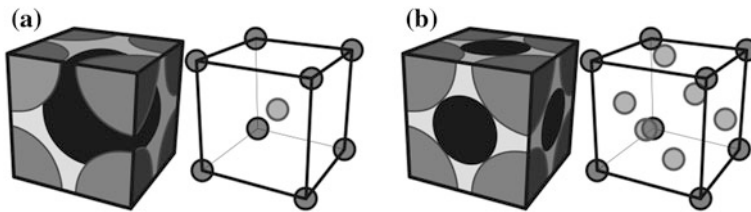
### 6.2.5 Steel and Transformation Kinetics

Steel are defined as alloys consisting mainly of iron and carbon and other alloying elements.

The chemical element iron (Fe) is one of the most common on Earth, however it does not occur in nature in a usable form but in mineral ore such as hematite or magnetite. It is the least expensive and most widely used metal. There are three allotropic forms of pure iron: faced centered cubic (fcc) austenite ( $\gamma$ ), body centered cubic (bcc) ferrite ( $\alpha$ ) and the hexagonal compact structure ( $\epsilon$ ). The ferrite or  $\alpha$ -iron is characterized by unit cells and belongs to the cubic system, where the three axes of the unit cell have the same length and are reciprocally perpendicular [13]. The lattice spacing of  $\alpha$ -iron is body-centered cubic. The austenite crystallizes in the cubic system. The unit cell has a cubic geometry with the iron atoms located at each cube edge and an atom on each face of the unit cell, *i.e.*, a face centered cubic structure [1]. The fcc structure, although more closed packed than the bcc structure, has larger “holes” or interstices, which are located at the center of the cube edges and are enclosed by six atoms in the form of an octagon (octahedral interstitial sites) and tetrahedral interstices. (Fig. 6.7)

The majority of steels depend on just two forms of iron:  $\gamma$  and  $\alpha$ . The transformation  $\gamma \rightarrow \alpha$  has associated a volume change, which may contribute to the formation of internal stresses during the phase transformation. The geometry of  $\alpha$ - and  $\gamma$ -phase is important to the solubility of carbon or nitrogen in the two-phases, in the diffusivity of the alloying elements and in the plastic deformation behavior [14].

Carbon content in steels ranges between 0.03 % and 2.0 % by weight and containing other elements like Si, P, Mn. The ferrous alloys with carbon content



**Fig. 6.7** a Ferrite and b Austenite crystal structure

exceeding 2.0 wt% are called cast irons [13, 14]. There are several products with varied chemical composition, shapes and sizes that are manufactured with iron and steel by casting, rolling and forging. Of all metals and alloys operated and produced worldwide, iron and steel represent more than 90 % of the annual production and it is by far the most affordable metal. In some applications, such as steel frames for large buildings, steel is the only material specified due to strength requirements.

Steels can be classified by different forms such as based on chemical composition, manufacturing methods, structural, tool applications, heat resistance, finishing methods (hot rolling, cold rolling, casting); product shape (plates, sheets); oxidation employed; microstructure (ferritic, austenitic, etc.); heat treatment and thermomechanical processing. However, chemical composition is the most widely used basis for designation. The most relevant codes defined by international institution to classify steels are the American Iron and Steel Institute (AISI) and the German Deutsches Institut für Normung (DIN).

The performance of steels are dependent of the properties associated with their microstructure: structural arrangements, volume fraction, grain size and morphology of the phases that constitute them. The study of the development of iron and steel starts in the equilibrium diagram Fe–C, as shown in Fig. 6.8, which provides an important basis on the knowledge of different types of alloy steels. The carbon content in the steel deeply alters the phase relationship between the microstructure and properties in the steel. The solid phases present in the phase diagram of Fe–Fe<sub>3</sub>C (Fig. 6.8) are:

- Austenite ( $\gamma$ ), an interstitial solid solution of carbon in the FCC iron crystal lattice. The solubility of carbon is much higher when compared with ferrite;
- Ferrite ( $\alpha$ ), an interstitial solid solution of carbon in BCC iron crystal structure, reaching the maximum solubility at 727 °C;
- $\delta$ -ferrite, with a crystal structure similar to alpha ferrite, but stable between 1394 and 1530 °C;
- Cementite (Fe<sub>3</sub>C), a intermetallic compound which has a carbon content of 6.67 % by weight.

There are important temperatures or critical points in the Fe–Fe<sub>3</sub>C phase diagram. The first one is the A<sub>1</sub> temperature, at 727 °C, where the eutectoid reaction occurs. Other temperature is A<sub>3</sub>, where  $\alpha$ -iron transforms in  $\gamma$ -iron. In pure iron, this



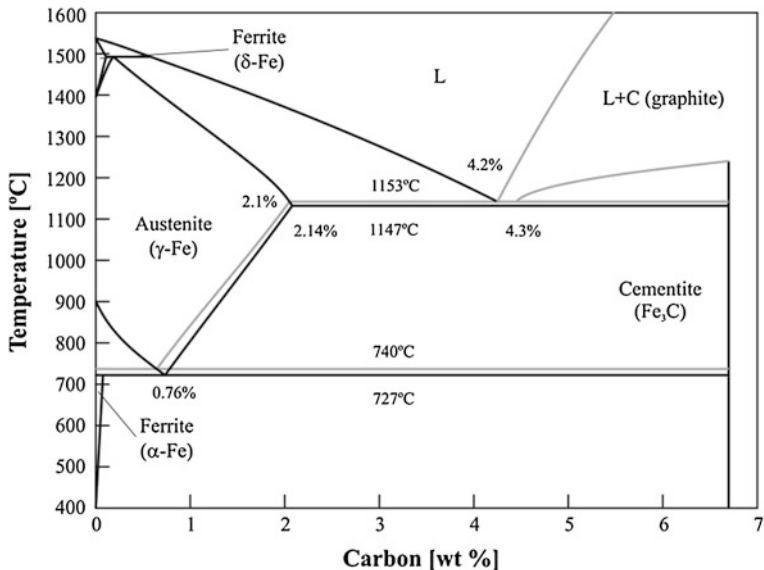


Fig. 6.8 Fe–Fe<sub>3</sub>C phase diagram [3]

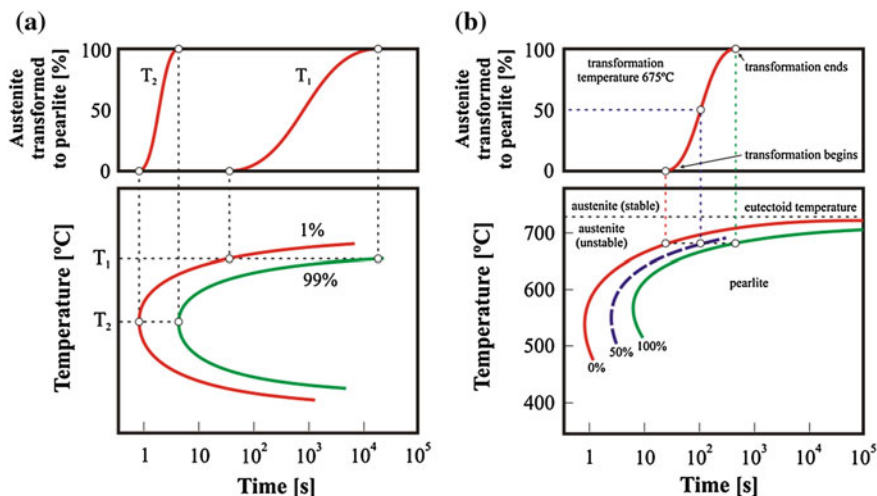
transformation happens at 910 °C, however, with the carbon addition, the temperature is lowered. The  $A_{cm}$  is the  $\gamma$ -iron transformation in  $\gamma + Fe_3C$  boundary.

The ferrite-cementite mixture is called pearlite and is formed during cooling. The pearlite structure consists in alternating ferrite and cementite lamellae.

The Fe-Fe<sub>3</sub>C phase diagram has three invariants: the peritectic, the eutectic and the eutectoid. However, only the eutectoid reaction occurs below 900 °C. Generally, eutectoid reaction occurs at 727 °C and is given by  $\gamma$  (0.77 wt.%C) = ferrite- $\alpha$  + Fe<sub>3</sub>C.

An important cause of the wide variety of microstructures in steels is the allotropic transformation which can occur with a variety of forms, where the atoms can move to achieve the change in the crystal structure. The transformation can happen either by breaking all the bonds and atomic rearrangement into another pattern (reconstructive transformation) or by homogeneous deformation of the initial pattern into a new crystal structure (displacive or shear transformation). All phase transformation in steels can be considered within these two mechanisms.

The understanding of phase transformations is decisive for the development of the microstructure of a material for a specific application. This implies a fully understanding of the phase transformation kinetics, *i.e.* the time and temperature dependences, including the transformation rates. To determine the reaction rate, it is necessary to measure the degree of transformation as function of time. The overall transformation kinetics strongly depends on the nucleation and growth rates and the impingements effects [15]. The isothermal phase transformation evolution (isothermal transformation kinetics) as a function of time and temperature for a metal alloy is given by the temperature–time–transformation (TTT)



**Fig. 6.9** **a** General temperature-time-transformation (TTT) diagram and **b** a TTT diagram for a steel. The formation of an isothermal transformation curve generated from a transformation versus logarithm of time measurements. Adapted from [1, 36]

diagram, illustrated in Fig. 6.9. The curves have a C shape due to the driving force for transformation, which is small at higher temperatures. However, the diffusion coefficient has a smaller value at lower temperatures. At intermediated temperatures, the combination of these two variables gives a maximum in the reaction rate. The start curve signals the beginning of the transformation (*e.g.* 1 %) and the finish curve indicates the end of the transformation (99 % curve) [16, 17].

Therefore, in steels, the transformation of austenite can be studied during continuous cooling, where physical measurements are applied (thermal analysis, dilatometry, etc.), although the results are sensitive to the applied cooling rate. The isothermal transformation shows a particular time-temperature-transformation (TTT) curve, with a defined C-shape for each specific steel.

In practice, steel is rarely quenched to a constant temperature and have an isothermal transformation. Most of the heat treatments of steel include continuous cooling of a specimen. Continuous-cooling-transformation (C-C-T) diagrams represent the transformation, temperature and time relationships during continuous cooling. Figure 6.10 shows a moderately rapid and slow cooling superimposed on a C-C-T for a eutectoid steel. The transformation begins after a period of time that corresponds to the intersection of the cooling curve with the start transformation curve and finalizes upon crossing the end transformation curve. The microstructures for these conditions are fine and coarse pearlite.

The hardening of steel through quenching is one of the most important technological processes. If the steel has a rapid quenching from austenite, there is no time for the eutectoid transformation by diffusion-controlled to occur and a martensitic structure will develop. Martensitic reaction is diffusionless. Therefore, the product

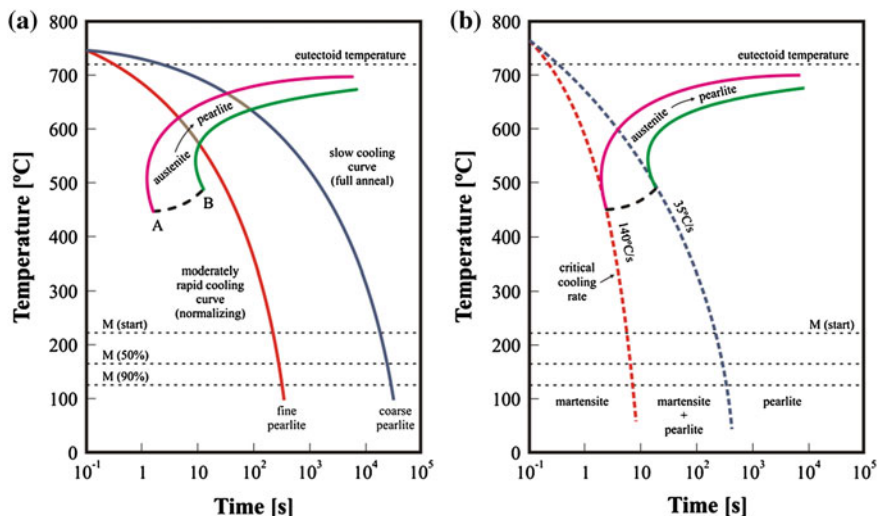
phase composition is similar to the parent phase. The martensitic transformation depends on several variables such as time and temperature. The transformation has a very fast velocity and starts at a fixed temperature  $M_s$  (Martensite start temperature). With a continuous variation of temperature, more material will transform until it reaches a temperature  $M_f$  (Martensitic finish temperature), where the transformation is completed. Another characteristic of the martensitic reaction is the reversibility of the transformation because the initial atomic configuration can be obtained continually. The reversibility is related with a temperature hysteresis and the reverse reaction begins at a temperature above  $M_s$  [2, 6].

The effect of the stress is also a factor in the martensitic transformation since the plastic deformation plays an important role in this type of reaction when compared with nucleation and growth transformations. At any given temperature, the use of plastic stress in the reaction range generally increases the amount of transformation. The deformation above  $M_s$  could result in the generation of the product phase. In general, the reverse transformation can be promoted in a similar form and an appropriate stress-state will produce a transformation below the temperature at which it begins spontaneously. If the parent phase has been submitted to cold-work in a temperature where it is stable (above  $M_s$ ), the subsequent deformation inhibits the transformation. In that case, and if the deformation temperature is not high enough to allow self-annealing, the  $M_s$  temperature decreases and the volume of the reaction obtained at any temperature is reduced.

The composition and the shape of the product phase in the martensitic transformation are other characteristics of these reactions. Although the parent crystal and the new crystals formed have the same chemical composition, the volume changes are regular. The crystals of martensite are flat plates, which became thinner at the end and develop a lenticular cross-section. The martensitic plates orientation follows the original lattice; the plane of the lattice on which they are developed is named habit plane. In martensitic transformation, there is a fixed relation between the orientation of the parent phase and the product phase. A single martensite plate could be a single crystal or could have two twin orientations [2, 6].

There is a critical rate for quenching, for the continuous cooling of a steel, that characterizes the minimum value of quenching rate that forms a complete martensitic microstructure. When this rate is included in C-C-T diagram, it will fail the nose of the curve at which the pearlitic reaction starts, as shown in Fig. 6.10b. It is also possible to observe that, for rates greater than the critical rate, only the martensitic microstructure occurs. Pearlite develops for low cooling rates.

It is the knowledge of this transformation phenomena (including temperature rates, etc.) that allow to perform heat treatments with the aim of retrieving specific microstructures and, consequently, specific properties. Heat treatment in steels is referred to the heating and cooling operations to modify the material properties. Heat treatments can be applied in casting, welding and other processes. The main objectives of a heat treatment are: relieve internal stresses set up during hot-working, cold-working, welding and casting processes; to increase the resistance to wear, heat and corrosion in an alloy by softening to improve the machinability; to improve the hardness of a tool steel by modification of grain structure after a



**Fig. 6.10** Continuous cooling transformation (C-C-T) diagrams for a eutectoid iron-carbon alloy: **a** for a moderately rapid and slow cooling, **b** dependence of the final microstructure on the transformations during cooling (adapted from [1])

hot-working and eliminate former heat treatment procedures; to remove strain hardening developed into cold-working, improving its ductility; to refine a steel microstructure after hot-working; to create a single phase in stainless steels, etc.

The most common heat treatments processes are annealing, normalizing, spheroidizing, quenching and tempering.

In annealing, the alloy steel is maintained at an elevated temperature for a period of time and then slowly cooled. This process is used to increase ductility, softness or to produce a specific microstructure.

The normalizing is used as a final treatment for carbon steels adding more strength to the steel, when compared with annealing, due to the faster cooling rate produced. This is a type of annealing treatment which is employed to refine the grains and create a uniform size distribution.

The spheroidizing heat treatment improves the machinability. The process requires several hours at approximately 30 °C below  $A_1$ , where the  $Fe_3C$  microstructure is changed into large spherical particles.

The quenching process is referred to cooling from the temperature range at such rate that the transformations in the primary and bainite ranges are overcome and martensite is formed. This induces a high strength and hardness in steels. To obtain the desired cooling rate during quenching, several cooling methods are used and, as a cooling medium, water, air or oil can be employed.

The principal objective of tempering is to obtain a disperse structure at a degree of cooling. In low carbon steels, quenching acts as tempering. When specific steels are quenched in oil, the microstructure is developed during transformation within the bainite range that is more disperse than the one formed after cooling in air.

### 6.3 Thermomechanical Modelling of Steels Through Classical Engineering Approaches

The classical engineering approach to model a material, even when it shows a heterogeneous microstructure, is to consider it macroscopically. Therefore, numerical models prepared for full thermomechanical processes at large deformations can be applied to simulate the macroscopic behavior of the material. Nowadays, these models can be found in the large majority of simulation software by the finite element method. These models considers a mathematical formulation based on a hypoelastic model and an additive decomposition of the rate of deformation tensor is assumed, *i.e.*,

$$\mathbf{D} = \mathbf{D}^e + \mathbf{D}^{th} + \mathbf{D}^{vp}, \quad (6.12)$$

where  $\mathbf{D}^e$ ,  $\mathbf{D}^{th}$  and  $\mathbf{D}^{vp}$  are the elastic, thermal and viscoplastic parts of the rate of deformation tensor  $\mathbf{D}$ , respectively. The thermal part of the rate of deformation tensor is defined by

$$\mathbf{D}^{th} = \alpha^* \dot{T} \mathfrak{S} \quad \text{with} \quad \alpha^* = \alpha + \frac{d\alpha}{dT} (T - T_0), \quad (6.13)$$

where  $\alpha$  is the coefficient of thermal expansion (CTE) of the material,  $\dot{T}$  is the temperature rate and  $T_0$  is a reference temperature. Considering the thermal effects, the elastic part of the rate of deformation tensor can be defined as

$$\mathbf{D}^e = \mathbf{C}^{-1} \overset{\nabla}{\boldsymbol{\sigma}} + \dot{T} \frac{\partial \mathbf{C}^{-1}}{\partial T} \boldsymbol{\sigma}, \quad (6.14)$$

where  $\mathbf{C}$  is the temperature-dependent elastic isotropic tensor

$$\mathbf{C} = 2\mu(T)\mathfrak{S} + \lambda(T)\mathbf{1} \otimes \mathbf{1} \quad (6.15)$$

and  $\overset{\nabla}{\boldsymbol{\sigma}}$  is an objective rate of the Cauchy stress tensor  $\boldsymbol{\sigma}$ .  $\mathfrak{S}$  is the fourth order identity tensor. The constitutive equation for  $\mathbf{D}^{vp}$ , *i.e.*, the flow rule, is given by

$$\mathbf{D}^{vp} = \frac{3}{2} \frac{\dot{\bar{\epsilon}}^p}{\bar{\sigma}} \boldsymbol{\sigma}', \quad (6.16)$$

where  $\boldsymbol{\sigma}'$  is the deviatoric stress tensor,  $\bar{\sigma}$  is the von Mises equivalent stress and  $\dot{\bar{\epsilon}}^p$  is the equivalent plastic strain rate, considered to be dependent on (i) the equivalent stress, (ii) the temperature and (iii) an evolutive internal state variable  $s$  that attempts to model some phenomena derived from the microstructure. The use of internal variable(s) is a stratagem to include some microstructure information without modeling the microstructure. Consequently, the computational effort can be acceptable. However, it should be mentioned that the majority of the classical models do not include internal variables that intend to introduce microstructure information. Generally, these internal variables define some macroscopic information (that is the case of, for example, the backstress tensor).

Data from thermomechanical tests suggest that the behavior of some steels can be accurately represented by a power law, while others are better represented by an exponential dependence. In order to accommodate both dependencies the adopted specific form for the equivalent plastic strain rate can be [18, 19]

$$\dot{\varepsilon}^p = A \exp\left(-\frac{Q}{R_g T}\right) \left[ \sinh\left(\zeta \frac{\bar{\sigma}}{s}\right) \right]^{1/m}. \quad (6.17)$$

In the previous constitutive relation,  $A$ ,  $m$  and  $\zeta$  are material parameters and  $R_g$  is the universal gas constant. The previous equation is an example of a macro-structural model prepared for thermomechanical problems. Other simpler model is the Norton model

$$\dot{\varepsilon}^p = \left[ \frac{\bar{\sigma} - Y - Q[1 - \exp(-b\varepsilon^p)]}{N} \right]^n, \quad (6.17a)$$

where  $Y$  is the yield stress and  $b$  and  $N$  are constants of the material. The activation energy term can also be considered to be dependent of the temperature, *i.e.*,

$$Q = \begin{cases} Q_0 [1 + \log(\frac{T}{T_f})] & \text{if } T \leq T_f \\ Q_0 & \text{if } T > T_f \end{cases}. \quad (6.18)$$

$Q_0$  is the activation energy in steady-state conditions and  $T_f$  is a behaviour transition temperature. The internal state variable  $s$  and its evolution are generally defined by the universal following relation

$$\dot{s} = \dot{\varepsilon}^p h(\bar{\sigma}, s, T) - \dot{r}(s, T) \quad (6.19)$$

where  $h(\bar{\sigma}, s, T)$  is associated to micro structural phenomena such as hardening and dynamic recovery, whereas  $\dot{r}(s, T)$  is related to static restoration effects. The classical models represent a full thermomechanical solution to the problem. The solution of the temperature field is given by the heat transfer equation

$$\rho c \frac{\partial T}{\partial t} + \text{div}(\mathbf{q}_k) = q_i + q_t, \quad (6.20)$$

where  $\rho$  and  $c$  represent the density and the specific heat, respectively.  $\mathbf{q}_k = -\mathbf{k} \text{grad}(T)$  is the conduction heat flux governed by the Fourier law and  $q_i$  and  $q_t$  define the plastic heat generation rate and other heat sources, respectively.  $\mathbf{k}$  is the thermal conductivity tensor. In the case of heat treatments, the thermal problem is driven by the surface boundary condition of convection.

These classical models can be implemented in a finite element code. Due to the strong non-linearities caused by the thermal and mechanical properties, the solution of the problem can only be achieved slowly. For that reason, generally the previous equations are implemented in an implicit formulation, where iteration cycles are performed in order to guarantee a final equilibrium solution. There are simulation codes that use numerical prediction/correction algorithms. The prediction stage, which uses a semi-implicit or explicit integration scheme, provides a

first estimation for the temperature, strain and stress fields. The correction stage, performed with a Newton–Raphson implicit scheme, improves the solution until a satisfactory result is found. Due to high computational effort, the thermomechanical coupled problem can be solved with a staggered approach. In this staggered approach the information between the thermal and mechanical problem is passed at each increment. This solution has demonstrated to be very efficient in different kinds of thermomechanical problems.

### ***6.3.1 Water Spray Quenching: Example of a Classical Approach***

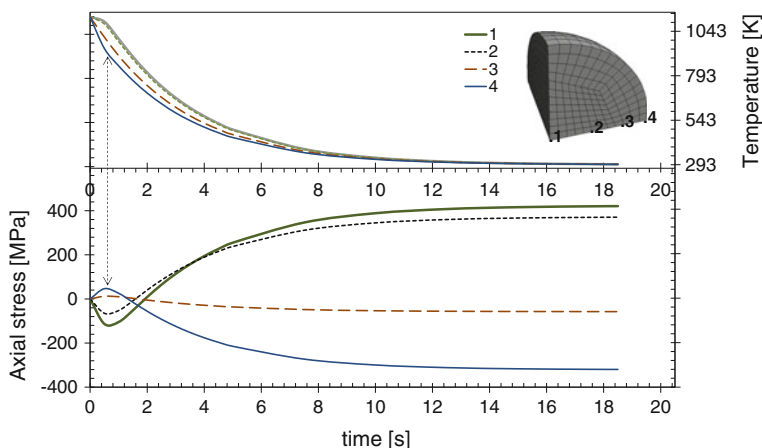
An example of application of such classical approach is the simulation of the quenching heat treatment. Three-dimensional models were employed in the simulation of spray water quenching of both cylinders and spheres of an austenitic stainless steel [20].

Figure 6.11 shows the evolution of the temperature and the axial stress for selected points of the sphere specimen radius, including the points at the center and the surface. As expected, it can be seen that cooling is faster for the points closer to the surface in contact with the quenching fluid. The compressive stress revealed at point 1 at the start of the process (0–2 s) transformed into tensile stress at subsequent stages. Points 1 and 2 followed qualitatively similar evolutions. Point 4, at the surface, presented the higher compressive stress of the solid sphere. However, this point presented tensile stresses of  $\approx 50$  MPa at the early stages of the process.

It can be seen in Fig. 6.11 that at the beginning of the quenching process tensile stresses were produced on the outside of the sphere and compressive on the inside. For  $t > 2$  s stresses were reversed. This phenomenon is due to the fact that initially the temperature gradient between the center and surface increases and the outside zone tends to contract more quickly. Then, the regions next to the surfaces are in tension and the inside in compressive, as indicated in Fig. 6.11. After 2–3 s of the process, when the temperature field becomes homogeneous and the inside tends to cool more quickly, the stress distribution is inverted. These stresses correspond to the situation where the inside tends to contract but it is opposed by the involving region. Consequently, resulting stresses are tensile in the center and compressive on the surfaces.

Figure 6.12 shows the resulting residual stress fields across radial line ( $z = 0$ ) after the process of spray water quenching. The results obtained by the classical model confirm the experimental results measured using neutron diffraction [21] and the analytical model [22] also presented in Fig. 6.12.

The simulation of the quenching process described in this text proved to be in accordance with the principles and physical phenomena that describe the heat treatment processes. It also demonstrates the suitability of the classical models. The results achieved are in good agreement with the results obtained by experimental and analytical methods.



**Fig. 6.11** Evolution of the temperature and the axial stress during the quenching process

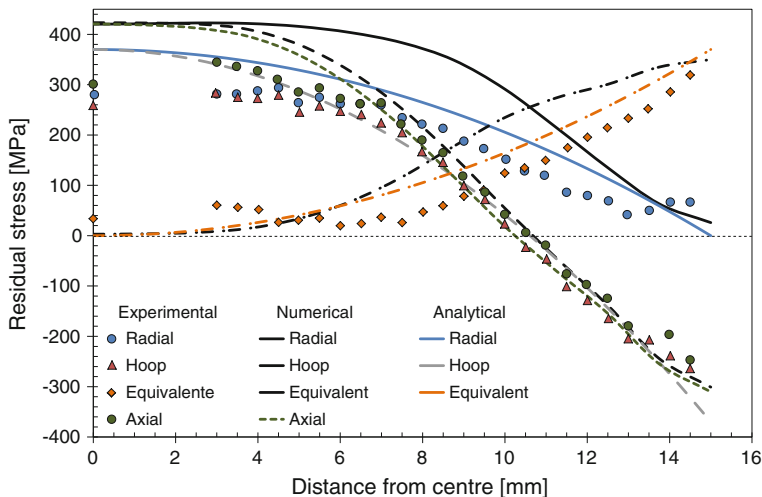
## 6.4 Modeling Steels Through Modern Analyses: The Use of Microstructural Information

In the last few decades, several phenomenological theories have been developed in order to describe the mechanics and thermodynamics of thermomechanical processes on the macroscopic level, as presented in the previous section. Nevertheless, these classic models have reached their limits in various aspects. Extensions of these models should take into account the formation of microstructures and the microheterogeneity of the underlying multiphase materials. The macroscopic response functions should be determined by appropriate averages of microscopically associated field quantities over evolving microstructures.

Scale-effects play a major role in modern micromechanical applications. Microstructure is indeed crucial, since thermomechanical behavior typically is the result of the interaction of complex substructures on several length scales. To understand some heat treatments and subsequent forming phenomena in high performance steels, models which are more closely related to physics and material science and able to take into account the microstructural behavior of the material are needed. These models rely strongly on variational formulations for which effective mathematical and numerical concepts have been developed only recently [23, 24].

Modern engineering applications require the reduction of structure weight while improving safety properties. Therefore, advanced high strength steels play an important role since they offer solutions to these demands. Due to their microheterogeneity, high strength steels pose outstanding challenges with respect to material modeling.





**Fig. 6.12** Calculated residual stresses in a 30 mm diameter quenched steel solid sphere compared to experimental [21] and analytical [22] solutions

Prospects for manipulating microstructure to achieve enhanced performance demands models that distinguish between nucleation, migration, absorption/desorption, trapping, and bypass or annihilation of dislocations at various material length scales that manifest work hardening behavior. Most of these phenomena cannot be considered in isolation, which is a hallmark of plasticity—it is a highly coupled phenomenon, in general, with important effects of both short and long range character attributed both to the physics of dislocation cores and long range interactions of dislocation arrays.

Considering that the mechanical behavior of high performance steels (such as multi-phase steels) is attributable to their microstructure, the modeling of the mechanical behavior of these steel materials has to be done based on microstructural levels. Nevertheless, macroscopic properties are also required for industry engineers.

Due to large advances in computational power, both in hardware and software, it is nowadays possible to use the microstructural information in the modeling of steels. Diverse methodologies are generally used to reproduce the behavior of multiphase steels. One of them, known as homogeneous strategy, uses the microstructural information as an average value. For a multiphase steel, the different phases are considered as volume fractions of a homogeneous material in which its behavior is a weighted average of the behavior of the constituent phases [25]. This kind of models can consider all sort of phases and grain size representation. However, in simulation codes, the phases and grain size are not discretized and, consequently, grain and phase boundaries are not considered. The thermal and metallurgical phenomena, such as phase transformation kinetics, are defined analytically in the model but the distributions of the phases and residual stresses due to different material thermal expansion coefficients are not represented [26].

Other options are available concerning the model of the microstructural phenomena. In this decade, crystal plasticity based models [27] and the so-called self-consistent models [28] have proven to be very useful and accurate reproducing valuable properties such as grain orientation and anisotropy. However, grain boundaries, phase distribution or transformation and occurrence of local heterogeneities cannot be represented in these models. The prediction of the creation, growth and role of grain boundaries is of the utmost importance for the microstructural behavior of the polycrystalline material in heat treatments and subsequent forming behavior. It should be highlighted that grain boundaries may act either as barriers for dislocation glide, or conversely as sources for bulk dislocations.

Only recently, multiphase transformations have been considered in finite element (FE) simulations of large deformation processes. There are now models that explain macroscopically the observed plasticity induced by the volume change between the hard and soft phases and models that use the preferred orientations and microscopic observations to explain the volume mismatch between phases [26].

In opposition to this kind of models are the multiscale homogenization models type. For instance, the asymptotic expansion homogenization method is an excellent methodology to model physical phenomena on media with periodic microstructure and a useful technique to study the mechanical behavior of structural components built with these kinds of materials.

The detailed numerical modeling of the mechanical behavior of multiphase material structures, considering the full micrometallurgical representation of the constituent phases, often involves cumbersome computational costs. Consequently, the use of homogenization methodologies can lead to significant reduction in computation efforts. These techniques allow the substitution of a heterogeneous medium for an equivalent macroscale homogeneous medium allowing the use of macrostructural behavior laws obtained from microstructural information. On the other hand, multiphase steel materials typically have heterogeneities with characteristic dimensions much smaller than the dimensions of the structural component itself. If the distribution of the heterogeneities can be described as periodical, it can be approximated by a periodical repetition of volume cell, representative of the microstructural details of the multiphase material. In this approach the multiphase material can be considered approximately as a composite material. These models can represent grain size and grain boundaries. When using with the finite element method (FEM), it is possible to analyze the different phase boundary effects. Nevertheless, the transformation kinetics phenomena are very difficult to take into account. Additionally, these models require specific numerical boundary conditions in order to fulfill the requirements of microscale periodicity.

In recent years, phase-field models [29] were able to simultaneously simulate nucleation, grain growth and coarsening and provide a visual impression of the development of a structure. However, these methods only take in account the phase kinetic and energetic equations discarding the mechanical issues. Although these methods are able to compute quantitative aspects of the evolution of

microstructure without explicit intervention, there are some assumptions which are inconsistent with practical experience [30]. Additionally, these only model the microstructural information. Therefore, they should be used coupled with the finite element method in order to simulate real technological thermomechanical processes. The phase-field model can be used to reproduce the transformation kinetics phenomena which is very difficult to calculate by FEM.

Examples of modern thermomechanical models are presented in the following section.

### 6.4.1 Homogeneous Multiphase Transformation Models

For models that consider microstructural phenomena, including volume expansion due to phase transformation, and where the different phases are considered as volume fractions of a homogeneous material in which its behavior is a weighted average of the behavior of the constituent phases, the thermal properties must be also written as volume fractions of its constituents. Therefore, the enthalpy changes of phases need to be accounted for due to the temperature increase by latent heat generation. The density and the specific heat, which are defined as weight average of the  $n_p$  constituent phases, use the information of the volume fraction of each phase, *i.e.*

$$\rho = \sum_{i=1}^{n_p} X_i \rho_i \quad \text{and} \quad C_p = \sum_{i=1}^{n_p} X_i C_{p,i} \frac{\rho_i}{\rho}. \quad (6.21)$$

The generated specific internal energy (per unit of volume) is given as

$$\dot{q} = \sum_{i=1}^{n_p} \Delta H_i \frac{\delta X_i}{\delta t}, \quad (6.22)$$

where  $\Delta H_i$  is the enthalpy change of phase  $i$  during the phase transformation process. However, in these models, the general heat transfer equation (Eq. 6.20) is usually used.

The kinetics of phase transformation must provide the equations that define the rate of nucleation and growth of new phases. The Avrami equation [31] is generally used to describe the rate of nucleation and growth of a newly formed phase decomposed from the austenite grain boundaries. However, the equation, originally proposed for isothermal conditions, must be applied in non-isothermal transformation kinetics by assuming that the temperature evolution is divided into small time steps. Therefore, the thermo-kinetics can be considered as isothermal (isokinetic reaction). The classical JMAK approach (see Eqs. 6.9–6.11) covers isothermal diffusional transformations and uses the thermodynamical equilibrium fraction ( $X^{eq}$ ) for each phase as a parameter. This value is determined from the

equilibrium phase diagram with known temperature and chemical composition (see Fig. 6.8). The equation is then extended to non-isothermal transformation kinetics considering the cooling curve as a combination of a sufficiently large number of isothermal reaction steps.

The equation for the transformed fraction  $X_i$  of phase  $i$  at the current time step  $n + 1$  can then be defined as

$$X_{i,n+1} = X_{i,n+1}^{\text{eq}} (1 - e^{-k_i(t^{\text{eq}} + \Delta t)^{n_i}}) \quad (6.23)$$

from an equivalent transformation time  $t^{\text{eq}}$ , which defines the phase fraction  $X_{i,n}$  of the previous time step, and time increment  $\Delta t$ . The thermodynamical equilibrium fraction of phase  $i$  evolves with the process and it is dependent of the material's chemical composition and temperature. Numerically, the parameter  $X_{n+1}^{\text{eq}}$  works as a boundary for the fraction of the phase  $i$  and can be determined from continuous cooling diagrams, as the one seen in Fig. 6.12. The parameters  $k_i$  and  $n_i$  are empirically obtained constants that also depend on the nature of nucleation and growth, chemical composition and temperature. The temperature dependence of these parameters minimizes the errors that can arise by the use of an extension of the JMAK equation for continuous cooling. The equivalent transformation time is calculated by

$$t^{\text{eq}} = \left[ -\frac{1}{k_i} \ln \left( 1 - \frac{X_{i,n}}{X_{i,n}^{\text{eq}}} \right) \right]^{1/n_i}, \quad (6.24)$$

that should be used for the diffusion phase fraction that is increasing, such as ferrite, pearlite or bainite. During processes of autenitization, a similar equation is used for the volume fraction of austenite.

As mentioned, the martensite phase transformation is difusionless. Therefore, the martensite transformation can be described using the Koistinen-Marburger relationship [32], where the martensite volume at a given temperature  $T$  below the martensite starting temperature  $M_s$  is given as

$$X_m = X_{\text{ra}} (1 - e^{-c(M_s - T)}), \quad (6.25)$$

where  $X_{\text{ra}}$  is the volume fraction of retained austenite and  $c$  is a material parameter that is considered as 0.01 for steels. As the difusionless equation does not take time into account, it can be considered as an increment of the martensite volume fraction until all the remaining retained austenite is transformed.

The temperatures at which the phase transformations start are dependent of the chemical composition and grain size. The values for these temperatures can be found in diagrams as Fig. 6.8 or can be calculated analytically using empirical functions of the chemical composition and grain characteristics.

The transformations in steel do not occur in isolation. However, as a simplification, homogeneous models generally consider sequential transformations. For simultaneous transformations there are some works based in the Avrami theory

such as the works of Jones and Bhadeshia [33] (model for simultaneous transformations of austenite) or Rios and Villa [34] (a kinetic methodology suitable for the case in which transformations take place simultaneously or sequentially and that are able to deal with position-dependent quantities and not relying on extended volume).

The homogeneous transformation models use thermomechanical models not so different of the macroscopic models used in classical approaches. However, these use the phase transformation information coming from the microstructure and defined in the previous equations. In viscoplasticity, the strain decomposition is the same as for the elasto-plastic model. Nonetheless, due to the phase transformation phenomenon, the strain decomposition must account for the volumetric strain increment coming from the transformation itself. Therefore,

$$\dot{\boldsymbol{\varepsilon}} = \dot{\boldsymbol{\varepsilon}}^e + \dot{\boldsymbol{\varepsilon}}^p + \dot{\boldsymbol{\varepsilon}}^{\text{tr}}. \quad (6.26)$$

This tensor, which also includes the thermal expansion of the material, is a linear mixture of strain increments of existing phases. Consequently, it can be written as

$$\dot{\boldsymbol{\varepsilon}}^{\text{tr}} = \sum_{i=1}^{n_p} \left[ \frac{1}{3} \rho_{i,\text{ref}} \left( \frac{1}{\rho_{i,n+1}} - \frac{1}{\rho_{i,n}} \right) \right] \mathbf{I}, \quad (6.27)$$

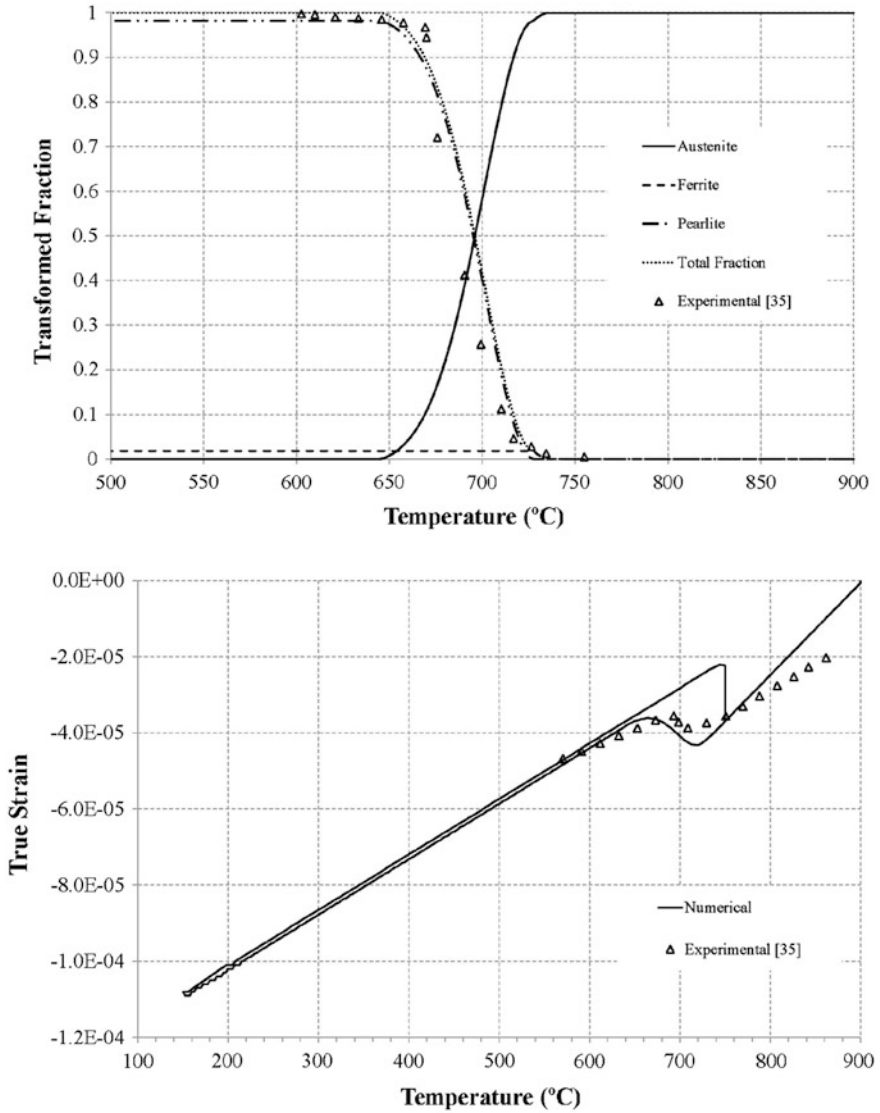
where  $\rho_i$  is the density of the phase  $i$  and  $\mathbf{I}$  is the identity tensor. Due to the presence of multiple material phases, the isotropic macroscopic stress can be expressed as a function of all existent phases by

$$\bar{\boldsymbol{\sigma}}_{\text{iso}} = \sum_{i=1}^{n_p} X_i \bar{\boldsymbol{\sigma}}_{\text{iso},i}(\bar{\boldsymbol{\varepsilon}}, T). \quad (6.28)$$

In this global model, different models can be used for each microstructural phase, from the pure elastic to more complex thermoelastic-viscoplastic models. In these models, the equivalent strain  $\bar{\boldsymbol{\varepsilon}}$  and the temperature must be taken in account. It is usual to consider the martensite material as elastic material considering that its yield stress is very high relatively to the other phases' properties.

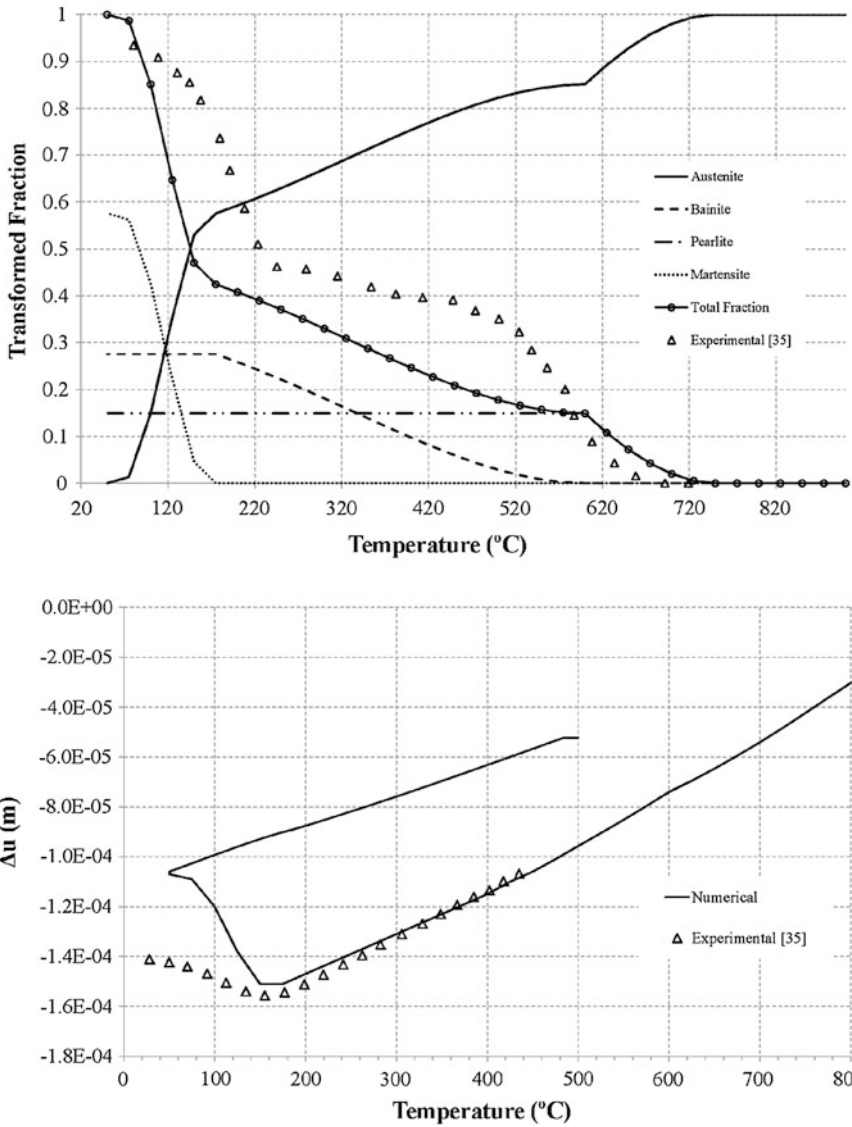
#### **6.4.2 Dilatometer Simulation, Austenitization and Annealing: Examples**

A continuous cooling rate of 5 °C/s and subsequent autenitization was applied to a dilatometer with initial temperature of 900 °C. In a second example, a continuous cooling rate of 50 °C/s followed by an annealing treatment is applied to the dilatometer. The behavior of a homogeneous material model presented in the previous section is analyzed and compared with experimental results [34].



**Fig. 6.13** Transformed phase fraction for a continuous cooling rate and autenitization. Dilatometric curve

Figure 6.13 shows the phase transformations and dilatometric curve, respectively, for the first example considered. The slope change at about 720 °C shows the thermal expansion due to phase transformation. As can be seen, the numerical results are in agreement with the experimental ones showing the good results obtained by this type of modern engineering analysis.



**Fig. 6.14** Transformed phase fraction for the continuous cooling and annealing. Dilatometric curve

In Fig. 6.14 the predicted and experimental transformed phase fraction and the dilatometric curve for the second example are shown. Although the model does not accurately predict the transformed phase fraction, it presents a good approximation in terms of shape and slope of the curves.

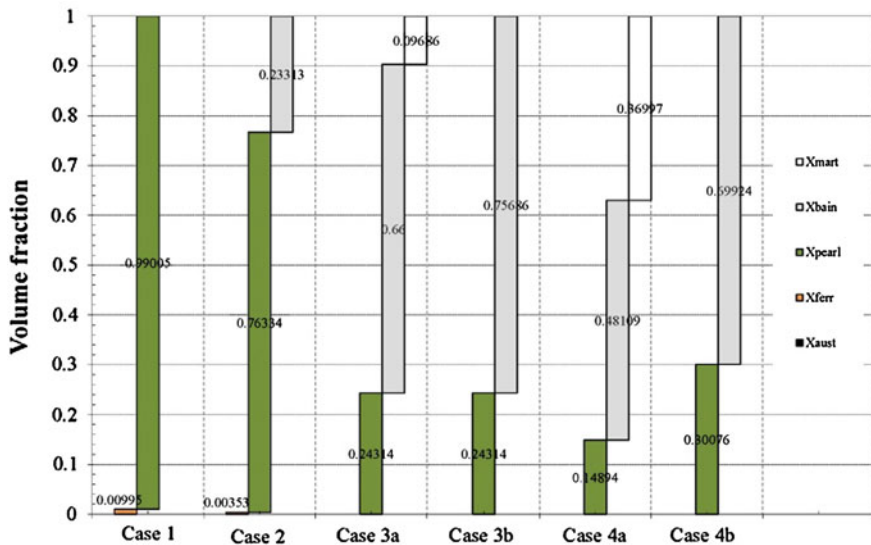


Fig. 6.15 Final microstructure composition for the heat treatment cases

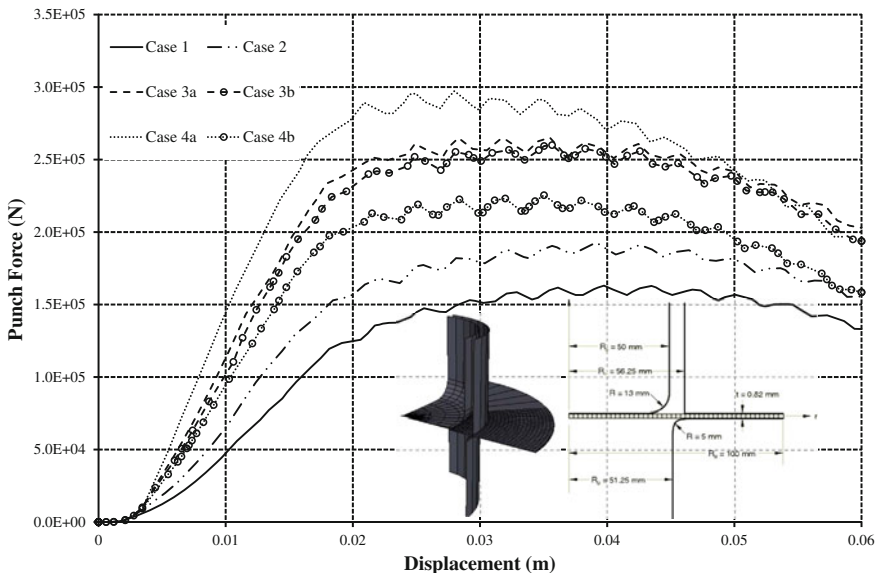
### 6.4.3 Heat Treatments for Metal Forming Operations: Examples

For industry engineers, the benefits of having an accurate model prepared for multiphase steels are not valid if the model cannot be used in the simulation of thermomechanical technological process, such as warm metal forming. Therefore, in this example, the presented model will be used in complex simulation involving contact, friction, springback phenomena, etc. Simulation of stamping processes does not create major difficulties, i.e. geometry of products and forming defects like wrinkles or ruptures can be predicted with good accuracy. However, FEM simulations encounter problems when complex constitutive models are used. Additionally, the calculation of final displacements due to springback when using a multiphase model is used is a complex task.

In this example, different homogeneous heat treatments were considered in order to obtain different multiphase steels. These heat treatments consist in quenching and subsequent annealing. However, different cooling rates were considered (50–16 °C/s). Figure 6.15 depicts the volume fraction for each heat treatment. The material volume fraction after the constant cooling are represented as cases 3a and 4a for cases 3 and 4, respectively. As can be seen, the subsequent annealing in these cases lead to the transformation of the martensite volume fraction and, therefore, demonstrating the metastable behavior of this phase.

The heat treatments analyzed are applied to model the material of the blank in the deep drawing of a cylindrical cup. The punch force versus displacement for



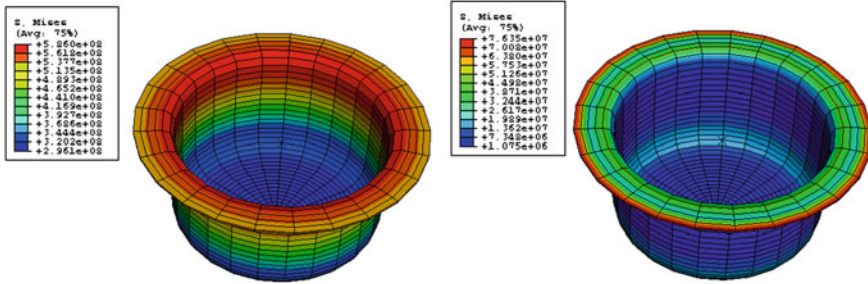


**Fig. 6.16** Punch force versus punch displacements for the diverse heat treatments. Geometry of the drawing tool

each case is depicted in Fig. 6.16. The difference of the punch force–displacement curve between 3a and 3b is small due to the small content of martensite that is transformed in the annealing treatment, not changing significantly the overall stiffness of the material. However, the 4a–b transformation has led to a significant change in the highest observed punch force. In this case, the martensite was also transformed into pearlite due to the higher annealing temperature and, consequently, a softening of the material. The higher punch force was registered in case 4a. This result can be attributed to the higher martensite and bainite volume fraction which leads to a higher global hardness of the material. Figure 6.17 present the stress distribution in the formed part before and after the springback stage.

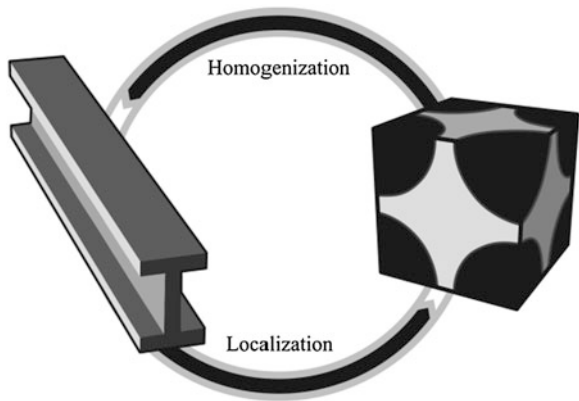
### 6.4.4 Multiscale Models Using Homogenization and RVEs

The mechanical properties of multi-phase (MP) steels are the result of their microscopic structure. Therefore, micromechanical analysis of MP steels allows the study of the overall behavior of the material based on the known properties of its constituents. However, the constituents of a MP steel can only be observed and analyzed in a micro-scale and, generally, they show a random distribution.



**Fig. 6.17** Stress field of the von Mises equivalent stress after the deepdrawing and the springback stages

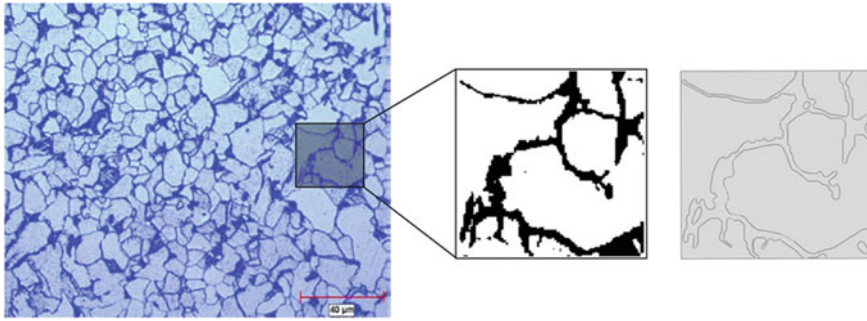
**Fig. 6.18** Homogenization and localization in multiscale models



Therefore, it is possible to consider a volume element of the material as representative of the material.

In micromechanical approaches, different length scales are considered (the microscale and the macroscale) and each of these scales contributes to the stress and strain field in an inhomogeneous material. The scales are assumed to be sufficiently different so that the stress and strain fields in the microscale only influences the macroscale behavior through their volume averages. Additionally, the gradients of the stress and strain fields and compositional gradients of the macroscale are not significant at the microscale.

In Periodic Microfield Approaches or Representative Volume Elements (RVE) models, the inhomogeneous material is approximated by an infinitely extended model with a periodic phase arrangement which is under the action of far field mechanical loads or uniform temperature fields. These approaches can be used to perform the material characterization of inhomogeneous materials in the nonlinear range and allow to obtain detailed information on the local stress and strain fields.



**Fig. 6.19** Representative Volume Element or representative unit cell obtained with a micrograph

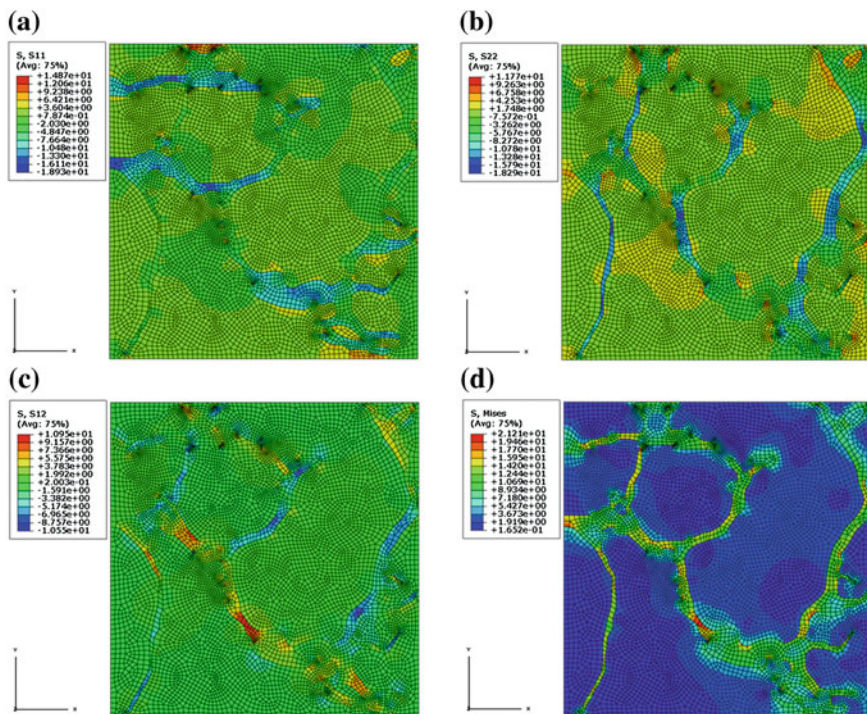
In the RVE models, each volume element contributes the same increment of the displacement vector, leading to a linearly varying homogenized displacement. To allow a realistic representation of the microgeometries, it is necessary to choose and generate a suitable volume element and appropriate boundary conditions. (Fig. 6.18)

It should be noticed that the use of these models requires that specific boundary conditions are prescribed to assure the perfect tiling between adjacent volume elements. The periodic boundary conditions make use of the translatory symmetries of a geometry to handle any physically valid deformation state of the cell. By pairing and linking the corresponding degrees of freedom in opposite faces, one can assure that the cells fit into each other, both in deformed and undeformed shape.

#### ***6.4.5 Cooling Tests Using Multiscale Models: Example***

In this example, a dual-phase (DP) steel with 30 % of martensite is used. As shown in Fig. 6.19, a DP Steel consists of a ferrite matrix containing a hard second phase, usually islands of martensite. In this thermomechanical test, a constant cooling was applied to a RVE model, with a starting temperature of 150 °C. This initial temperature was chosen because the martensite starting temperature was defined as 157 °C and, being a metastable phase, higher temperatures would dissolve the martensite phase.

In Fig. 6.20a, it can be seen that the higher values are located in the martensite ramifications reaching a maximum of  $\sim 21$  MPa. Fig. 6.20b and c represent the stresses along the  $x$  and  $y$  directions, respectively. It can be seen that the higher compressive stresses are located in or close to the martensite, while most part of the ferrite is under tensile stress.



**Fig. 6.20** **a**  $\sigma_{xx}$ , **b**  $\sigma_{yy}$ , **c**  $\sigma_{xy}$  and **d** equivalent stress in the cooling of the microstructure of the multiscale model

### 6.4.6 Other Modern Models

The previous models do not reproduce both grain boundary and phase transformation effects. In order to include both phenomena, the previous multiscale model can couple a formulation that is able to calculate phase transformation in each finite element. However, the structure of finite element method (FEM) is not well suited for simulation of diffusion phenomena. Other numerical methods, such as phase-field methods (PFM), are more suitable for this kind of calculations. Therefore, some models couple FEM with PFM in the microstructural scale.

In a FEM-PFM model, and within the microstructural representation as representative volume element (RVE), the material grains (particles) are explicitly discretized by finite elements. Therefore, each finite element is considered as a fraction of a grain. Thermomechanical, metallurgical (including transformation kinetics, grain size, etc.) and diffusion equations are implemented at the microlevel as FEM and PFM in order to reproduce these phenomena. (Fig. 6.20)

A finite element representing an elementary part of the material has also an energetic (thermodynamics) evolution equation attached given by the PFM (thermodiffusion equations that reproduce grain dynamics including recovery,

recrystallization and grain growth). This energetic evaluation (including temperature and thermal gradients, Gibbs free energy, activation energy, etc.) defines the metallurgical phase-change during the heat treatment (not only considering phase starting temperatures). Finite element surfaces are used, when defined, as grain boundaries allowing grain regrowth.

These models also allow the study of non-uniform deformation within the individual crystals of a polycrystal of steels. For this case, even anisotropy levels can be analyzed microstructurally and used macroscopically. Such approaches are also used to study grain-scale heterogeneous deformations that lead to the formation of macroscopic shear bands in plane strain compression.

## 6.5 Trends in Thermomechanical Modeling of Multiphase Steels

Although there were significant developments in the numerical modeling of thermomechanical processes of multiphase steels, there are many challenges yet to be solved. Here, some are presented.

In the field of heat treatments several issues should be addressed in order to reproduce the real process numerically. The aim of heat treatments is to achieve material properties and microstructures that could not be obtained otherwise. It is due to heat treatments that new impressive materials, such as multi-phase steels, can be manufactured. However, strain and microstructural morphology control during heat treatment processes still resorts, almost exclusively, to empiric experience. This is also due to the large difficulty of following experimentally all phenomena (micro and macro) of the thermomechanical process and to assess the resultant residual stresses without destructive testing techniques. It is true that this fact stresses out the need for numerical analyses of heat treatment processes as well as thermomechanical and metallurgical models that are able to accurately predict the phenomena and evolution of the behavior of the material. However, without more experimental information it is very difficult to develop more accurate numerical models. Additionally, the thermal boundary conditions of these processes are not fully understood. The temperature-dependent surface heat transfer coefficient during heat treatments and steel manufacture processes is a function of the geometry, slope and roughness of the material, temperature and temperature rate, etc. Without the precise knowledge of these parameters, the heat treatment processes will not be precisely reproduced by numerical simulation.

Other challenge is the use of multiple micro and macro scales depending on the phenomena one wishes to analyze. One of the problems of multiscale analyses is the selection of the appropriate scale. The mechanical behavior of MP-steels is attributed to their microstructure. Consequently, modeling the mechanical behavior of the MP-steel materials has to be done based on microstructural levels, which are many,

but only at the phase level the material can be considered almost as an isotropic continuum. Phenomena related to dislocation are analyzed at an atomic scale. However, grain boundaries phenomena are addressed at grain size scale. Nevertheless, macroscopic properties are also required for industry engineers. In multi-scale approaches, the automatic selection of an appropriate scale is a challenge.

Predictions for manipulating microstructure to attain enhanced performance demand models that distinguish between nucleation, migration, absorption/desorption, trapping, and bypass or annihilation of dislocations at various material length scales that manifest work hardening behavior. Most of these phenomena cannot be considered in isolation, which is a hallmark of plasticity. These are extremely coupled phenomena, in general, with important effects of both short and long range character attributed both to the physics of dislocation cores and long range interactions of dislocation arrays. Therefore, a challenge would be the development of a model that could address all these phenomena in a totally coupled structure.

**Acknowledgments** Part of his work was co-financed by the Portuguese Foundation for Science and Technology via project PTDC/EME-TME/122287/2010 and by FEDER via the “Programa Operacional Factores de Competitividade” of QREN with COMPETE reference: FCOMP-01-0124-FEDER-020517.

## References

1. Callister WD, Rethwisch DG (2012) Fundamentals of materials science and engineering an integrated approach. John Wiley & Sons, New York
2. Christian JW (2002) The theory of transformation in metals and alloys. Pergamon Press, Oxford
3. Campbell FC (2008) Elements of metallurgy and engineering alloys, ASM International
4. Porter DA, Easterling E (1992) Transformations in metals and alloys. Chapman & Hall, London
5. Shackelford JF (2008) Introduction to materials science for engineers, 7th edn. Prentice Hall, New Jersey
6. Gür CH, Pan J (2009) Handbook of thermal process modeling of steels. CRC Press, Boca Raton
7. Balluffi RW, Allen SM, Carter WC (2005) Kinetics of materials, department of materials science and engineering. MIT, Wiley-Interscience, Cambridge
8. Cahn RW, Hanssen P (1996) Physical metallurgy, vol I. North-Holland, Amsterdam
9. Kolmogorov AN (1937), Izv Akad Nauk SSR Ser Fiz Mat Nauk 3:355 (in Russian)
10. Johnson WA, Mehl RF (1939) Trans Am Crystallogr Assoc 135:416
11. Avrami M (1941) J Phys Chem 9:177
12. Sumpter ML (2004) Johnson-Mehl-Avrami kinetics of intracellular ice formation in confluent tissue constructs. PhD Thesis, Georgia Institute of Technology
13. Krauss G (2005) Steels: processing, structure, and performance, ASM International
14. Bhadeshia HKDH, Honeycombe R (2006) Steels: microstructure and properties, 3rd edn. Elsevier
15. Sourmail T (2004) Simultaneous precipitation in austenitic stainless steels, PhD Thesis, University of Cambridge

16. Bhadeshia HKDH (1999) Kinetics of simultaneous transformations, Proceedings of solid-solid phase transformations'99, (JIMIC-3), Kyoto, Japan, 1445. Lecture available <http://www.msm.cam.ac.uk/phase-trans/teaching.html>
17. Mittemeijer EJ (2010) Fundamentals of materials science—the microstructure—property relationship using metals as model systems. Springer, Berlin
18. Brown SB, Kwon HK, Anand L (1989) An internal variable constitutive model for hot working on metals. *Int J Plast* 5:95–130
19. Anand L (1985) Constitutive equations for hot working of metals. *Int J Plast* 1:213–231
20. MacKenzie AC, Moakler M (1973) On the relaxation of residual stress fields by thermal stress relief. In: 2nd international conference pressure and vessel technology, San Antonio, Texas pp. 1167–1178
21. Hossain S, Truman CE, Smith DJ, Daymond MR (2006) Application of quenching to create highly triaxial residual stresses in type 316H stainless steels. *Int J Mech Sci* 48(3):235
22. Andrade-Campos A, Teixeira-Dias F (2007) Numerical analysis of triaxial residual stresses in quenched 316H stainless steel. *Mater Sci Forum* 553:7–14
23. Bartels S, Carstensen C, Hackl K, Hoppe U (2004) Effective relaxation for microstructure simulations: algorithms and applications. *Comput Methods Appl Mech Eng* 193:5143–5175
24. McDowell LD (2010) A perspective on trends in multiscale plasticity. *Int J Plast* 26:1280–1309
25. Andrade-Campos A (2010) The influence of the material heat treatment and phase transformations in forming processes. *Int J Mater Form* 3:1139–1142
26. Caseiro JF, Oliveira JA, Andrade-Campos A (2011) Thermomechanical modelling strategies for multiphase steels. *Int J Mech Sci* 53:720–733
27. Wei YJ, Anand L (2004) Grain-boundary sliding and separation in polycrystalline metals: application to nanocrystalline fcc metals. *J Mech Phys Solids* 52:2587–2616
28. Brassart L, Doghri I, Delannay L (2009) Self-consistent modeling of DP steel incorporating short range interactions. *Int J Mater Form* 2:447–450
29. Huang C-J, Browne D (2006) Phase-field model prediction of nucleation and coarsening during austenite/ferrite transformation in steels. *Metall Mater Trans A* 37:589–598
30. Qin RS, Bhadeshia HKDH (2010) Phase field method. *Mater Sci Technol* 26:803–811
31. Avrami MA (1939) Kinetics of phase change-I. General theory. *J Chem Phys* 7:1103–1109
32. Koistinen DP, Marburger RE (1959) A general equation prescribing the extent of the austenite-martensite transformation in pure iron-carbon alloys and plain carbon steels. *Acta Metall* 7:59–60
33. Jones SJ, Bhadeshia HKDH (1997) Modelling the development of microstructure in power plant steels. *Acta Mater* 45:2911–2920
34. Rios P, Villa E (2011) Simultaneous and sequential transformations. *Acta Mater* 59:1632–1643
35. Ashby MF, Shercliff HR, Cebon D (2007) Materials: engineering, science processing and design. Butterworth-Heinemann, Oxford
36. Boyer (ed) (1977) Atlas of isothermal transformation and cooling transformation diagrams. American Society for Metals, p. 369

# Chapter 7

## Optimal Real-Time Management for Automated Production Lines

Joao Figueiredo

**Abstract** Today manufacturing is highly decentralized from the company headquarters to local Production sites, taking advantages from local resources such as labor costs, raw materials, infrastructures, etc. Economic scale factors motivate the implementation of large scale plants. These huge plants imply great number of distributed production lines. The decentralized environment increases the need for complex management tools that enable the complete on-line system supervision and control. This chapter presents an optimal real-time management system for automatic production plants, that incorporates low-level communications between processes—PLCs (Programmable Logic Controllers), a centralized optimization and supervisory tool that manages the entire system and remote communication between system's administrator and production processes (internet, GSM mobile communications, ...). This concept is a two layer management architecture where inner loops are performed by PLC networks and the outer optimization loop is performed by a SCADA system (Supervisory Control And Data Acquisition), combined with complex mathematical tools that enable the system to run real-time complex computational algorithms to optimize the hourly target production.

### 7.1 Introduction

Today with the worldwide communication development, the need for decentralized system management is increasing tremendously. This strategy reaches different fields, from agriculture, to industry, building automation [1–9].

In the economic field the big companies locate plants in low-cost lands and research and commercial activities close to customer industrialized lands.

---

J. Figueiredo (✉)  
CEM-IDMEC, University of Évora, R. Romão Ramalho,  
59 7000-671 Évora, Portugal  
e-mail: jfig@uevora.pt



In engineering, especially in the industrial field, the motivation to optimize resources forces the communication between decentralized systems in order to reach better allocation of resources, minimizing the waste of raw materials, reducing production costs, etc.

In this context the communication between remote systems is increasing tremendously, not only in new built systems but also when reengineering is brought to old systems. The reengineering of old plants brings new intelligence to these systems by introducing automation solutions in their processes. These new instrumentation systems (intelligent sensors and actuators) allow the plants to communicate their actual state to centralized control units, allowing a real-time decision process [10, 11].

In this chapter a two-level hierarchical control strategy is developed and applied to a manufacturing unit. In this strategy the first control loop is assigned to locally independent PLCs (Programmable Logic Controllers) connected in a network, exchanging control signals with an upper level control loop, managed by a SCADA supervisory system (Supervisory Control And Data Acquisition) [12]. The SCADA application performs an optimal allocation of production resources, taking into account the minimization of the assembly costs that are being on-line monitored.

The developed strategy is internet available allowing the remote management of the entire plant. Additionally to the internet remote access, each process can be connected independently through the GSM mobile network.

## 7.2 Plant Setup

The industrial plant studied in this chapter is composed by several automatic production lines.

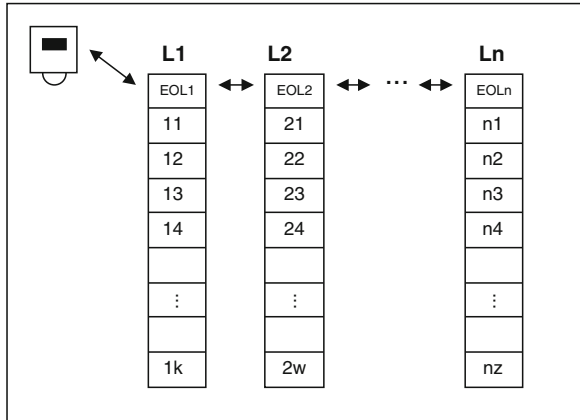
This structure provides flexibility to the plant manager as he can quickly adapt to different production volumes. Figure 7.1 shows an automatic plant composed by  $n$  production lines (L1–Ln) where each line is composed by several automatic machines with own processes, controlled by independent PLCs.

Each automatic line is connected to a general management platform through a typical PLC master–slave network. The master PLC communicates with a SCADA system that enlarges the system communication capabilities, allowing on-line monitoring and control, events recording, alarm management, etc. [12].

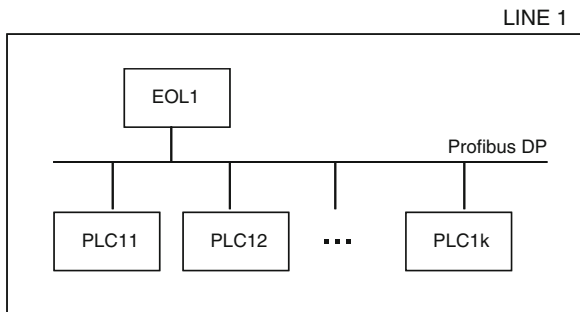
The set of machines in one production line constitutes an autonomous assembly unit. In each assembly unit there exists an industrial master–slave network connecting all PLCs from the several automatic machines. The EOL (End-Of-Line machine that runs the final tests) is configured as the master PLC, for each assembly unit. All other machines, in the same production line, are configured as PLC slaves (see Fig. 7.2).

To complete the overall management of the production plant, a second network layer has to be added in order to allow the communication between the different

**Fig. 7.1** Plant layout with n automatic production lines



**Fig. 7.2** Industrial profibus-DP master-slave network (automatic production line)

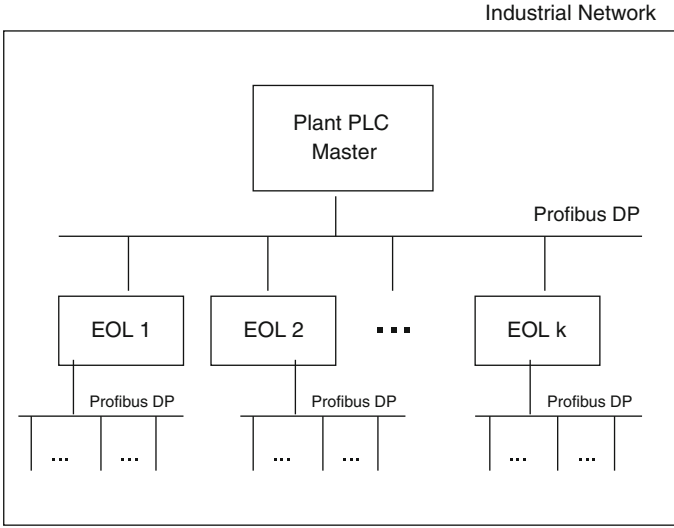


independent lines. This communication channel is implemented through the connection of all EOLs (End of Line machines) in a second Profibus hub, as illustrated in Fig. 7.3.

### 7.3 Control Strategy

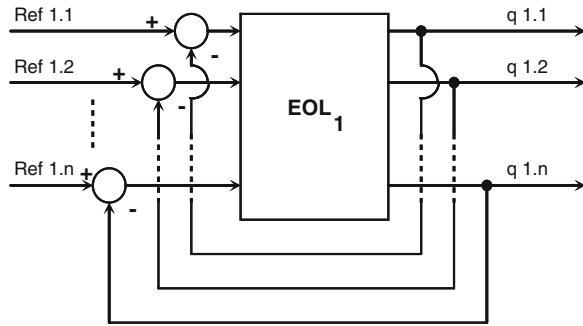
#### 7.3.1 Description of the Control Strategy

The strategy developed in this chapter is commonly known as a two-level hierarchical control as it integrates a first control loop that is managed by local PLCs with a second loop that is performed by a SCADA supervisory system monitoring globally the several distributed assembly lines. Figures 7.4 and 7.5 illustrate this control strategy: The inner control loop (first loop)—Fig. 7.4—and the outer control loop—Fig. 7.5.



**Fig. 7.3** Two-layer profibus-DP industrial network

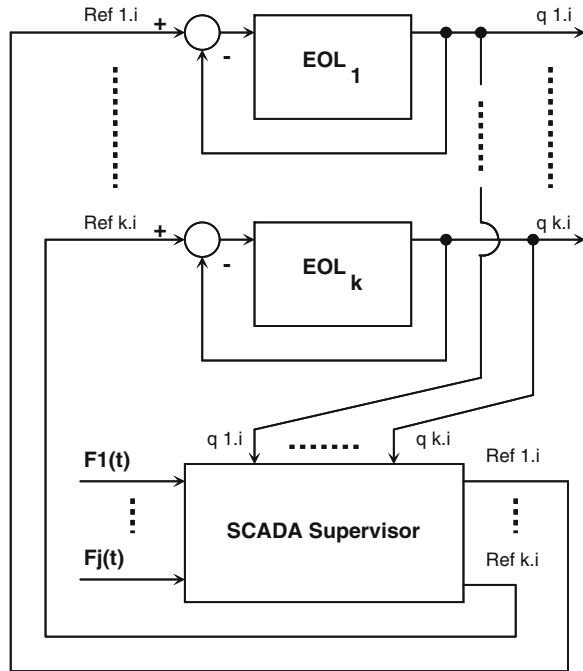
**Fig. 7.4** First loop control—master PLC from distributed production lines



Applying this strategy to a complex Production Plant that is instrumented and monitored through a SCADA supervisory system, one can manage globally the entire net of automatic lines that control locally each own assembly unit.

The upper level control law, having a global system overview, generates the set of references for each automatic line avoiding possible conflicts in the logistics of pre-manufactured parts. The input functions for the upper control loop are mentioned as global inputs  $(F_1(t), \dots, F_j(t))$ , which are worked out through the pre-optimization of a functional, in order to allocate the proper quantities to each Production Line, assembly machine, etc., regarding the constraints of the available resources (see Sect. 7.3.3).

**Fig. 7.5** Second loop control—SCADA supervisory control



### 7.3.2 PLC Network and SCADA Supervisor

The developed strategy is implemented through a 2-layer PLC network.

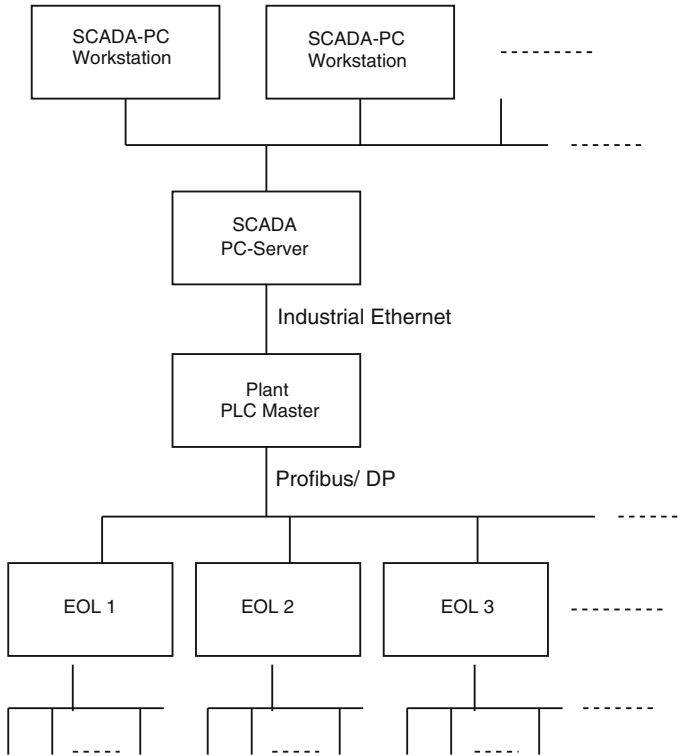
The first network layer is composed by the Plant master PLC and by the set of EOL-PLCs from each Production Line (Fig. 7.3).

The second network layer is composed by all other PLCs from each Production Line connected to the correspondent EOL-machine, as it was previously illustrated in Fig. 7.2, referring Production Line 1.

The complete Plant network includes a SCADA system connected to the Plant master PLC, through an Ethernet communication interface, which is illustrated in Fig. 7.6.

Each EOL hosts several control programs whose selection is made either locally, via an HMI (Human Machine Interface) or remotely, via the Plant Master PLC. The Plant Master PLC is connected to the server PC, where the SCADA application is running, via Ethernet protocol.

The server PC is simultaneous a SCADA server and an internet server, as the implemented SCADA application is web enabled. All process variables are available at the SCADA PC as these variables are on-line available through a Profibus/DP connection protocol (see Fig. 7.6).



**Fig. 7.6** PLC network and SCADA supervisor

### 7.3.3 SCADA Application Development

A Supervisory Control and Data Acquisition (SCADA) System is used as an application development tool that enables system integrators to create sophisticated supervisory and control applications for a wide variety of technological domains, mainly in the industry field. The main feature of a SCADA system is its ability to communicate with control equipment in the field, through the PLC network. As the equipment is monitored and data is recorded, a SCADA application responds according to system logic requirements or operator requests.

In the developed control strategy, the SCADA application performs the outer loop of the operational plant management. At this outer loop several complex control structures can be used to manage the overall system dynamics.

In this study an optimal allocation of production resources is performed taking into account the minimization of the assembly costs, that are being on-line monitored.

A traditional production allocation is made off-line, on a planning level, considering standard costs and standard capacities. Following this planning concept

the usual optimization process to allocate the best quantities to each line and shift, follows the minimization of the typical static functional:

$$\min J = \frac{\sum_i c_{1i}w_{1i} + \sum_j c_{2j}w_{2j} + \sum_l c_{3l}w_{3l}}{F_{Total}} \quad (7.1)$$

Subjected to:

$$\sum_i w_{1i} + \sum_j w_{2j} + \sum_l w_{3l} = F_{Total} \quad (7.2)$$

where:

$i$  = number of lines working on 1st shift;

$j$  = number of lines working on 2nd shift;

$l$  = number of lines working on 3rd shift;

$c_{1i}$  = 1st shift production cost, associated with line  $i$ ;

$w_{1i}$  = number of parts produced in line  $i$ —1st shift;

$c_{2j}$  = 2nd shift production cost, associated with line  $j$ ;

$w_{2j}$  = number of parts produced in line  $j$ —2nd shift;

$c_{3l}$  = 3rd shift production cost, associated with line  $l$ ;

$w_{3l}$  = number of parts produced in line  $l$ —3rd shift;

$F_{Total}$  = total demanded parts from customers.

This traditional process can lead to very poor results when quality costs, or unexpected line downtimes occur. Actually both of the previous referred problems occur very frequently and they are not taking into account on the traditional production planning. In order to improve the traditional methodology it is needed an automatic on-line output production monitoring, in order to reflect the line efficiency on the optimization process.

In order to solve this problem, this study follows a two step approach. On a first step, one uses the previous off-line method (Eqs. 7.1–7.2) to allocate the total production quantity per shift (Eq. 7.3):

$$\sum_i w_{\alpha i} = F_{\alpha} \quad (7.3)$$

On a second step, one uses the potential of the SCADA supervisory platform to integrate the monitoring of the real production costs on the optimization problem.

The selected functional to allocate the proper hourly quantities to each production line is presented below (Eqs. 7.4–7.9). The minimization of the functional  $J_k$  allocates the optimal hourly number of production parts to each line, for each shift (8 h shift production). Considering a shift  $\alpha$ , the on-line optimization problem becomes:

$$\min J_{\alpha|k} = \frac{\sum_i c_{\alpha i|k} w_{\alpha i|k}}{F_{\alpha|k}} \quad (7.4)$$

with the constraints:

$$c_{\alpha i|k} = \frac{w_{\alpha i|_{\max}} \times (cP_{\alpha i} + cI_i + cM)}{w_{\alpha i|_{k-1}}} \quad (7.5)$$

$$\sum_i w_{\alpha i|k} = F_{\alpha i|k} \quad (7.6)$$

$$F_{\alpha i|k} = \frac{F_{\alpha} - \sum_{k=1}^{H_p} \left( \sum_i w_{\alpha i|k} \right)}{H_r} \quad (7.7)$$

$$H_r = 8 - H_p \quad (7.8)$$

$$w_{\alpha i|k} \leq w_{\alpha i|_{\max}} \quad (7.9)$$

where:

$i$  = number of lines working on  $\alpha$  shift;

$c_{\alpha i|k}$  = Production cost of line  $i$ ,  $\alpha$ -shift, hour  $k$ ;

$w_{\alpha i|k}$  = Production parts allocated to line  $i$ ,  $\alpha$ -shift, hour  $k$ ;

$w_{\alpha i|_{k-1}}$  = good parts produced by line  $i$ ,  $\alpha$ -shift, in the previous hour ( $k-1$ );

$w_{i|_{\max}}$  = max. production capacity for line  $i$ ,  $\alpha$ -shift,

$cP_{\alpha i}$  = Personal costs for line  $i$ ,  $\alpha$ -shift/part,

$cI_i$  = Infrastructure costs for line  $i$ /part,

$cM$  = Material costs/part,

$H_r$  = remaining production hours,

$H_p$  = produced hours,

$F_{\alpha}$  = Total production required for shift  $\alpha$ ,

$F_{\alpha i|k}$  = number of parts to be produced in hour  $k$ ,

Analysing the minimization criterion, it is clear that the change in the hourly line production costs depends mainly on the number of good parts produced in the previous hour ( $w_{\alpha i|_{k-1}}$ ).

This value informs the algorithm about the production level of each line, with 1 h time-delay. This time-delay can be changed by the system administrator (1 h, 30, 15 min).

This model simulates actually what happens in real production plants, where the unexpected rising of the production costs is directly dependent on the quantity of good parts produced. The greater the scrap level, the line downtime, the non-availability of assembly parts, etc., the more the production costs rise.

The optimization algorithm implemented for the on-line allocation of line production, that runs on the SCADA outer loop, cannot be implemented directly on a usual SCADA system, as this complex controller needs mathematical operations that are not present at available SCADA systems. In this chapter it is

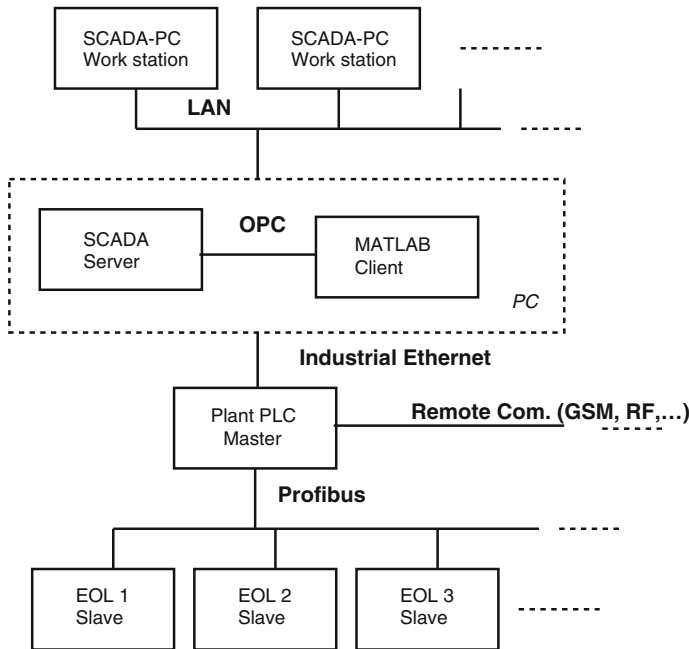


Fig. 7.7 Architecture of the plant-prototype

developed a strategy to couple the SCADA system [12] with the MATLAB software [13], where the optimal line hourly production is continuously recalculated.

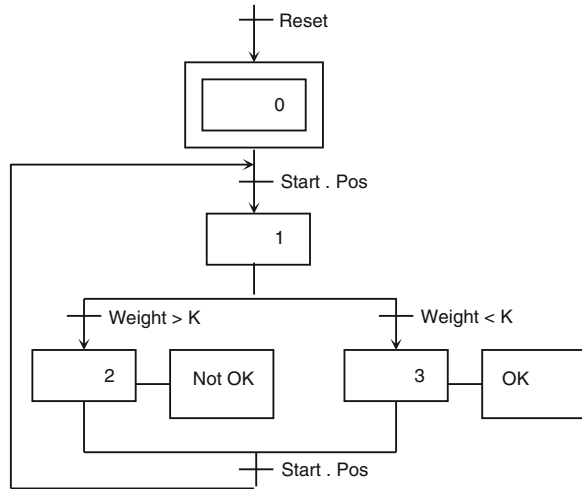
So it was necessary to develop a communication channel between the SCADA [12] and the Matlab software [13], where the optimization algorithm runs.

The used communication channel is the OPC protocol (Object Linking and Embedding –OLE- for Process Control) [14]. This protocol is based on standard specifications developed in 1996 by a task force from industrial automation [15]. This standard specifies the communication of real-time data among several control devices from different manufacturers. This protocol provides the exchange of data between two independent software programs (Server and Client) running simultaneous at the implemented platform. In this paper the Matlab software initiates the communication, as it is the Client and the SCADA software responds to Client's requests (Server attributions) [14, 15].

Figure 7.7 illustrates the communication flow that was developed to implement the optimization algorithm at the Outer loop control (Eqs. 7.4–7.9). In this figure we see the coexistence of four different communication protocols (LAN, OPC, Ethernet, PROFIBUS) working simultaneously at different levels of the developed platform.



**Fig. 7.8** Grafcet for weight measurement



### 7.3.4 PLC Application Development

At the inner loop of the developed strategy (PLC level), several algorithms have been developed. These algorithms were built using the Grafcet methodology—Sequential Function Chart [16]. The designed algorithms were implemented using the Ladder Diagram language [17]. Two main process algorithms were developed: Weight measurement and Dimension measurement. As an application example, Fig. 7.8 shows a simple process for selecting parts having a weight greater than  $K$  [N]. The Variable “Pos” informs that a part is available on the weighing-machine.

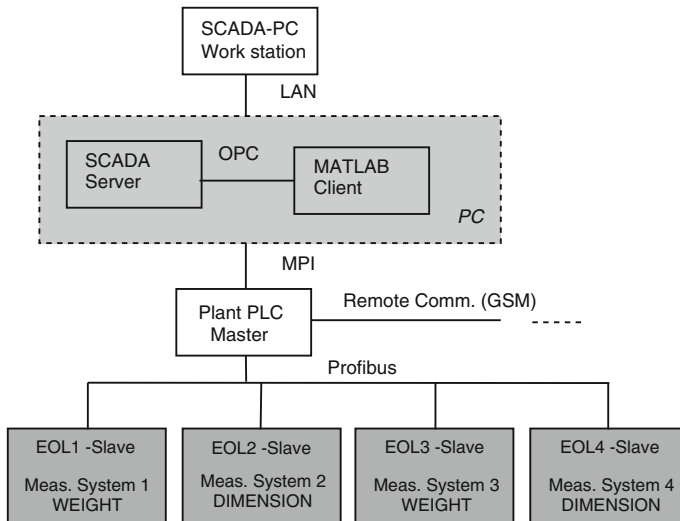
## 7.4 Experimental Setup

The developed application to monitor and control automatic Plants had been implemented on an experimental setup with the following software and hardware requirements.

The PLC network implemented had four PLCs: one PLC for each Line. Lines 1 and 3 measure Weight. Lines 2 and 4 measure Dimension. Figure 7.9 shows the architecture of the built Prototype.

### 7.4.1 Software Requirements

The Scada system was developed over the platform Siemens Simatic WinCC V7.0 [12]. The optimization algorithm runs on the MATLAB Optimization Toolbox [18].



**Fig. 7.9** Plant prototype architecture

The software used for the PLC programming was the Siemens Simatic Step 7 [17], with the Simatic S7 Prosave V5.5 [19] needed for the communication between the Scada system and the PLC network.

### 7.4.2 Hardware Requirements

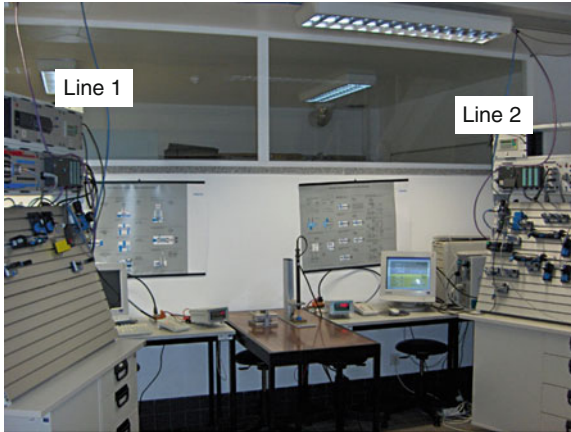
Referring hardware characteristics each PLC (Plant Master and EOL slaves) was composed by the following Siemens modules:

- Slot1 = Power supply PS 307 2A
- Slot2 = Processor CPU 315
- Slot4 = Communication module CP 342–5
- Slot5 = Digital card DI8/DO8xDC24 V/0,5A
- Slot6 = Digital card DI8/DO8xDC24 V/0,5A
- Slot7 = Analogue card AI4/AO2x8/8bit

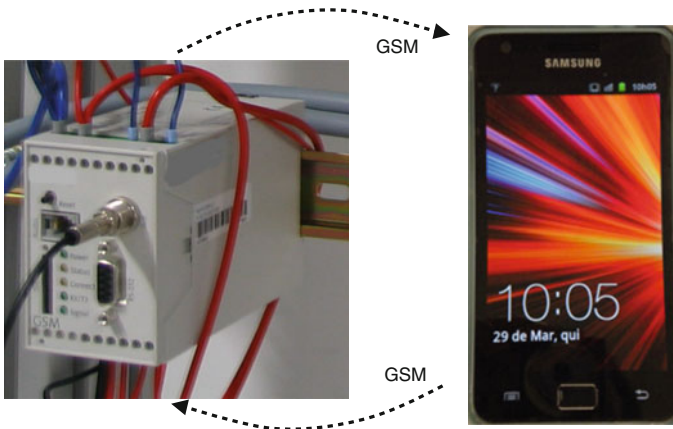
Additionally, the Plant Master PLC has a modem for GSM communication that provides the system capacity to communicate through the mobile phone network.

Figure 7.10 shows an overview of the implemented prototype. Figure 7.11 shows the GSM modem used for the system communication through the mobile phone network. Figure 7.12 shows the implemented architecture for the establishment of the communications: Plant Master PLC—GSM Modem—Mobile phone.

The sensors used were standard industrial sensors for weight and dimension. Their main technical characteristics are shown below:



**Fig. 7.10** Plant prototype with two production lines (EOL 1 and 2)

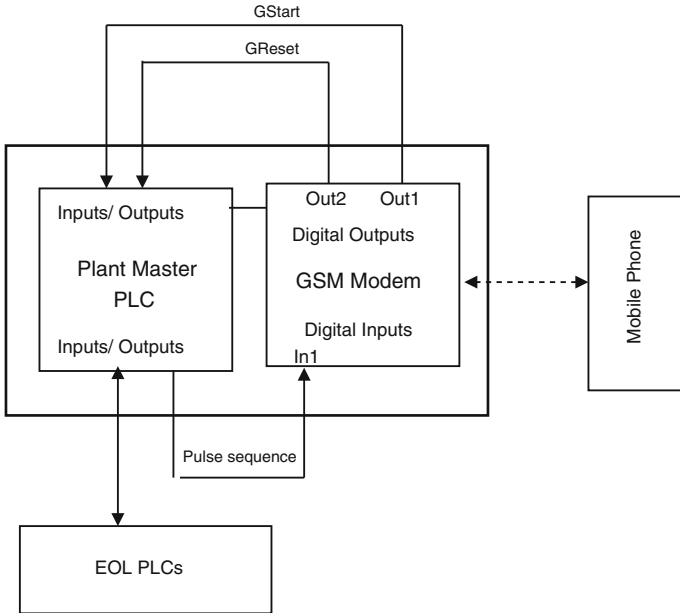


**Fig. 7.11** GSM modem for mobile phone communications

- weight sensor: Burster Ring Load cell 8438E (0...2000 N);
- dimension sensor: Burster Gaging Transducer DC-DC 87350 (0...50 mm);
- position inductive sensors: Festo SME-8-K-LED-24;
- position inductive sensors: Festo SMEO-4U-K-LED-24.

The actuators used were standard industrial pneumatic actuators. Their main technical characteristics are shown below:

- Cylinders type A: Festo ADVU 12-50-PA;
- Cylinders type B: Festo DSNU-10-80-P-A



**Fig. 7.12** Designed architecture for the communication GSM-EOL (PLC)

All PLCs responsible for dimension and weight measurements, communicate through an industrial Profibus-DP network [20]. This network was established through a dedicated communication module (Siemens CP 342-5), as mentioned earlier.

The two measurement systems (weight and dimensions) implemented in the present prototype are shown in Fig. 7.13.

### 7.5 Results

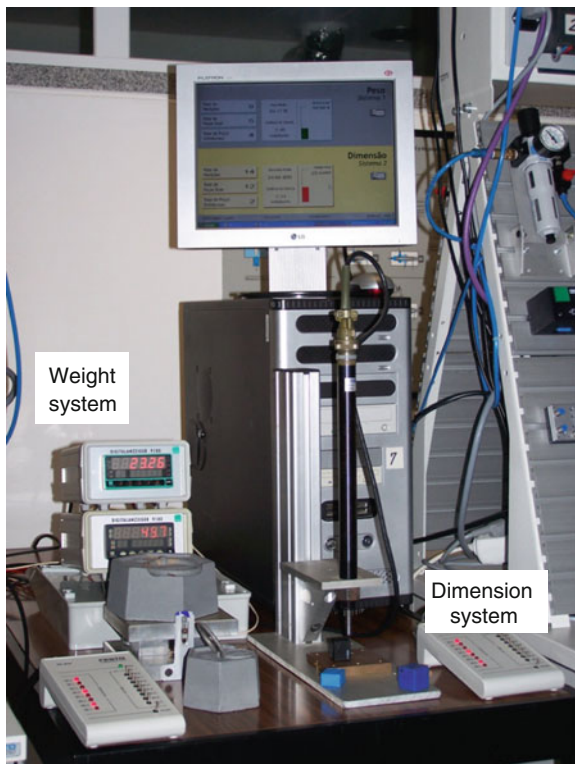
In this section two types of results are illustrated: Optimization Results and Operational Results. Each one has proper objectives to be fulfilled.

With the Optimization Results one tests the ability of the on-line management tool to integrate the low level information coming from the EOL-PLCs about the production output, and evaluating the minimization of the functional described in Sect. 7.3.3 (Eqs. 7.4–7.9), performing the optimal resources allocation of a distributed production Plant.

In these tests we evaluate also the compatibility of the several communication protocols present in the developed application (LAN, OPC, Ethernet, PROFIBUS).

The Operational Results show mainly the information made available at the several developed Graphical User Interfaces (GUI) of the application.

**Fig. 7.13** Measurement systems for dimension and weight



### ***7.5.1 Results from the Optimization Algorithm***

The optimization problem described in Eqs. 7.4–7.9, was solved through the MATLAB Optimization Toolbox, using the standard algorithm “fmincon” [18].

The tested assembly plant was composed by 4 automatic production lines, grouped in two sets of different processes (Weight and Measurement). A good part produced had to be processed by both processes (Weight and Measurement) without any prescribed sequence.

The main parameters considered for the field test are shown in Table 7.1. Table 7.2 illustrates the events considered to evaluate the performance of the developed on-line optimization system.

The obtained results for the 6 h test performed are illustrated in Tables 7.3 and 7.4 and in Figs. 7.14 and 7.15.

Observing the results we can assure the good performance of the developed methodology. These results illustrate the important role of on-line management of production processes in order to minimize production costs. In fact, taking only in account the traditional off-line planning methodology, we came easily to the wrong production allocation, which would lead to allocate the maximal production

**Table 7.1** Experimental parameters

	EOL1	EOL2	EOL3	EOL4
	Weight	Dimension	Weight	Dimension
$w_{zilmax}$	1800	2200	1600	1500
$cP_{zi}$ (€/part)	0.20	0.18	0.35	0.26
$cl_i$ (€/part)	0.15	0.07	0.10	0.06
$cM$ (€/part)	0.42	0.42	0.42	0.42
$F_z$	13200			
mean $F_z$ /hour	2200			

**Table 7.2** Experimental events

	EOL1	EOL2	EOL3	EOL4
	Weight (%)	Dimension (%)	Weight (%)	Dimension (%)
Real efficiency (%)*; $H_p = 1$	99	98	99	99
Real efficiency (%)*; $H_p = 2$	<b>55</b>	99	99	99
Real efficiency (%)*; $H_p = 3$	<b>34</b>	99	99	99
Real efficiency (%)*; $H_p = 4$	99	<b>46</b>	99	99
Real efficiency (%)*; $H_p = 5$	99	99	99	99
Real efficiency (%)*; $H_p = 6$	99	99	99	99

\*Real efficiency = produced parts/allocated parts

level to lines EOL1 and EOL3 (Weight 1 and Measurement 2) which have the best nominal efficiency values (see Table 7.1). But, when we consider the real plant, with technical problems (line downtimes) and quality costs (scrap level), then those theoretically more efficient lines become worst than the other lines that are running normally. The rise of the production costs in lines EOL1 and EOL2, motivated by the less effective produced parts, turns Lines EOL3 and EOL4 to be more efficient at different time intervals, as it is illustrated in Figs. 7.14 and 7.15.

## 7.5.2 Operational Results

The SCADA system used to implement this monitoring and control strategy permits the selective access of the application, depending on the user's responsibility degree. In this study we developed three user levels: Operators, Supervisors and Administrators.

The Scada main interface between the system and the user are the application images. The image building is a functionality of all SCADA systems and its main function is to allow the user a quick visual identification of all system functional characteristics. An easy identification of the system Inputs and Outputs permits the user an effective monitoring and a quick actuation on the process when it is necessary.

**Table 7.3** Experimental results—weight

	EOL1			EOL3			Mean hourly demand	Hourly allocated qty. (EOL1 + EOL3)
	Hourly allocated	Hourly produced	Efficiency (%)	Hourly allocated	Hourly produced	Efficiency (%)		
$H_p = 1$	1800	1780	98.9	400	398	99.5	2200	2200
$H_p = 2$	1800	992	55.1	422	418	99.1	2200	2222
$H_p = 3$	1312	440	33.5	1600	1587	99.2	2200	2912
$H_p = 4$	1394	1392	99.9	1600	1590	99.4	2200	2994
$H_p = 5$	1800	1793	99.6	612	610	99.7	2200	2412
$H_p = 6$	1800	1790	99.4	409	407	99.5	2200	2209

Hourly allocated =  $w_{zilk}$

Hourly produced =  $w_{zilk-1}$

**Table 7.4** Experimental results—dimension

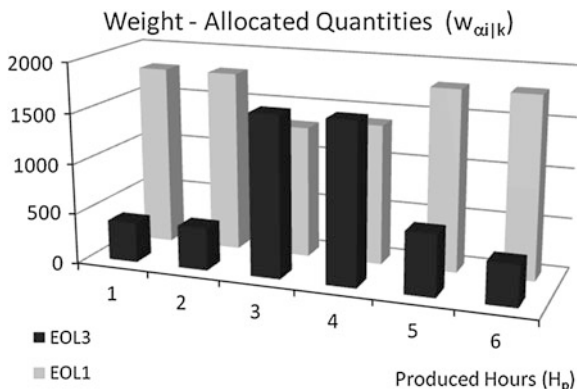
	EOL2		EOL4		Mean hourly demand	Hourly allocated qty. (EOL1 + EOL3)
	Hourly allocated	Hourly produced	Efficiency (%)	Hourly allocated		
$H_p = 1$	2200	2155	98.0	0	2200	2200
$H_p = 2$	2200	2172	98.7	45	2200	2245
$H_p = 3$	2200	2189	99.5	29	2200	2229
$H_p = 4$	2200	1007	45.8	12	2200	2212
$H_p = 5$	1193	1187	99.5	1500	2200	2693
$H_p = 6$	2200	2173	98.8	741	2200	2941

Hourly allocated =  $w_{zilk}$

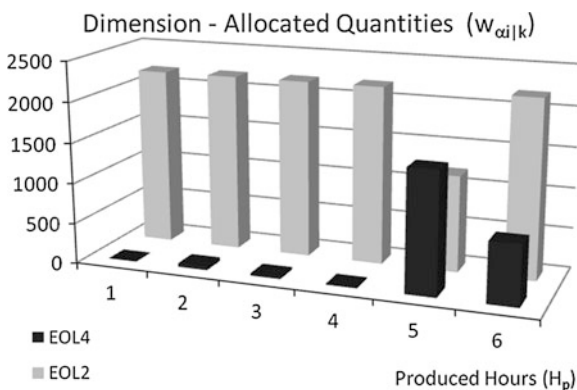
Hourly produced =  $w_{zilk-1}$



**Fig. 7.14** Overview of the complete production at EOL1 and EOL3



**Fig. 7.15** Overview of the complete production at EOL4 and EOL2



The developed application used important features that we named animated images. These images change geometry characteristics and colours when their digital variables change the state (on/off).

Several SCADA menus were built. The main characteristics of a Scada Menu is to be simple, explicit and quick on transmitting the information to the operator or to the Production Line manager.

Some of the developed Graphical User Interfaces (GUI) are shown in Figs. 7.16 and 7.17.

As this SCADA platform is web enabled, all the GUI displayed data is also on-line accessible through the internet.

In Fig. 7.16 it is shown an overview of the complete Plant production, with the cumulative information regarding the two production lines (EOL1-Weight measurement and EOL2-Dimension measurement). The on-line available information referring statistical data from both production lines is: number of total produced parts, number of good parts, number of bad parts, measurement average value. This menu presents additionally the last measurement value, indicating the quality criteria (good/bad part).

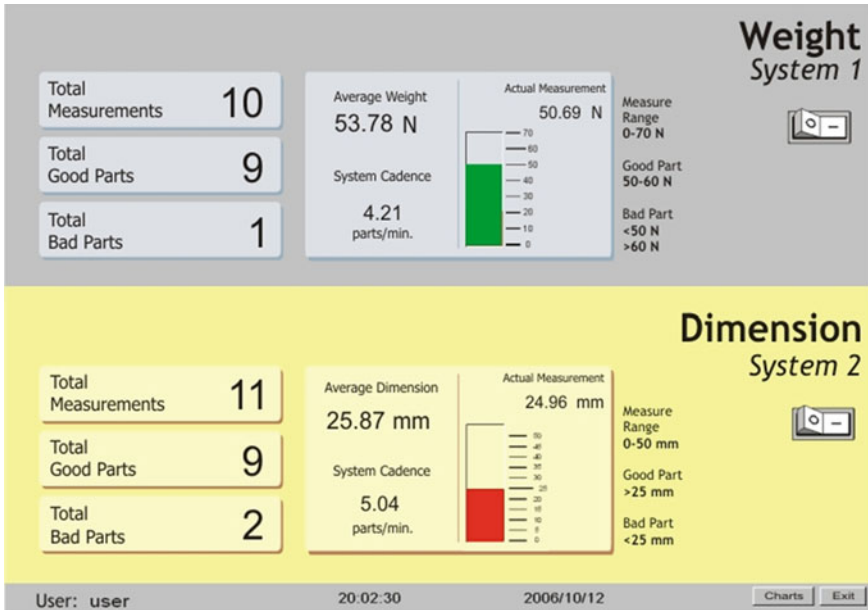


Fig. 7.16 Main graphical user interface

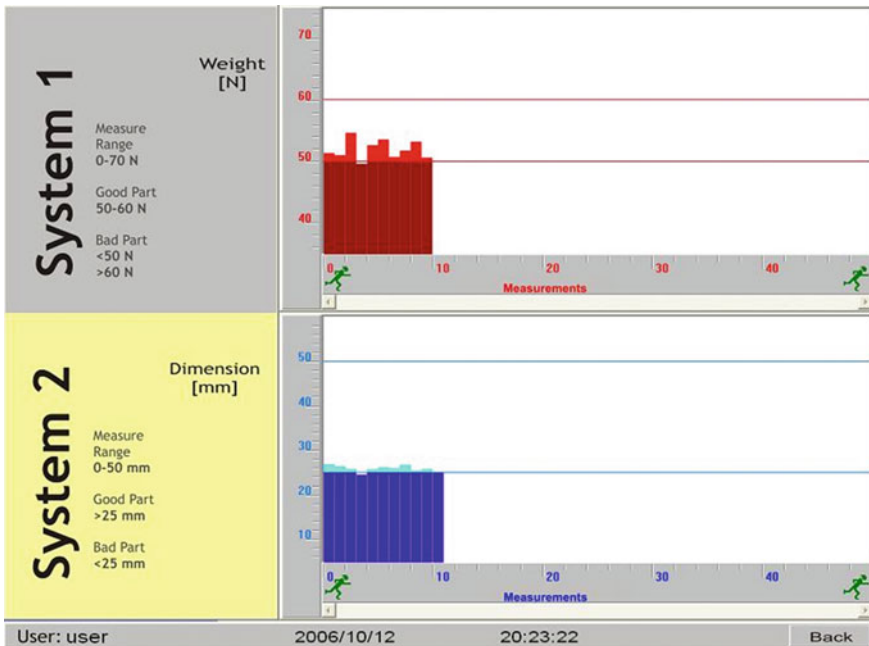


Fig. 7.17 Graphical information of the cumulated produced parts

In Fig. 7.17 it is shown the cumulative produced data from both measurement systems (weight and dimension) in a graphical format. This Menu identifies also the quality criteria for accepted or rejected parts.

## 7.6 Conclusions

Modern production plants imply great number of production lines that are distributed along big surfaces. This decentralized environment increases the need for complex management tools that enable a complete on-line overview about the system.

This study presents a concept for an industrial operational management tool that incorporates low-level communications between processes—PLCs (Programmable Logic Controllers) and remote communications between system administrators and production processes (internet, GSM mobile communications, etc.). This concept is a two layer management architecture where inner loops are performed by low level PLC Master–Slave networks and the outer loop is performed by a SCADA supervisory system.

At the SCADA supervisory level an optimal allocation of production resources is performed taking into account the minimization of the assembly costs that are being on-line monitored.

This model simulates actually what happens in real production plants, where the unexpected rising of the production costs is directly dependent on the quantity of good parts produced. The greater the scrap level, the line downtime, the non-availability of assembly parts, etc., the more the production costs rise.

**Acknowledgments** This work was partially supported by the Fundacao Ciencia e Tecnologia, through IDMEC under LAETA.

## References

1. Silva P, Botto M, Figueiredo J, Rijo M (2007) Model predictive control of an experimental water canal. In: Proceedings of the IFAC ECC'07 European control conference, Kos–Greece
2. Figueiredo J, Sá da Costa J (2007) A concept for an operational management system for industrial purposes. In: Conference of the proceedings of the IEEE international symposium of intelligent signal processing (WISP'2007), Madrid, Paper 1-4244-0830-X/07
3. Dey A, Abowd G, Salber D (1999) A context-based infrastructure for smart environments, managing interactions in smart environments. Springer, Dublin, pp 114–130
4. Silva C, Sousa JM, Runkler T, Palm R (2005) Soft computing optimization methods applied to logistic processes. *Int J Approx Reason* 40:280–301
5. Carpanzano E, Jovane F (2007) Advanced automation solutions for future adaptive factories. *Ann CIRP* 56/1:435–438

6. Urdaneta G, Colmenares J, Queipo N, Arapé N, Arévalo C, Ruz M, Corzo H, Romero A (2007) A reference software for the development of industrial automation high-level applications in the petroleum industry. *Comput Ind* 58:35–45
7. Figueiredo J, Sá Costa J (2012) A SCADA system for energy management in intelligent buildings. *Energy Buildings* (Elsevier) 49:85–98. <http://dx.doi.org/10.1016/j.enbuild.2012.01.041>
8. Figueiredo J, Sá Costa J (2008) Operative platform applied to building automation. *Comput-Aided Civil Infrastruct Eng* (IOS Press) 23:639–653. <http://dx.doi.org/10.1111/j.1467-8667.2008.00566.x>
9. Figueiredo J, Martins J (2010) Energy production system management—renewable energy power supply integration with building automation system. *Energy Conversat Manag* (Elsevier) 51:1120–1126. <http://dx.doi.org/10.1016/j.enconman.2009.12.020>
10. Neumann P (2007) Communication in industrial automation—What is going on? *Control Eng Pract* 15:1332–1347
11. Ozdemir E, Karacor M (2006) Mobile phone based SCADA for industrial automation. *ISA Trans* 45(1):67–75
12. SIEMENS, Simatic WinCC V7.0, 2008
13. The MathWorks, Matlab 7.6.0.324 (R2008a), The MathWorks Inc. 2008
14. <http://www.opcfoundation.org>
15. Santos R, Normey-Rico J, Gomez A, Arconada L, Moraga C (2005) OPC based distributed real time simulation of complex continuous processes. *Simul Model Pract Theory* 13:525–549
16. Grafset Le (1991) De Nouveaux concepts. GREPA, Cépaduès Éditions, France
17. Simatic S7-300—Ladder Logic (LAD) for S7-300, SIEMENS, 2001
18. The MathWorks, Matlab—Optimization Toolbox, 2008
19. Simatic S7 Prodave S7—Toolbox for PGs and PCs, SIEMENS, 2001
20. Simatic Net—NCM S7 for Profibus/FMS SIEMENS, SIEMENS, 12/2001

# Chapter 8

## Modeling and Optimization of Mechanical Systems and Processes

Ramón Quiza, Gerardo Beruvides and J. Paulo Davim

**Abstract** This chapter reviews the most commonly used techniques used for modeling and optimizing mechanical systems and processes. Statistical and artificial intelligence based tools for modeling are summarized, pointing their advantages and shortcomings. Also, analytic, numeric and stochastic optimization techniques are briefly explained. Finally, two cases of study are developed in order to illustrate the use of these tools, the first one dealing with the modeling of the surface roughness in a drilling process and the other one, on the multi-objective optimization of a hot forging process.

### 8.1 Empirical Modeling of Mechanical Systems and Processes

#### 8.1.1 Introduction

A model is an idealized representation of a real system. Models can be either physical, such as architectural scale models (Fig. 8.1a), maps (Fig. 8.1b) and nomograms (Fig. 8.1c), or logical-mathematical such as computer-based geometrical representations (Fig. 8.2a), mathematical equations or systems of equations (Fig. 8.2b) and algorithms (Fig. 8.2c).

---

R. Quiza (✉) · G. Beruvides  
Department of Mechanical Engineering, University of Matanzas,  
Autopista a Varadero km 3½, 44740 Matanzas, Cuba  
e-mail: quiza@umcc.cu

J. P. Davim  
Department of Mechanical Engineering, University of Aveiro,  
Campus Santiago, 3810-193 Aveiro, Portugal

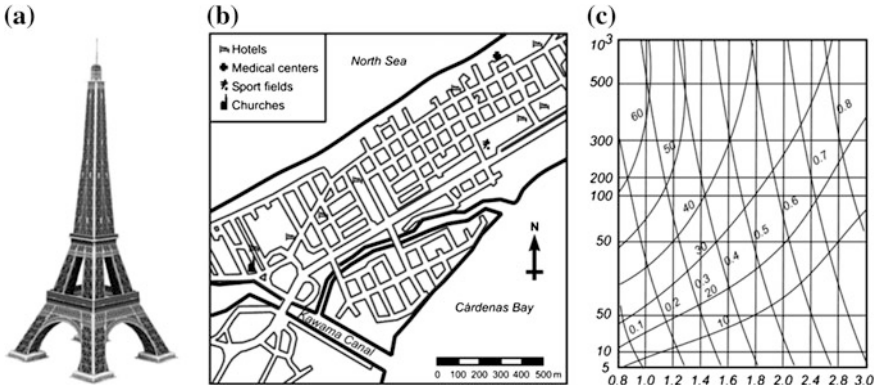


Fig. 8.1 Physical models. **a** Scale model. **b** Map. **c** Nomogram

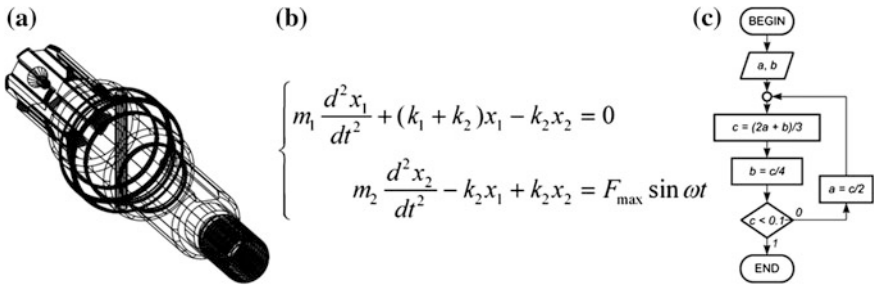


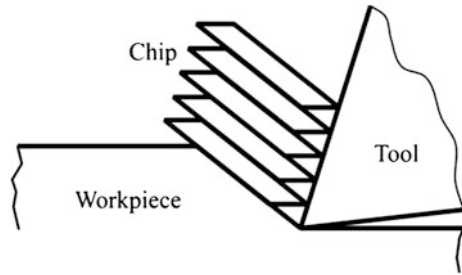
Fig. 8.2 Logical-mathematical models. **a** Computer-based geometric model. **b** System of equations. **c** Algorithm

Of course, this representation can be achieved only by disregarding the non important features of the studied system. For example, in modeling the behavior of beam under flexion, information about the color of the surface is completely irrelevant and can be excluded from the model.

Every model is composed by a set of inputs, a set of outputs and a set of relationships between them. By considering the nature of these inputs and outputs, models can be classified as qualitative and quantitative. For example, the Piispänen model of the chip formation in the orthogonal cutting process (Fig. 8.3) is a well-known qualitative model. On the contrary, the Oxley predictive machining theory is a quantitative model for the same purpose.

Another way for classifying a model is by considering the source of the relationships between inputs and outputs. If these relationships are based on natural laws the model is called phenomenological. An example of this kind of model is the equation describing the position,  $h$ , of a body freely falling from an initial height,  $h_0$ :

**Fig. 8.3** Piispanen chip formation model



$$h(t) = h_0 - \frac{gt^2}{2}; \tag{8.1}$$

where  $t$  is the time and  $g$  is the gravitational acceleration. The before-mentioned expression can be obtained directly from the universal gravitation law and second Newton law.

On the other hand, a model is considered to be empirical if the relationships between the inputs and outputs in based on experimental data. An example of empirical model is the Johnson–Cook equation:

$$\sigma = (A + B\varepsilon^n) \left( 1 + C \ln \frac{\dot{\varepsilon}}{\dot{\varepsilon}_0} \right) \left[ 1 - \left( \frac{T - T_0}{T_M - T_0} \right)^m \right]; \tag{8.2}$$

with describe the thermoviscoplastic behavior of metals. The coefficients of the model,  $A$ ,  $B$ ,  $C$ ,  $n$  and  $m$ , must be determined from experimental data.

Although the classification into phenomenological and empirical is useful from the theoretical point of view, practical applications, especially in engineering fields, usually combine both classes models.

### 8.1.2 Foundations of the Empirical Modeling

A mathematical empirical model relates a nonrandom  $n$ -input variable,  $\mathbf{x} \in \mathbb{R}^n$ , with a random variable scalar output variable,  $Y(\mathbf{x})$ . As in a multi-output model the outputs can be separately handled, and, therefore, it can be considered as multiple single-output models, the previous definition is general enough [5].

The model is said to be true if it verifies the expression:

$$Y(\mathbf{x}) = \Phi(\mathbf{x}) + \varepsilon(\mathbf{x}); \tag{8.3}$$

where  $\Phi(\mathbf{x})$  is a parameterized function:

$$\Phi(\mathbf{x}) \triangleq \{f(\mathbf{x}, \boldsymbol{\beta}), \mathbf{x} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^q\}; \tag{8.4}$$

and  $\varepsilon(\mathbf{x})$  is a random variable with zero expectation.

The values of the parameters of the model,  $\boldsymbol{\beta}$ , are estimated from a dataset of  $N$  input–output pairs  $\{\mathbf{x}^{(k)}, y^{(k)}\}_{k=1\dots N}$ , where  $y^{(k)}$  are measured values (realizations) of the random variable  $Y(\mathbf{x})$ .

The goal of the modeling procedure is not only to estimate the regression  $\Phi(\mathbf{x})$ , but also to compute the value of a confidence interval for the regression, that is the value of a random interval with a chosen probability to contain the regression [12].

### 8.1.3 Linear Regression Models

The simplest regression models are represented by the linear equation:

$$\Phi(\mathbf{x}) \triangleq \{\mathbf{x}^T \boldsymbol{\beta}, \mathbf{x} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^n\}, \quad (8.5)$$

In this case, for the whole data set can be written as [3]:

$$\mathbf{y} = [\mathbf{x}] \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (8.6)$$

where  $[\mathbf{x}] = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}]^T$  is the nonrandom input matrix;  $\mathbf{y} = [y^{(1)}, y^{(2)}, \dots, y^{(N)}]^T$  is the random output vector and  $\boldsymbol{\varepsilon} = [\varepsilon^{(1)}, \varepsilon^{(2)}, \dots, \varepsilon^{(N)}]^T$  is a random vector with zero expectation, also known as residuals. The values:

$$\hat{\mathbf{y}} = [\mathbf{x}] \boldsymbol{\beta}; \quad (8.7)$$

are known as predicted values of the model.

The values of the model parameters,  $\boldsymbol{\beta}$ , are estimated by using the least squares method, which minimizes the quadratic cost function:

$$J(\boldsymbol{\theta}) = \frac{1}{2}(\mathbf{y} - [\mathbf{x}] \boldsymbol{\beta})^T (\mathbf{y} - [\mathbf{x}] \boldsymbol{\beta}); \quad (8.8)$$

obtaining the estimations,  $\hat{\boldsymbol{\beta}}$ , through the expression:

$$\hat{\boldsymbol{\beta}} = ([\mathbf{x}]^T [\mathbf{x}])^{-1} [\mathbf{x}]^T \mathbf{y}. \quad (8.9)$$

The  $(1-\alpha)100$  % confidence interval for the predictions of the linear regression model can be expressed as:

$$\mathbf{x}^T \hat{\boldsymbol{\beta}} \pm t_{N-n}(1 - \alpha/2) \sqrt{s^2 \mathbf{x}^T ([\mathbf{x}]^T [\mathbf{x}])^{-1} \mathbf{x}}; \quad (8.10)$$

where  $t_{N-n}$  is the inverse of the Student cumulative distribution with  $N-n$  degrees of freedom, and  $s^2$  is an unbiased estimator of the variance of the residuals,  $\boldsymbol{\varepsilon}$ :

$$s^2 = \frac{\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}}{N - n}. \quad (8.11)$$



In order to know the goodness of fit of a model, some analyses must be carry out. The first one, and probably the best known, is the coefficient of determination,  $R^2$ , which can be determined by:

$$R^2 = 1 - \frac{SS_{\text{err}}}{SS_{\text{tot}}}; \quad (8.12)$$

where:

$$SS_{\text{err}} = (\mathbf{y} - [\mathbf{x}]\boldsymbol{\beta})^T (\mathbf{y} - [\mathbf{x}]\boldsymbol{\beta}) \quad (8.13)$$

is the sum of squares of residuals, and:

$$SS_{\text{tot}} = (\mathbf{y} - \bar{\mathbf{y}})^T (\mathbf{y} - \bar{\mathbf{y}}) \quad (8.14)$$

is the total sum of squares. Here:

$$\bar{\mathbf{y}} = \frac{1}{N} \sum_{k=1}^N y^{(k)} \quad (8.15)$$

is the average of the observed output values.

The coefficient of determination can be interpreted as the amount of variation in the response (output variable) that can be explained by the fitted model. Nevertheless, it cannot be assumed that a regression with a high coefficient of determination is automatically a good model. For example, models based in high noisy data (Fig. 8.4a) often have a low  $R$ -squared even if they properly describe the relationship between the variables. On the contrary, a model can have a high coefficient of determination but its trend does not match with the observed data (Fig. 8.4b).

Another important test in the statistical significance of the model. In this test the null hypothesis is that all the model coefficients are equal to zero:

$$\begin{cases} H_0 : \beta_i = 0, \forall i \in \{1, 2, \dots, n\} \\ H_1 : \exists i \in \{1, 2, \dots, n\} | \beta_i \neq 0 \end{cases} \quad (8.16)$$

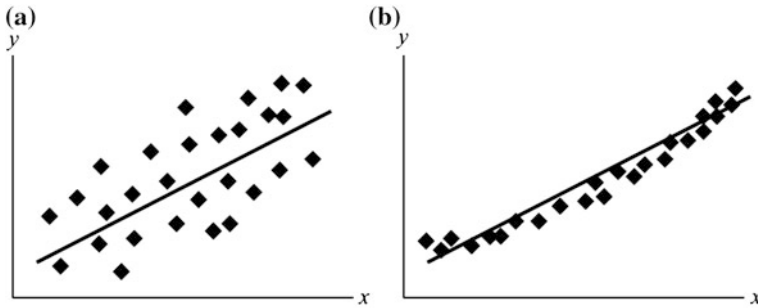
This hypothesis can be rejected, with a  $(1-\alpha)100\%$  of statistical confidence, if the value of the cumulative probability of the statistic:

$$F_0 = \frac{(SS_{\text{tot}} - SS_{\text{err}})/(n-1)}{SS_{\text{err}}/(N-n)}; \quad (8.17)$$

in a Fisher distribution with  $(n-1)$ , and  $(N-n)$  degree of freedom, is greater than  $1-\alpha$ :

$$F_{n-1, N-n}^{\text{cdf}}(F_0) < 1 - \alpha; \quad (8.18)$$

This test is often presented in a tabular form in the so-called analysis of variance (ANOVA) (Table 8.1).



**Fig. 8.4** Goodness of fit with high and low noisy data. **a** Good fitted model for high noisy data. **b** Poorly fitted model for low noisy data

**Table 8.1** ANOVA of a regression model

Source	Sum of squares	Degrees of freedom	$F_0$	$p$
Model	$SS_M = SS_T - SS_E$	$DF_M = n - 1$	$F_0 = (SS_M/DF_M)/(SS_E/DF_E)$	$1 - F_{n-1, N-n}^{cdf}(F_0)$
Error	$SS_E$ (Eq. 8.13)	$DF_E = N - n$		
Total	$SS_T$ (Eq. 8.14)			

A third analysis is the significance of each coefficient of the model. This test allows to know which input variables have not a significant influence on the variability of the output variable and, consequently, can be removed from the model. The hypothesis for this test considers that the given coefficient is equal to zero:

$$\begin{cases} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{cases} \tag{8.19}$$

The test of this hypothesis is carry out by computing the statistics:

$$t_{0,i} = \frac{\hat{\beta}_i}{\sqrt{s^2 C_{ii}}}; \tag{8.20}$$

where  $C_{ii}$  are  $i$ -th diagonal elements of the matrix:

$$[C] = ([\mathbf{x}]^T [\mathbf{x}])^{-1}. \tag{8.21}$$

If the cumulative probability in a t-Student distribution, with  $N - n$  degrees of freedom, is greater than  $1 - \alpha/2$ :

$$t_{N-n}^{cdf}(t_{0,i}) > 1 - \alpha/2; \tag{8.22}$$

then the null hypothesis can be rejected with a  $(1 - \alpha)100$  % of statistical confidence and this coefficient can be removed from the model.

Finally, the last important aspect that can be analyzed in a regression model is the distribution of the residuals. If the model is true, it can be expected that the residuals will be normally distributed. It is also important to analyze the unusually large residuals. In order to do that, the residuals are transformed by :

$$r_i = \frac{\varepsilon_i}{\sqrt{s^2(1 - H_{ii})}}; \quad (8.23)$$

where  $H_{ii}$  are the  $i$ -th diagonal elements of the matrix:

$$[\mathbf{H}] = [\mathbf{x}][\mathbf{x}^T[\mathbf{x}]^{-1}[\mathbf{x}]. \quad (8.24)$$

The values  $r_i$  are known as internally studentized residuals. On the contrary, if the standard error is determined by excluding the  $i$ -th point:

$$t_i = \frac{\varepsilon_i}{\sqrt{\tilde{s}^2(1 - H_{ii})}} \quad (8.25)$$

where:

$$\tilde{s}^2 = \frac{SS_E - \varepsilon_i^2 / (1 - H_{ii})}{N - n - 1}; \quad (8.26)$$

then, the values residuals,  $t_i$ , are called externally studentized.

As the studentized residuals must have a normal distribution with zero mean and standard deviation equal to one, the 99 % of the points must be contained in the interval  $(-3, 3)$ . Therefore, any point outside this interval (Fig. 8.5) can be considered as a possible unusual residual and must be carefully analyzed in order to determine if this behavior is purely random or if it has another cause (for example a wrong experimental execution).

Finally, it must be pointed out that some data, although not being intrinsically linear, can be transformed in such a way that can be described by a linear regression. This process is known as linearization.

For example, if the relationship between an input variable,  $x$ , and an output variable,  $y$ , is expected to be modeled in the form:

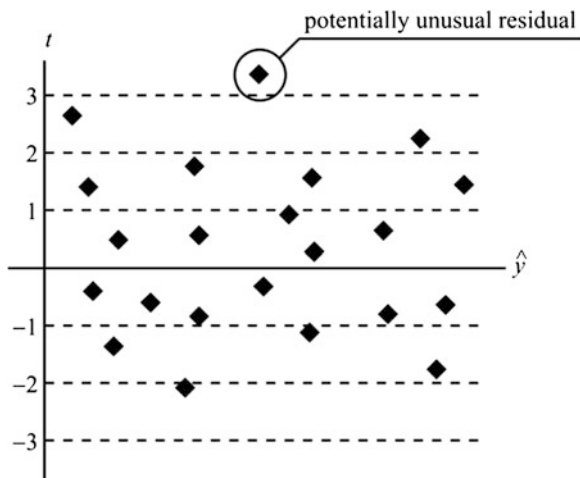
$$y = Cx^n; \quad (8.27a)$$

where  $C$  and  $n$  are specific constants (which is a very common behavior for a lot of natural phenomena), by applying a logarithm to both sides of the equation, the following linear model is obtained:

$$\ln y = \ln C + n \ln x; \quad (8.27b)$$

which can fitted as explained.

**Fig. 8.5** Residuals distribution



### 8.1.4 Nonlinear Regression Models

Although most of the models used in science and engineering are linear or can be linearized, some of them are strictly nonlinear. In these kind of models, the amount of parameters,  $q$ , can be different than the number of input variables,  $n$ , but always less than the number of input–output pairs in the dataset,  $N$ , in order to prevent mathematical indetermination [15].

For this models, some numerical methods are used, instead Eq. 8.9, for determining the values of the model parameters. Some of the most common are the Gauss–Newton algorithm, the gradient descent and the Levenberg–Marquardt algorithm [5].

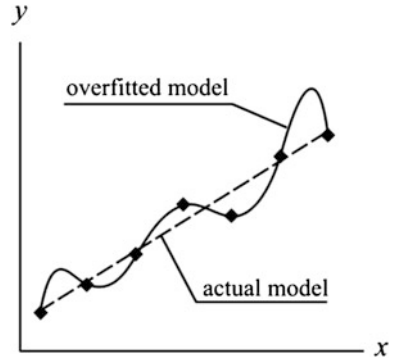
In order to determine the confidence intervals for the predictions, in the nonlinear models, the following equation, resulting of a first order linearization by using Taylor expansion, is used [12]:

$$f(\mathbf{x}, \boldsymbol{\beta}) \pm t_{N-q}(1 - \alpha/2) \sqrt{s^2 \mathbf{z}^T ([\mathbf{z}]^T [\mathbf{z}])^{-1} \mathbf{z}}; \quad (8.28)$$

where  $[\mathbf{z}]$  is the Jacobian matrix of the nonlinear function  $f(\mathbf{x}, \boldsymbol{\beta})$ , whose elements can be computed by:

$$z_{ij} = \left. \frac{\partial f_i(\mathbf{x}, \vec{\beta})}{\partial \beta_j} \right|_{\vec{\beta}=\hat{\vec{\beta}}}. \quad (8.29)$$

An important issue in the use of nonlinear models is the overfitting, which take place when the model fits not only the relationship between the input and output variables but also the noise present in the data. This unwanted phenomenon causes a lack of generalization capability of the model and takes place, mainly when the number of parameters becomes too large (Van der [1]).

**Fig. 8.6** Overfitting

For example, Fig. 8.6 shows a dataset (points) which is a realization of a linear relationship (dashed line) between the variables  $x$  and  $y$ , plus some noise. Nevertheless if a nonlinear regression with an excessive number of parameters (solid line) is fitted, the model, although passing through all the points does not reflect properly the nature of the relationship. Avoiding overfitting is an important aspect in highly complex intelligent models, such as neural networks.

### 8.1.5 Artificial Intelligence-based Models

Artificial intelligence (AI) refers to a wide group of tools and paradigms aiming to understand and simulate the human intelligence and other complex natural systems. Although far from these final objectives, IA-based techniques have found a wide spectrum of applications in several knowledge branches.

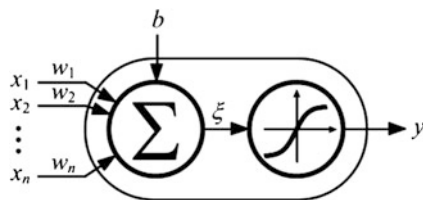
For modeling purposes, the most used AI techniques are the neural networks, the fuzzy logic and neuro-fuzzy systems, and the probabilistic methods for uncertain reasoning. These topics will be briefly reviewed in the following paragraphs.

#### 8.1.5.1 Neural Networks

Artificial neural networks are the most popular and well-established AI paradigm. This technique, inspired in the network structure of the biological brains, has been extensively used for classification tasks and function approximation.

The whole idea of the artificial neural network is based on the concept of artificial neuron, which is a rough mathematical simulation of the biological neuron. The McCulloch-Pitts model (Fig. 8.7) is a typical neuron model. It can be considered as a computing unit with a input  $n$ -vector,  $\mathbf{x} \in \mathbb{R}^n$ , which are weighted and linearly combined (Eq. 8.30) and then transformed by some transfer function, usually a sigmoid (Eq. 8.31a), an hyperbolic tangent (Eq. 8.31b), or simply a linear function (Eq. 8.31c) to give a scalar output,  $y$

**Fig. 8.7** McCulloch-Pitts neuron model



$$\xi = b + \sum_{k=1}^n w_k x_k; \quad (8.30)$$

$$y(\xi) = \frac{1}{1 + e^{-\xi}} \quad (8.31a)$$

$$y(\xi) = \tanh(\xi) = \frac{e^{\xi} - e^{-\xi}}{e^{\xi} + e^{-\xi}}; \quad (8.31b)$$

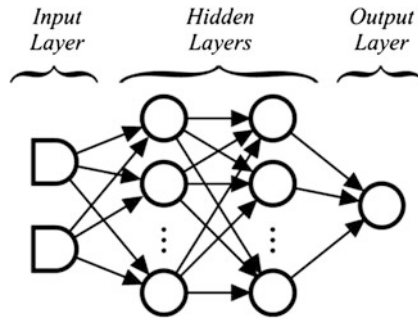
$$y(\xi) = \xi. \quad (8.31c)$$

To the linear combination is often added a scalar term,  $b$ , which is called bias or threshold, representing the predisposition of the neuron to be activated.

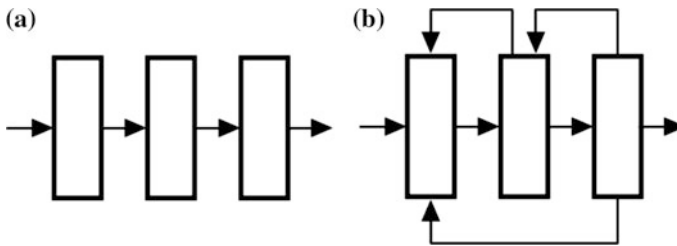
Neurons are combined to form a network, in such a way that the outputs of some networks are the inputs of other ones. Neurons are topologically arranged in layers (Fig. 8.8). The input layer just takes the input values without any mathematical processing; actually, this layers is conventionally included only for hardware implementation purposes. The hidden layers receive this name because neither their inputs nor their outputs are gives in the training dataset. Finally the output layer give the response of the network.

If neurons in some layers only receive inputs from neurons located in previous layers the networks is called feed-forward (Fig. 8.9a). Examples of this structure are the multilayer perceptron (MLP) and the radial basis function networks (RBFN). On the other hand, if connections exist from any layer to previous ones, the network is called recurrent (Fig. 8.9b). Examples of this architecture are de Hopfield network and de Elman networks.

The process of select the values of the free parameters (weights and biases) of the network is known as training or learning process. There are two main approaches for training a neural network: if the whole set of input–output pairs are known, a supervised learning is carried out. The most popular supervised learning algorithm is the error backpropagation, which is used for training MLP's and is, actually, a generalization of the delta rule, a gradient descent algorithm. On the contrary, when the output information of the training dataset is not completely known, unsupervised learning algorithms must be used. This kind of learning is used principally for classification task and is widely used in self-organized maps (SOM) and adaptive resonance theory (ART) networks.



**Fig. 8.8** Neural network topology



**Fig. 8.9** Feed-forward versus recurrent architecture. **a** Feed forward architecture. **b** Recurrent architecture

### 8.1.5.2 Fuzzy and Neuro-Fuzzy Systems

Fuzzy logics, unlike traditional or Boolean logics, deals with uncertain relationships. In this approach, an statement can have a degree of truth ranging from zero to one, while in Boolean logic this degree of truth can be just zero (false) or one (true). The so-called membership functions determines the degree of membership of some element to some fuzzy subset. For example, for three membership functions called “LOW”, “MODERATED” and “HIGH” (Fig. 8.10) a cutting speed of 150 m/s will have degrees of truth of 0.75, 0.25 and 0, respectively.

A fuzzy inference system is a set of IF–THEN fuzzy rules, assembled together to offer some response to a given input. A fuzzy inference system (Fig. 8.11) is composed by a fuzzifier that converts the crisp inputs into fuzzy values. Then the inference engine applies the IF–THEN rules to produce a fuzzy output representing the response of the system and, finally, this fuzzy set is transformed back into a crisp value by the defuzzifier. The membership functions used in the three steps form the fuzzy knowledge base.

Although fuzzy inference systems have been successfully applied for solving many control and modeling problems, their main drawback is the selection and tuning of the membership functions.

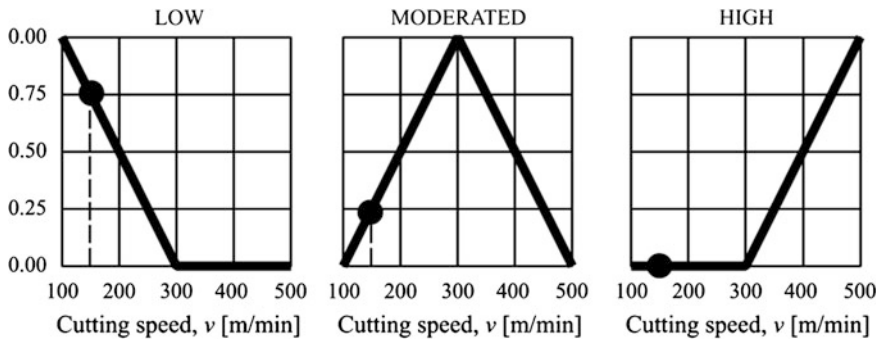
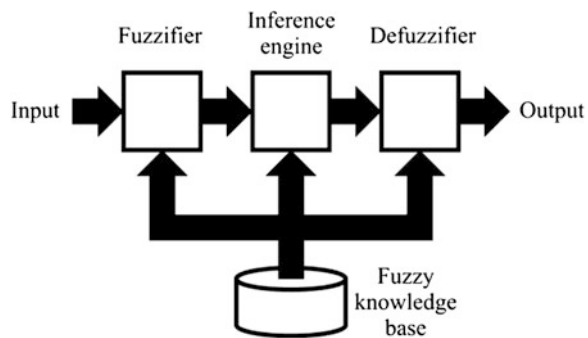


Fig. 8.10 Example of membership functions

Fig. 8.11 Fuzzy inference system



An approach for solving this problem is the adaptive neuro-fuzzy inference system (ANFIS) which is an inference system organized by layers, some of them using fuzzy rules and other being tuned by an adaptive learning process, like a neural network.

### 8.1.5.3 Probabilistic Methods for Uncertain Reasoning

This group of techniques deal with uncertainty. Although they seem to be similar to fuzzy logic, both approaches are fundamentally different. On the first hand, probability theory makes predictions about events from a state of partial knowledge. On the other hand, fuzzy logic deals with situations where vagueness is intrinsic. In probability theory the truth is unknown but absolute while in fuzzy logic it is relative by nature.

The most widely applied probabilistic techniques, in the fields of modeling mechanical systems are the Bayesian networks and the Hidden Markov models. A Bayesian network is a directed acyclic graph consisting in a set of nodes, representing random variables, and a set of directed edges, representing their conditional dependencies [10].



The dependencies in a Bayesian network can be adaptively determined from a dataset through a learning process. The objective of this training is to induce the network with the best description of the probability distribution over the dataset, and can be considered as an unsupervised learning because the values of the attributes are not supplied in the dataset [4].

Hidden Markov models consist of a set of hidden states that form a chain described by a transition probability distribution over these states and an associated set of emission probability distribution for the observed symbols [6]. Unlike Bayesian networks, hidden Markov models can be represented as cyclic graphs and they have the ability to model the temporal evolution of signals [9].

The learning process in a hidden Markov model aims to find, for a given a set of output sequences, the corresponding best chain. Although there is not a known suitable algorithm for solving this problem exactly, several approaches have been proposed for obtaining local solutions in an computationally efficient way [7].

## 8.2 Multi and Single-Objective Optimization of Mechanical Systems and Processes

### 8.2.1 Basic Concepts on Optimization

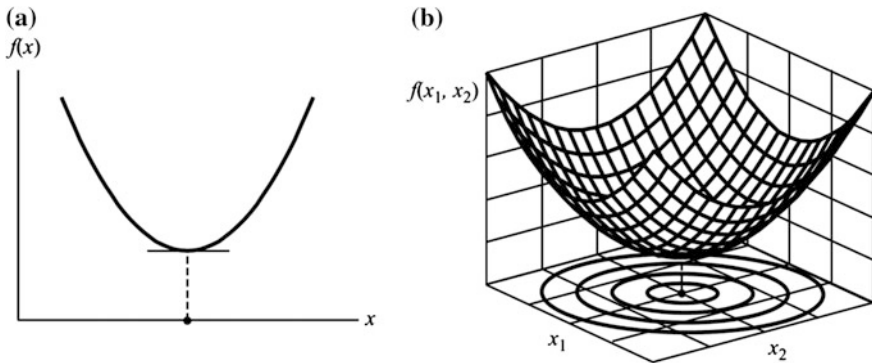
Optimization is the problem of, given some scalar function,  $f(\bullet)$  (called objective function), of a vector variable,  $\mathbf{x} \in \Omega \subset \mathbb{R}^n$  (called decision variable), finding the value  $\mathbf{x}^*$  such that  $f(\mathbf{x}^*) \leq f(\mathbf{x})$  for all  $\mathbf{x} \in \Omega$ , and also fulfill the conditions  $g_i(\mathbf{x}) \leq 0$ ,  $i = 1, \dots, m$  (called inequality constraints) and  $h_i(\mathbf{x}) \leq 0$ ,  $i = 1, \dots, p$  (called equality constraints).

Actually this definition refers only minimization but it is not a limitation as any maximization problem can be transformed into a minimization one just by multiplying the objective function by minus one.

Some practical problems, however, require the simultaneous minimization (or maximization) of several objective functions. This is called multi-objective or multi-criteria optimization.

There are two main approaches for solving a multi-objective optimization. The first one, called a priori approach, is carried out by supplying information about the preferences between the objectives before executing the optimization. In this group are included the linear and nonlinear aggregation of objectives, the lexicographic method and the goal programming.

In the second approach, called a *posteriori*, no information is supplied about the preferences between the objectives. The optimization process is carried out and then the decision maker choose the most convenient solution from a set of non-dominated solutions that are optimal in the wide sense that there is no other solution in the considered search space that improve at least one objective function without detriment of another function [8]. The methods of this approach are based on obtaining of the so-called Pareto front, which contains these non-dominated solutions [16].



**Fig. 8.12** Stationary points. **a** Scalar decision variable. **b** Vector decision variable

### 8.2.2 Analytic Optimization Techniques

Analytic are the oldest and the most exact optimization methods. They are based on determine the stationary points, i.e., the points where the first derivative of the objective function is zero (Fig. 8.12a):

$$\frac{df}{dx} = 0. \quad (8.32)$$

When the decision variable is not scalar but vector, stationary points exists where all the components of the gradient of the function are zero (Fig. 8.12b):

$$\nabla f = \frac{\partial f}{\partial x_1} \mathbf{e}_1 + \dots + \frac{\partial f}{\partial x_n} \mathbf{e}_n = \mathbf{0}. \quad (8.33)$$

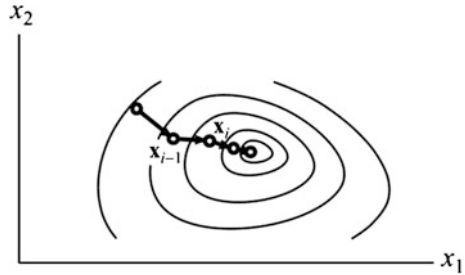
As this condition is necessary but not sufficient, the second derivative (or the Hessian matrix, for vector decision variable) must be checked in order to know if the given stationary point corresponds to a maximum, a minimum or a saddle point.

Analytic techniques has a strong mathematic foundation but only work fine in relatively simple problems. Another limitation is in constrained optimization, although some methods such as the Lagrange multipliers have been developed with this purpose.

### 8.2.3 Numeric Optimization Techniques

Numeric optimization techniques are also based on the gradient of the objective function, but unlike analytic methods do not require the computation of the roots of derivatives, which are often transcendental equations.

**Fig. 8.13** Iterative optimization method



Numeric techniques start from some point,  $\mathbf{x}_0$ , and computes iteratively new points,  $\mathbf{x}_i$ , from previous point,  $\mathbf{x}_{i-1}$ , by following the gradient until some stop condition is reached (Fig. 8.13).

Into the numeric optimization techniques two major groups exists: those that uses the Hessian matrix (such as the Newton's method) or those that uses the gradient (for example, the conjugate gradient and the gradient descent). Often the derivatives are not directly evaluated but approximated using finite differences.

Two main drawbacks have the iterative methods. In the first place, the selection of the start point heavily influences the convergence of the method. The other shortcoming is that the found solution may be a local optimum instead a global one. Furthermore, for a successful application of most of the iterative methods, the objective function must be continuous and differentiable.

### 8.2.4 Heuristic Optimization Techniques

Many optimization problems in engineering, especially on mechanical and manufacturing systems, do not fulfill the conditions of continuity, differentiability and unimodality, required for applying the analytical or numeric methods. For solving this kind of problems, a group of techniques, known as heuristics, have been developed.

Heuristics, unlike analytical or numeric methods, do not rely on a solid mathematic foundation; they are inspired on natural systems, either physical or biological, and try to find near-optimal solutions, which although being different from the actual optimum, are good enough to be applied in practical situations.

Currently there are a lot of heuristics for optimization; however, the most popular are the following:

- **Simulated annealing:** Is based on the annealing processes in metals and other lattice structures, where the systems is conduced to a minimum energy state. The method starts from some point and, in each iteration creates and evaluates some neighbor points. If the objective function is lower in the new point than in the actual one, this is replaced. However, this replacement is not deterministic

buy random, depending on some prescribed parameter called temperature by analogy with the physical annealing process [14].

- Evolutionary optimization: Is inspired in the natural evolution of biological organisms. This methods carry out parallel searches which start from an initial solution set (called population) and create, in each iteration, a new “child” population that “inherits” the characteristics of the best “parents”. Both the selection of the “parents” and the creation of “children” include random processes. Evolutionary algorithms include two main branches, evolutionary strategies (European school) and genetic algorithms (American school) [13].
- Swarm intelligence: Is based on the behavior of natural decentralized systems, composed by a group of individuals which work together to achieve some common goal. This paradigm comprises a lot of algorithms, including as the most popular, but not limited to, ant colony optimization, cuckoo search and particle swarm optimization [2].

It must be remarked that all these methods have some random behavior, which tries to avoid the finding of a local minimum instead the global one. For this reason, these approaches are often referred to as stochastic optimization.

## 8.3 Cases of Study

### 8.3.1 Modeling of the Surface Roughness in a Drilling Process

#### 8.3.1.1 Problem Statement and Experimental Setup

This case of study deals with a drilling process of a Ti-6Al-4 V alloy with a hardness of 265 HV and a tensile strength of 993 MPa. 6 mm-diameter DIN 6537 carbide (grade GC 1010) drills were used.

The objective of the study was to determine the model relating the surface roughness of the elaborated holes with the cutting speed,  $v$ , the feed rate,  $f$ , and the used amount of lubricant,  $L$ . Experiments were carried out considering three levels of cutting speed (20, 40 and 80 m/min), four levels of feed rate (0.015, 0.030, 0.060 and 0.120 mm/rev) and three levels of lubrication (0: without lubrication, 0.5: minimum quantity lubrication [MQL], and 1: abundant lubrication). For abundant lubrication an emulsion of 10 % of mineral oil was used, with the flow rate defined by the machining center. For MQL biodegradable vegetable oil with a flow rate of 50 mL/h.

The roughness,  $R_z$ , was measured as established by ISO 4287, by using a wave system roughness tester. Experimental results are shown in Table 8.2.

**Table 8.2** Experimental data of roughness

Lubrication	$v$ [m/min]	$f$ [mm/rev]	$R_z$ [ $\mu\text{m}$ ]		
			1	2	3
0.0	20	0.015	10.1	11.6	12.5
0.0	40	0.015	8.1	7.1	8.3
0.0	20	0.03	10.2	10.6	9.6
0.0	40	0.03	9.3	5.7	7.7
0.0	80	0.03	6.8	5.2	8.3
0.0	20	0.06	6.1	6.4	6.6
0.0	40	0.06	4.7	4.6	5.5
0.0	80	0.06	9.7	7.1	6.9
0.0	20	0.12	4.4	4.7	4.4
0.0	40	0.12	9.1	9.3	8.4
0.0	80	0.12	8.8	7.3	8.3
0.5	20	0.015	3.7	3.6	4.6
0.5	40	0.015	5.5	5.2	6.1
0.5	80	0.015	4.9	4.9	5.6
0.5	20	0.03	11.3	12.0	13.6
0.5	40	0.03	5.3	5.9	5.9
0.5	80	0.03	9.3	7.7	10.4
0.5	20	0.06	4.1	4.2	4.6
0.5	40	0.06	4.8	4.2	3.5
0.5	80	0.06	8.0	5.7	5.0
0.5	20	0.12	5.6	5.1	5.1
0.5	40	0.12	5.9	7.0	7.2
0.5	80	0.12	10.3	10.6	11.0
1.0	20	0.015	4.9	3.8	5.0
1.0	40	0.015	6.0	6.7	9.0
1.0	80	0.015	7.2	4.5	5.4
1.0	20	0.03	2.8	3.9	4.6
1.0	40	0.03	10.0	8.4	11.9
1.0	80	0.03	5.1	11.9	12.2
1.0	20	0.06	4.4	5.5	8.5
1.0	40	0.06	6.9	6.5	9.6
1.0	80	0.06	8.9	6.7	7.3
1.0	20	0.12	6.1	5.2	4.3
1.0	40	0.12	5.9	5.9	8.2
1.0	80	0.12	9.8	11.2	8.8

### 8.3.1.2 Regression Models

Three statistical models were tried in order to describe the relationship between the experimental factors (lubrication,  $L$ ; cutting speed,  $v$ ; and feed rate,  $f$ ) and the maximum roughness,  $R_z$ . The first fitted model was a linear regression, obtaining the equation:

**Table 8.3** Analysis of variance of the lineal regression of  $Rz$ 

Source	Sum of squares	DoF	Mean square	$F$ -ratio	$P$ -value
Model	14.65	3	4.88	0.940	0.4354
Residual	161.9	31	5.22		
Total	176.5	34			

$$Rz = 6.363 - 0.7018L + 24.19 \times 10^{-3}v - 0.4310f; \quad (8.34)$$

The  $R^2$  statistic has a value of 0.083, indicating that the model as fitted explains only the 8.3 % of the variability in  $Rz$ . The standard error of the estimate is 2.285 and the mean absolute error is 1.706.

Table 8.3 shows the analysis of variance of the linear regression of  $Rz$ . Since the  $P$ -value is greater than 0.05, there is not a statistically significant relationship between the variables at the 90 % or higher confidence level. Finally, analyzing the significance of the model coefficients (Table 8.4) it can be noted that all the terms are not statistically significant at the 90 % or higher confidence level.

Considering all the above mentioned aspects, it can be concluded that the linear regression models does not properly describe the relationship between the studied variables.

Other tried model was a quadratic regression which includes not only the linear term but also their second order interactions. The obtained equation has the form:

$$Rz = 11.15 + 2.889L^2 + 8.194 \times 10^{-6}v^2 + 155.1f^2 + 0.05842Lv + \dots \quad (8.35)$$

$$\dots + 20.54Lf + 0.8290vf - 7.349L - 0.05471v - 69.88f$$

The  $R^2$  has a value of 0.338, meaning that the model as fitted explains the 33.8 % of the variability in  $Rz$ . The standard error of estimation for this model is 2.161 and the mean absolute error is 1.460.

The analysis of variance for the quadratic regression (Table 8.5) shows that there is not a statistically significant relationship between the variables at the 90 % or higher confidence level, because the  $P$ -value is greater than 0.1.

The analysis of coefficient significance (Table 8.6) shows that only the parameters  $vf$  and  $L$  are statistically significant at the 90 % confidence level; so, all the other parameters can be removed from the model.

It can be concluded that, although a bit better than the linear regression, the quadratic regression is also unsuitable for describing the relationship between the studied variables.

Finally, a general linear model was used for describing the relationship between the variables. In this model, the lubrication,  $L$ , was considered as a categorical factor while the speed,  $v$ , and feed,  $f$ , were considered as quantitative factors. The obtained expression has the form:

**Table 8.4** Coefficient significance in the linear regression of  $R_z$

Parameter	Estimate	Standard error	$t$ -statistic	$P$ -value
Constant	6.363	1.079	5.899	0.000
$L$	-0.702	0.955	-0.735	0.468
$v$	0.024	0.016	1.540	0.134
$f$	-0.431	9.631	-0.045	0.965

**Table 8.5** Analysis of variance of the quadratic regression of  $R_z$

Source	Sum of squares	DoF	Mean square	$F$ -ratio	$P$ -value
Model	59.74	9	6.638	1.420	0.232
Residual	116.8	25	4.671		
Total	176.5	34			

**Table 8.6** Coefficient significance in the quadratic regression of  $R_z$

Parameter	Estimate	Standard error	$t$ -statistic	$P$ -value
Constant	11.14	2.798	3.984	0.000
$L^2$	2.889	3.090	0.935	0.359
$v^2$	$8.194 \times 10^{-6}$	$9.772 \times 10^{-6}$	0.008	0.993
$f^2$	155.1	332.961	0.466	0.645
$Lv$	0.05842	0.0374009	1.562	0.131
$Lf$	20.54	22.6906	0.905	0.374
$vf$	0.8290	0.373655	2.219	0.036
$L$	-7.349	3.81585	-1.926	0.065
$v$	-0.05471	0.104771	-0.522	0.606
$f$	-69.88	51.1926	-1.365	0.184

$$R_z = 6.012 + 0.6297c_1 - 0.5315c_2 + 0.02465v - 0.6521f;$$

$$\text{where } c_1 = \begin{cases} 1 & L = 0 \\ -1 & L = 1 \\ 0 & \text{otherwise} \end{cases} \quad ; \quad \text{and } c_2 = \begin{cases} 1 & L = 0.5 \\ -1 & L = 1 \\ 0 & \text{otherwise} \end{cases} . \quad (8.36)$$

This model has an  $R^2$  statistic equal to 0.113, so the model as fitted explains only the 11.3 % of the variability in  $R_z$ . the standard error of estimation is 2.287 and the mean absolute error 1.648.

From the analysis of variance (Table 8.7) it can be concluded that there is not a statistically significant relationship between  $R_z$  and the independent variables at the 90 % confidence level. A second ANOVA (Table 8.8) tests the statistical significance of each of the factors. Since all the  $P$ -values are greater than 0.05, there are no any statistically significant term at the 95 % confidence level.

As in the other cases, neither the general linear model is capable to properly describe the relationship between the studied variables.

**Table 8.7** Analysis of variance of the general linear model of  $Rz$ 

Source	Sum of squares	DoF	Mean square	$F$ -ratio	$P$ -value
Model	19.65	4	4.913	0.940	0.454
Residual	156.9	30	5.229		
Total	176.5	34			

**Table 8.8** Coefficients significance in the general linear model of  $Rz$ 

Source	Sum of squares	DoF	Mean square	$F$ -ratio	$P$ -value
L	7.819	2	3.910	0.75	0.4820
V	12.85	1	12.85	2.46	0.1274
f	0.02392	1	0.02392	0.00	0.9465
Residual	156.9	30	5.228		
Total	176.5	34			

It can be concluded that the statistical-based models are not able to deal with the studied data, due to the strong non-linearity and, probably, the high noise levels.

### 8.3.1.3 ANN-based Model

As an alternative, a neural network based model was fitted for the maximum roughness,  $Rz$ . For this purpose, a radial basis function network (RBFN) was selected, because it can model any nonlinear relationship, provided an enough number of neurons in the hidden layer and it can be trained very fast.

The selected network uses Gaussian transfer functions for computing the output,  $a_i$ , of the  $i$ -th neuron in the hidden layer:

$$a_i = f_i^{(h)}(\mathbf{x}) = e^{-\sum_{j=1}^N \frac{(u_{ij} - x_j)^2}{(\sigma_j / 0.8326)^2}}; \quad (8.37)$$

where  $\mathbf{x} = [x_1, \dots, x_N]^T$  is the vector of inputs;  $\mathbf{u}_i = [u_{i1}, \dots, u_{iN}]^T$  are the weights of the  $i$ -th neuron; and  $\sigma_i$  is the so-called spread of this neuron. In the studied cases, the same value of spread,  $\sigma = \sigma_1 = \dots = \sigma_M = 0.35$ , was selected for all the neurons in the hidden layer.

On the contrary, in the output layer uses linear transfer function for computing the output,  $y$ , of the single existing neuron:

$$y = f_i^{(o)}(\mathbf{a}) = \sum_{i=1}^M v_i a_i + b; \quad (8.38)$$

where  $\mathbf{v} = [v_1, \dots, v_M]^T$  are the weights and  $b$  the bias of this neuron.



In order to verify the generalization capabilities of the model, the data were randomly divided into a training set (31 elements) and a validation set (4 elements).

The training process was carried out iteratively, adding one neuron at a time, until reaching a prescribed sum squared error (*SSE*). In the analyzed case, a target *SSE* of 0.023 was used. The process stopped after 14 iterations (see Fig. 8.14).

The fitted model can be conveniently represented as a MATLAB function, which is shown in Code 8.1. As can be seen, the input values (lubrication, *L*; speed, *v*; and feed, *f*) are normalized in the interval [0, 1] by linear interpolation:

$$\tilde{x}_i = \frac{x_i - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})}. \quad (8.39)$$

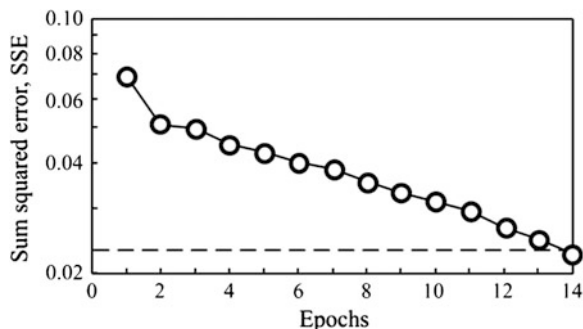
**Code 8.1** Implementation of the neural model for *Rz*

```
function Rz = rzann(LT, v, f);
    v = (v - 20)./(80 - 20);
    f = (f - 0.015)./(0.12 - 0.015);
    b1 = 2.3787;
    b2 = -0.0468;
    w1 = [ 0.0000  0.0000  0.0000; ...
          0.5000  1.0000  1.0000; ...
          1.0000  0.3333  0.1429; ...
          1.0000  1.0000  0.4286; ...
          0.0000  1.0000  0.4286; ...
          0.0000  0.3333  1.0000; ...
          1.0000  0.3333  1.0000; ...
          1.0000  1.0000  1.0000; ...
          0.5000  1.0000  0.0000; ...
          0.0000  1.0000  1.0000; ...
          1.0000  0.0000  0.0000; ...
          0.5000  0.0000  0.1429; ...
          0.5000  0.3333  0.4286; ...
          0.5000  0.0000  0.0000];
    w2 = [ 0.8520  0.6102  0.8709  0.4905 ...
          0.4459  0.5655  0.3846  0.4853 ...
          0.3321  0.2943 -0.4325  1.4281 ...
          -0.4826 -0.9994];
    x = [L, v, f];
    a1 = exp(-sum((w1.' - repmat(x.', 1, size(w1, 1))).^2).*(b1.^2)).';
    a2 = w2*a1 + b2;
    Rz = 3.77 + a2.*(12.3 - 3.77);
end
```

The neural model of *Rz* has a  $R^2$  statistic of 0.6757, meaning that the model as fitted explains the 67.6 % of the variability of *Rz*. This value is significantly higher than those computed for the statistical models. Furthermore, the standard error of estimation and mean absolute errors were 2.862 and 0.912, respectively.

The analysis of variance (Table 8.9) shows that there is a statistically significant relationship between the variables at the 90 % of confidence level.

**Fig. 8.14** Training process of the RBFN



**Table 8.9** Analysis of variance of the neural model of  $R_z$

Source	Sum of squares	DoF	Mean square	$F$ -ratio	$P$ -value
Model	112.0	15	7.46	2.07	0.0855
Residual	54.17	15	3.61		
Total	166.2	30			

In order to analyze the generalization capabilities of the model, a mean comparison between the residuals of the training and validation sets (Fig. 8.15), was carried out. The analysis gave a value of the  $t$ -Student statistic equal to  $-0.05837$ , with a corresponding probability value of 0.9538. Since the computed probability value is not less than 0.1, the null hypothesis (both means are equal) cannot be rejected with a 90 % of confidence level.

Finally, in Fig. 8.16, it is plotted a graphical representation of the obtained neural model and the corresponding experimental points.

Comparing the performance of the statistical regression models and the neural network, it can be concluded that the last one offers better predictions. This conclusion is supported by all the statistic test performed.

### 8.3.2 Modeling and Optimization of a Hot Forging Process

#### 8.3.2.1 Case Description

In the third case of study, the optimization of a hot forging process has been carried out. The part to be forged is a disk (Fig. 8.17) of AISI 1045 steel.

Two dies (bottom and top) were designed as Fig. 8.18 shows. The goals of the optimization process is to minimize simultaneously the billet forging temperature and force, as they have a strong influence in the economy of the process. As experimental factors, in this study were selected the forge temperature,  $T$ ; the forming speed,  $v$ , and the diameter of the cylindrical billet,  $D$  (as the volume is given, the height of the billet is determined by its diameter).

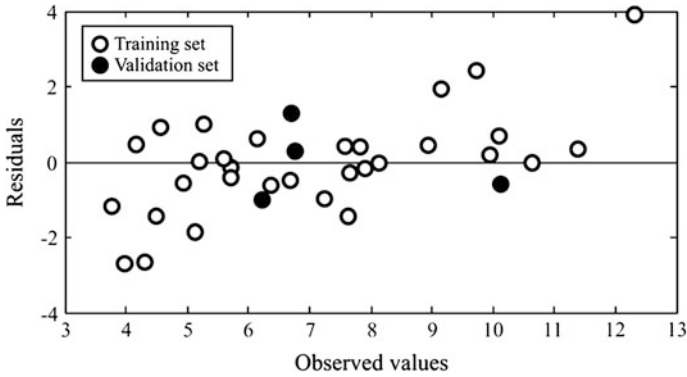


Fig. 8.15 Distribution of training and validation sets

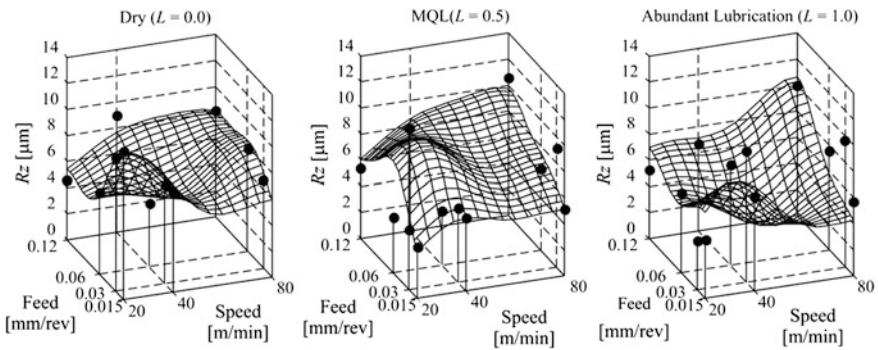


Fig. 8.16 Graphical representation of the neural model of  $R_z$

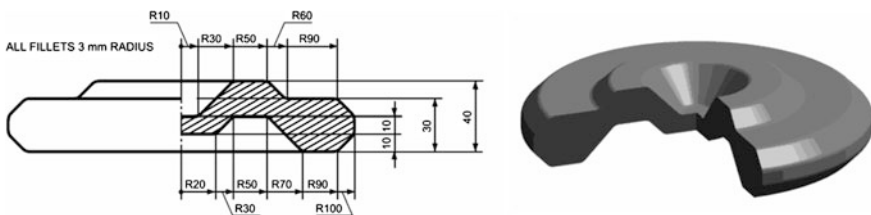


Fig. 8.17 Sketch of the finished part

8.3.2.2 FEM-based Simulations

In Table 8.10 the levels for these experimental factors are shown. With this factors and levels, a full factorial design was planned and finite element simulations were carried out in order to obtain the corresponding forging forces.

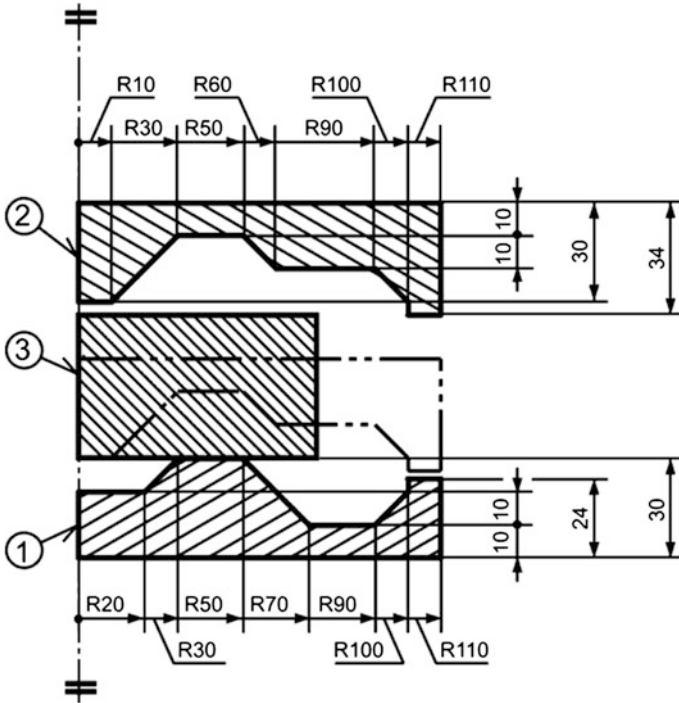


Fig. 8.18 Dies design (1: lower die, 2: upper die, 3: metal billet)

Table 8.10 Experimental factors in the hot forging process

Experimental factor	Level		
	Low	Medium	High
$T$ [°C]	900	1050	1200
$v$ [mm/s]	10	30	50
$D$ [mm]	100	150	200

The finite element-based modeling was established by considering the dies as rigid bodies. The plastic behavior of the workpiece material was considered as depending on the temperature, the strain and the strain rate (see Fig. 8.19). The temperature of the piece was considered as constant throughout all the process (isothermal forging). The workpiece mesh included from 2,000 to 3,000 elements, in every case. A friction coefficient of 0.25 was considered between the workpiece and dies.

In Fig. 8.20 it is shown the stress distribution through the forming process, for the medium levels of the factors (experimental point No. 14). Figure 8.21 plots the values of the pressing force through the time, for this experimental point. As can be seen, the maximum values of the pressing force are achieved at the end of the forging process, which match with the stress distribution (see Fig. 8.20d).

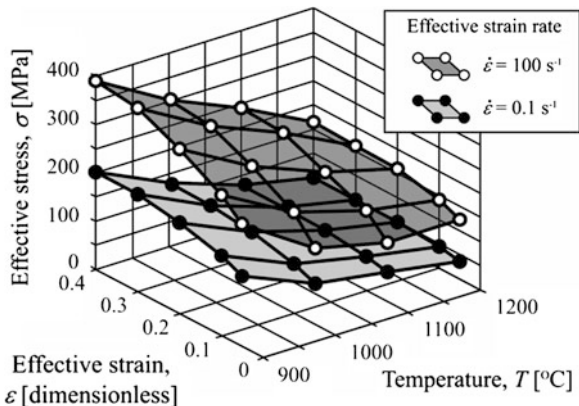


Fig. 8.19 Plastic behavior of the AISI 1045 steel

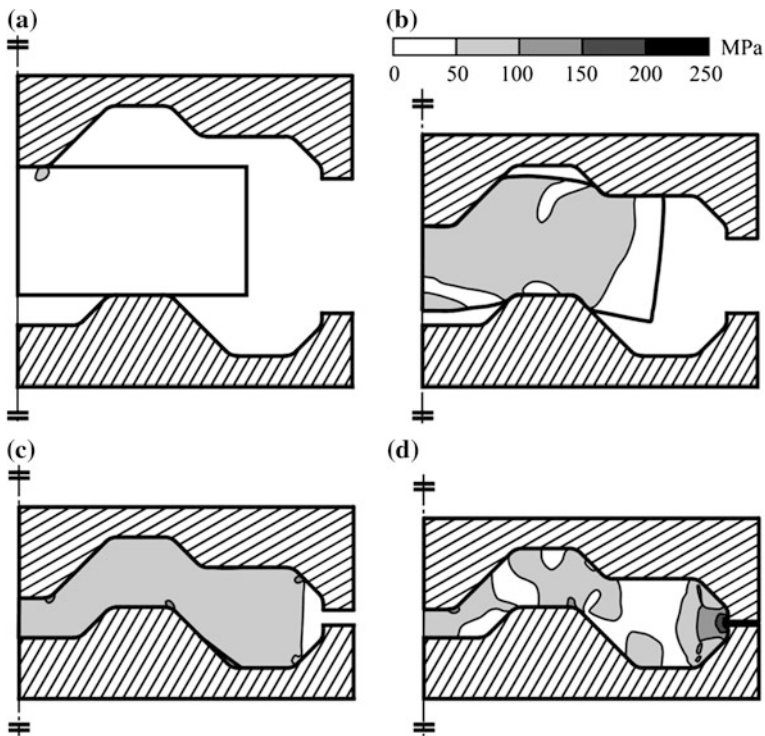
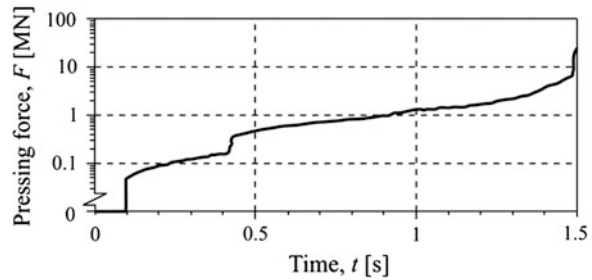


Fig. 8.20 Stress distribution through the process (experimental point No. 14). **a** Time: 0.09 s (stroke 2.8 mm). **b** Time: 0.70 s (stroke 22.3 mm). **c** Time: 1.38 s (stroke 4.6 mm). **d** Time: 1.50 s (stroke 45.0 mm)

**Fig. 8.21** Press force behavior through the time (experimental point No. 4)



**Table 8.11** Simulated pressing forces

No.	$T$ [°C]	$v$ [mm/s]	$D$ [mm]	$F_{\max}$ [MN]
1	900	10	100	45.0
2	1050	10	100	33.8
3	1200	10	100	18.2
4	900	30	100	54.6
5	1050	30	100	44.8
6	1200	30	100	19.8
7	900	50	100	64.0
8	1050	50	100	39.0
9	1200	50	100	23.7
10	900	10	150	51.6
11	1050	10	150	41.6
12	1200	10	150	19.0
13	900	30	150	59.4
14	1050	30	150	42.7
15	1200	30	150	23.8
16	900	50	150	58.6
17	1050	50	150	49.5
18	1200	50	150	27.4
19	900	10	200	55.7
20	1050	10	200	35.1
21	1200	10	200	19.1
22	900	30	200	55.2
23	1050	30	200	43.2
24	1200	30	200	24.0
25	900	50	200	60.8
26	1050	50	200	44.0
27	1200	50	200	26.8

In Table 8.11 the maximum values of the force, for each experimental point are shown.

**Table 8.12** Preliminary sensibility analysis of the coefficients

Parameter	Estimate	Standard error	<i>t</i> -statistic	<i>P</i> -value
Constant	148.2	5.866	25.25	0.0000
T	-11.23	0.005011	-22.40	0.0000
v	20.75	0.03758	5.521	0.0000
D	0.02333	0.01503	1.552	0.1343

**Table 8.13** Analysis of variance of the model of  $F_{\max}$ 

Source	Sum of squares	DoF	Mean square	<i>F</i> -ratio	<i>P</i> -value
Model	5414	2	2707	251.4	0.0000
Residual	528.4	24	10.8		
Total	5672	26			

**Table 8.14** Sensibility analysis of the coefficients

Parameter	Estimate	Standard error	<i>t</i> -statistic	<i>P</i> -value
Constant	151.7	5.5724	27.22	0.0000
T	-0.1123	0.005156	-21.77	0.0000
v	0.2075	0.03867	5.366	0.0000

### 8.3.2.3 Modeling of the Maximum Force

In order to model the maximum force, a linear regression analysis was performed. As can be noted from the sensibility analysis of the coefficients (Table 8.12), the *t*-Student test for diameter has an associated probability value greater than 0.10, so it is not statistically significant at 90 % or higher of confidence level.

After removing the diameter, a new regression analysis was carried out, giving the model:

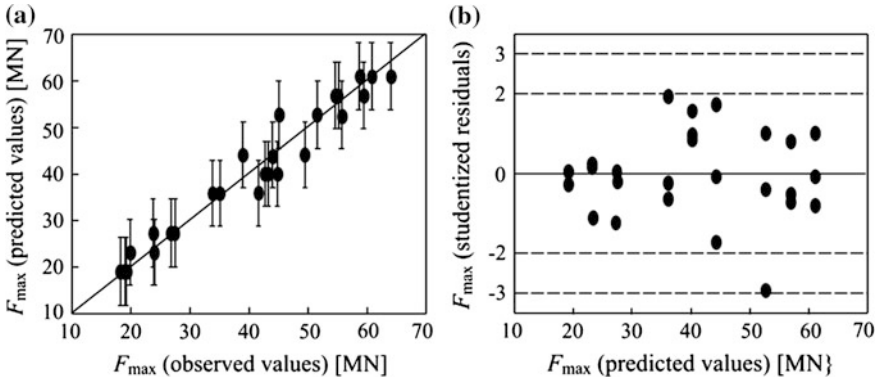
$$F_{\max} = 151.7 - 0.1123T + 0.2775v; \quad (8.40)$$

with and R-squared of 95.4 % and a mean absolute error of 2.36 MN.

From the ANOVA (Table 8.13) the probability values of the *F*-test for the model was lower than 0.01, indicating that there is a statistically significant relationship between the variables at 99 % of confidence level.

The sensibility analysis of the coefficient (Table 8.14) shows that all the parameters are statistically significant at 99 % of confidence level.

The graphical representation of the predicted values (Fig. 8.22a) shows that their confidence intervals contain the observed values in all the cases, indicating a good matching with the experimental results. Moreover, all the studentized residuals (Fig. 8.22b) are contained into the interval  $[-3, 3]$  and only one of them are outside the interval  $[-2, 2]$ .



**Fig. 8.22** Prediction capabilities of the force model. **a** Predicted versus observed values. **b** Studentized residuals

**8.3.2.4 Optimization of the Forging Process**

As the diameter has a significant influence on the process, as decision variables for the optimization problems were selected only the temperature,  $T$ , and the velocity,  $v$ . The objective functions were the forging time,  $t$ :

$$\tau = \frac{F_{\max}}{S}; \tag{8.41}$$

where  $S$  is the stroke (selected as 25 mm, corresponding the maximum diameter), and overall energy required for the process, computed as the sum of deformation work and thermal energy for heating:

$$E = F_{\max}S + cm(T - T_0); \tag{8.42}$$

where  $c$  is the specific heat of the workpiece material;  $m$  is the mass of the workpiece and  $T_0$  is the room temperature. For this specific problem,  $m = 5.81$  kg,  $c = 486$  J/(kg·K) and  $T_0 = 20$  °C. Both objectives must minimized.

The only considered constraint was the maximum force, which cannot surpass the allowable force by the forge machine:

$$F_{\max} \leq [F]; \tag{8.43a}$$

being  $[F] = 5\,500$  T = 54 MN. The constraints was rearranged, for efficiency purposes, into the form:

$$g(T, v) = \frac{F_{\max}}{[F]} - 1 \leq 0. \tag{8.43b}$$

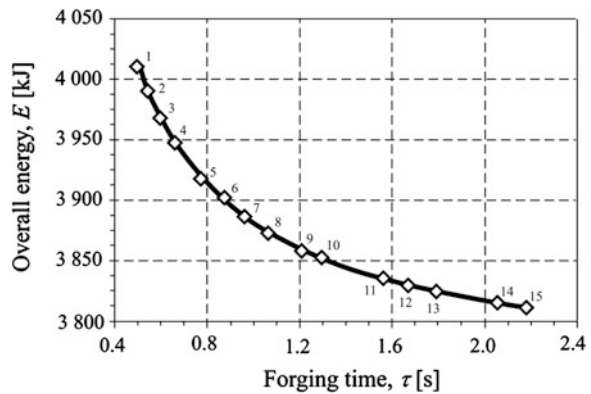
The optimization process was carry out by using a micro-genetic algorithm [11]. The static population size was established at 250; the dynamic population



**Table 8.15** Outcomes from the optimization

$v$ [mm/s]	$T$ [°C]	$E$ [kJ]	$\tau$ [s]
11.4	977	3811	2.18
12.2	1019	3816	2.06
14.0	1023	3825	1.79
15.0	984	3830	1.67
16.0	1085	3836	1.57
19.3	1057	3853	1.30
20.7	971	3859	1.21
23.5	971	3873	1.07
25.9	1020	3887	0.97
28.5	1160	3903	0.88
32.2	959	3918	0.78
37.8	987	3948	0.66
41.6	1004	3968	0.60
46.0	998	3991	0.54
50.0	977	4011	0.50

**Fig. 8.23** Graphical representation of the Pareto front



size, at 100; and the mutation likelihood, at  $10^{-4}$ ; The process was carried out for 20 evolutionary periods, each of them involving 50 epochs.

Table 8.15 shows the outcomes of the optimization process. As can be seen, all the points are non-dominated and all fulfill the constraint (Eq. 8.43b).

These solution can be represented as the Pareto front (Fig. 8.23), which should be used, by the decision-maker, to choose the most convenient combination. While the point 1 represents the highest forging time and the lowest energy requirements, point 15 includes the lowest forging time and the highest energy waste. All the other points are intermediate combinations. Evidently, an economical analysis based on the optimization results will help the decision-making process.

**Acknowledgment** The authors acknowledge the kind contribution of MSc. Rui Sendão and MSc. A. Festas in the execution of the experimental work on surface roughness of the titanium alloy drilling process.

## References

1. Aalst WMP, Rubin V, Verbeek HMW, Dongen BF, Kindler E, Günther CW (2010) Process mining: a two-step approach to balance between underfitting and overfitting. *Softw Syst Model* 9:87–111. doi:[10.1007/s10270-008-0106-z](https://doi.org/10.1007/s10270-008-0106-z)
2. Bonabeau E, Dorigo M, Theraulaz G (1999) *Swarm intelligence: from natural to artificial systems*. Oxford University Press, New York
3. Foster J, Barkus E, Yavorok C (2006) *Understanding and using advanced statistics*. SAGE Publishers, London
4. Friedman N, Geiger D, Goldszmidt M (1997) Bayesian network classifiers. *Mach Learn* 29:131–163. doi:[10.1023/A:1007465528199](https://doi.org/10.1023/A:1007465528199)
5. Grafarend EW (2006) *Linear and nonlinear models: fixed effects, random effects and mixed models*. Walter de Gruyter, Berlin
6. Lewis SG, Raval A, Angus JE (2008) Bayesian Monte Carlo estimation for profile hidden Markov models. *Math Comput Model* 47:1198–1216. doi:[10.1016/j.mcm.2007.07.002](https://doi.org/10.1016/j.mcm.2007.07.002)
7. Lyngsø RB, Pedersen CNS (2002) The consensus string problem and the complexity of comparing hidden Markov models. *J Comput Syst Sci* 65:545–569. doi:[10.1016/S0022-0000\(02\)00009-0](https://doi.org/10.1016/S0022-0000(02)00009-0)
8. Marler RT, Arora JS (2004) Survey of multi-objective optimization methods for engineering. *Struct Multidisc Optim* 26:369–395. doi:[10.1007/s00158-003-0368-6](https://doi.org/10.1007/s00158-003-0368-6)
9. Miao Q, Huang HZ, Fan X (2007) A comparison study of support vector machines and hidden Markov models in machinery condition monitoring. *J Mech Sci Technol* 21:607–615. doi:[10.1007/BF03026965](https://doi.org/10.1007/BF03026965)
10. Pernkopf N, Wohlmayr M (2013) Stochastic margin-based structure learning of Bayesian network classifiers. *Pattern Recogn* 46:4640471. doi:[10.1016/j.patcog.2012.08.007](https://doi.org/10.1016/j.patcog.2012.08.007)
11. Quiza R, Reis P, Davim JP (2006) Multi-objective optimization of cutting parameters for drilling laminate composite materials by using genetic algorithms. *Compos Sci Technol* 66:3083–3088. doi:[10.1016/j.compscitech.2006.05.003](https://doi.org/10.1016/j.compscitech.2006.05.003)
12. Rivals I, Personnaz L (2000) Construction of confidence intervals for neural networks based on least squares estimation. *Neural Netw* 13:463–484. doi:[10.1016/S0893-6080\(99\)00080-5](https://doi.org/10.1016/S0893-6080(99)00080-5)
13. Sarker R, Mohammadian M, Yao X (2003) *Evolutionary optimization*. Kluwer Academic Publishers, New York
14. Schneider JJ, Kirkpatrick S (2006) *Stochastic optimization*. Springer-Verlag, Berlin
15. Sha W (2006) Comment on prediction of the flow stress of 0.4C–1.9Cr–1.5Mn–1.0Ni–0.2Mo steel during hot deformation by R.H. Wu et al. [*J. Mater. Process. Technol.* 116 (2001) 211]. *J Mater Process Tech* 171:283–284. doi:[10.1016/j.jmatprotec.2005.07.004](https://doi.org/10.1016/j.jmatprotec.2005.07.004)
16. Utyuzhnikov SV, Fantini P, Guenov MD (2009) A method for generating a well-distributed Pareto set in nonlinear multiobjective optimization. *J Comput Appl Math* 223:820–841. doi:[10.1016/j.cam.2008.03.011](https://doi.org/10.1016/j.cam.2008.03.011)

# Chapter 9

## Implementing STEP-NC: Exploring Possibilities for the Future of Advanced Manufacturing

Kelvin Hamilton, Jean-Yves Hascoet and Matthieu Rauch

**Abstract** This chapter contains a summary of the current state of the ISO data model ISO14649 for Numerical Controller also known as STEP-NC. It details the reasons and need for an industrial STEP-NC paradigm shift by showing the benefits that would be immediately realizable using currently available tools and knowledge. Specific focus is given to the SPAIM application as it is one of the most advanced STEP-NC enabling applications available today that allows realizing those benefits. In considering the future possibilities of STEP-NC and the need for continued implementation, four important and complex topics are addressed. These topics would enable an increase in: interoperability through hybrid manufacturing environments, manufacturing supervision and traceability, flexibility and efficiency with high knowledge and information transfer as well as production optimization and simulation in multi-process manufacturing. Finally, a brief synopsis of the systems and components necessary for machine migration to STEP-NC using the SPAIM enabling application is given.

### 9.1 Introduction

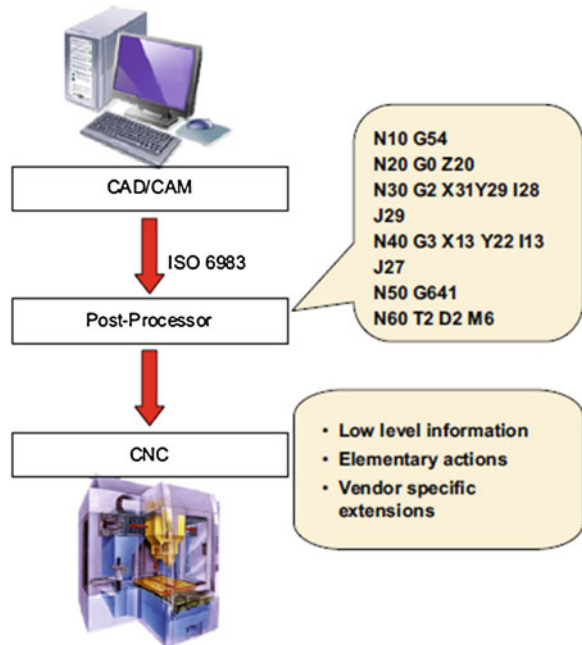
#### 9.1.1 STEP-NC, the Data Model

STEP-NC, known formally as the ISO 14649 standard, is an offshoot of the STEP standard (ISO 10303) which is centered around automation systems, representation of product data and exchange specifically between Computer Aided systems (CAx). However, unlike STEP, the distinction of STEP-NC is that it focuses primarily on the interaction between CAx systems and the numerical controller (NC).

---

K. Hamilton · J.-Y. Hascoet (✉) · M. Rauch  
IRCCYN UMR CNRS 6597, EC-Nantes, 1 Rue de la Noe, Nantes 44321, France  
e-mail: jean-yves.hascoet@irccyn.ec-nantes.fr; jean-yves.hascoet@ec-nantes.fr

**Fig. 9.1** Current G-code programming

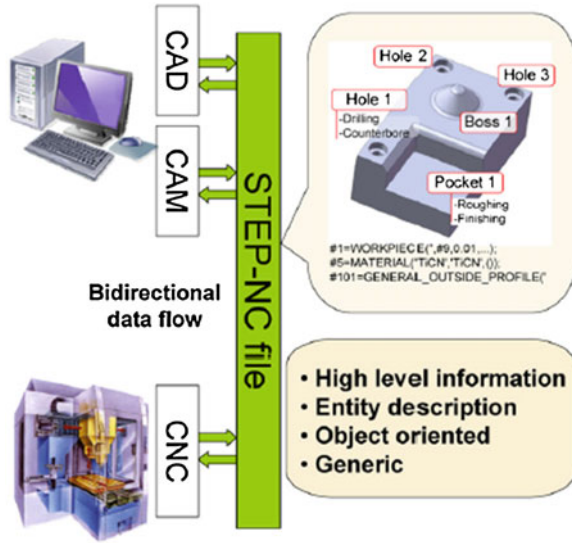


It aims to provide a seamless link with CAx systems while taking a strong perspective from the machine and machine-controller point of view. Using the defined backbone of EXPRESS data modeling language, STEP-NC shifts the focus towards a very important part of manufacturing, the machine-tools that are responsible for making the carefully designed parts.

The birth of STEP-NC is attributable to the need from industrial manufacturers as the revolution of the CNC machine continued. After having attained efficient and low cost productions with the CNC, manufacturers seek to meet the needs of the ever changing market by increasing flexibility, adaptability and improving productivity. To this end, the research community has been invaluable in providing new solutions, technological improvements related to cutting tools, machine-tools and CNC performances. However, little focus was placed on the current programming standard, which makes the link between the perfect CAD model and the real machined product as well as supporting machining orders and intelligence of CAM processing and simulations.

This current machine-tool programming standard is the ISO 6983 (G-codes) dating back to the early 1980s [1]. This standard with low level information describes elementary actions and tools moves, strongly reducing possibilities at the CNC level. Its linearly sequential nature (Fig. 9.1) breaks the CAD–CAM–CNC numerical chain and makes gathering feedback from the shopfloor difficult. The G-codes standard is one of the main limitations to flexibility and interoperability. To counter this limitation and to meet changing market needs, a new standard was required as manifested in the form of STEP-NC [2, 3].

**Fig. 9.2** New STEP-NC high level programming



STEP-NC provides new opportunities to support high level and standardized information from the design stage to fabrication by an NC controller. It allows bidirectional data flow between CAD/CAM and CNC without any information loss (Fig. 9.2). ISO 14649 diverges sharply from its ubiquitous predecessor, the G-code, because it does not describe the tool movements for a specific CNC machine as G-code does but rather provides a feature-based data model. A wide range of high level information is therefore made available such as feature geometry, cutting tool description, operation attributes and workplan. STEP-NC as a data model consists of several parts that contain information related to processes in general (Part 10) as well as process-specific parts. At present several processes have been defined including: Milling (Part 11), Turning (Part 12), and EDM (Parts 13 and 14). Other processes such as Additive Manufacturing are currently being defined.

A STEP-NC file, generally used for exchange, is not machine-tool specific and can be used on various machine-tool controllers. A new generation of intelligent controllers can interpret STEP-NC information to generate, simulate and optimize machining toolpaths among other things. The standard also includes a data model for inspection, which aims to provide a closed loop CNC machining environment. With the STEP-NC standard, the CNC controller becomes a central element in the design/manufacturing data chain and some intelligence is transferred from CAM to CNC. For example explicit toolpaths can be computed in the CNC controller itself with the help of an embedded CAM system or trajectory generator. Equally, a machine-tool functional model exists to provide critical information concerning the machine-tool and its controller.

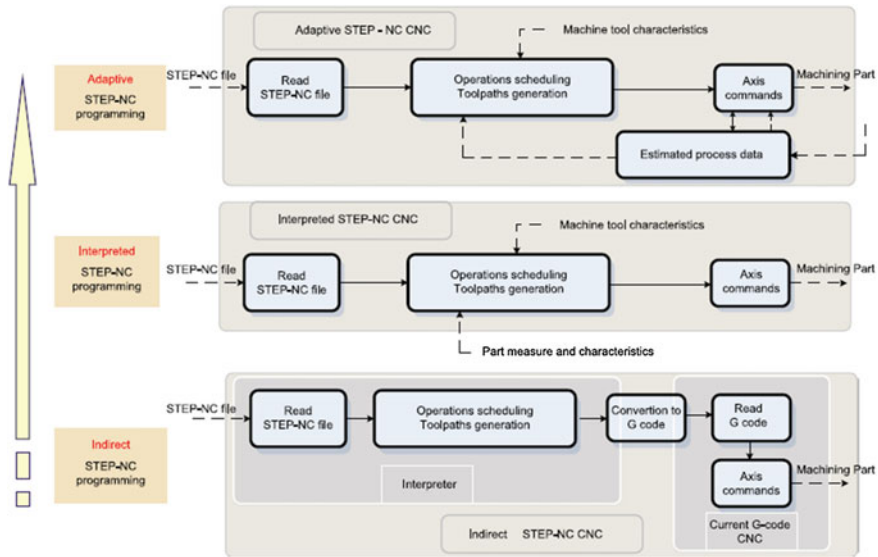


Fig. 9.3 Three-stage evolution towards advanced STEP-NC programming

### 9.1.2 STEP-NC, Step-by-Step Evolution

Although STEP-NC does promise new opportunities for machine-tool programming, a deep reorganization of the numerical chain at the CAM/CNC level as well as the human element (operator, programmers, CNC developers etc.) needs to take place. High level object-oriented information in STEP-NC has to be treated and executed on the machine-tool which at first requires an interpreter. The interpreter stands as a main part of the extended STEP-NC controller intelligence and carries out toolpaths generation and machining operations scheduling. This is the premise for STEP-NC migration and to realize it requires some new industrial mind shift and training. To this end, three-stages of evolution for STEP-NC in Fig. 9.3 were proposed as a way to gradually, with time and research, migrate the industrial mindset to the STEP-NC paradigm.

The first level is called *Indirect STEP-NC programming*. The use of STEP-NC with legacy NC controllers, that only read G-codes, is made possible. This level combines a standalone application to manipulate STEP-NC, alongside an interpreter embedded in the CNC. The behind-the-scenes STEP-NC is totally transparent from the user and the traditional operations she currently performs. This Indirect level is advantageous because it can very easily be spread on existing CNC equipment and they can profit from some STEP-NC benefits.

The second level is *Interpreted STEP-NC programming* where axis command is directly executed from STEP-NC. Tool and machine-tool functional model is taken into consideration for toolpaths generation and simulations. Integrated

bidirectional dataflow between CAD/CAM, as embedded systems, and CNC make this possible. External data such as probing results can be integrated in the toolpath generation process.

The third level is *Adaptive STEP-NC programming* where the NC controller estimates online process data and optimizes machining parameters and toolpaths in real time. This third level is the ultimate goal to achieve and relies on strong STEP-NC programming and interaction with associated systems.

Today, the ISO 14649 standard [4] is built as the Application Reference Model (ARM) of STEP-NC. It offers the opportunity to seriously think about the necessary content to meet the requirement of advanced programming. Some experimental controller platforms, demonstration tools and concept validation algorithms, described in Sect. 9.1.3, exist today as a way to convince the actors of the numerical chain of the interest of using STEP-NC.

### ***9.1.3 Experimental STEP-NC Enabled Prototypes***

Existing STEP-NC controllers can be sorted into these three categories of Indirect, Interpreted and Adaptive programming following the evolution levels STEP-NC.

Within the Indirect category and STEP-NC interpreters for G-code machines, (1) one of the first prototypes was realized within the context of the now terminated European STEP-NC project Esprit [5]. In this project, a STEP-NC file is generated by Catia (Dassault Systemes) and Open Mind CAD software. It is then interpreted by a modified Siemens 840D controller. Similarly, a STEP-compliant CNC interface (2) was proposed by the STEP Tools Company [6]. Their ST-Machine software generates ISO 10303-238 files (STEP-NC Application Integrated Model) and works as a front-end application on a current CNC controller. However, despite the generic nature conveyed by the use of ISO 10303-238 files, only a few capabilities envisaged by ARM (ISO 14649) are available. Another STEP compliant interpreter (3), developed at Loughborough University by Newman et al. was envisioned as an Agent-Based Computer Aided Manufacture system (AB-CAM) [7]. This prototype generates ISO 14649 part programs translated to G-code for machining applications. For NC milling applications, software-based CNC prototypes have been developed at the University of Auckland (4) by Wang et al. [8] and (5) by Minhat et al. [9] respectively. The first is an interpreter that stands as a front-end application to commonly used CNC controllers, translating STEP-NC data into G-codes. Conversely, the second prototype is an open CNC architecture based on STEP-NC and function blocks performing the task as a STEP-NC interpreter.

Within the Interpreted category, prototype development has been limited. In this category, the CNC controller does not require G-code as STEP-NC programming is totally integrated. The first prototype (6) was developed at the University of POSTECH in Korea [10]. The platform is based on several independent modules (STEP-NC file generator, toolpath generator, toolpath viewer, machine-tool driving

and control). The prototype enables direct control of the motion axes of a scale machine-tool dedicated to laboratory applications. It has not been implemented on an industrial machine-tool but proposes new solutions for axis command control. In a separate development by Xu, a CNC controller prototype (7) was implemented on a retrofitted CNC lathe and enables the realization of G-code free machining scenario [11].

Within the Adaptive category there are no prototypes implemented on any industrial machine-tool. As a first step to facilitate some of the expected features within this category, some proposals and framework have been introduced to help guide development. Shin et al. developed an interpreter for turning applications that converts G-codes files into STEP-NC files [12]. As a result, the introduction of STEP-NC standard for a company does not necessitate reprogramming all the existing machining files. Alternately, STEP-NC based machining optimization was proposed by Xu to optimize the machining parameters at CNC level using the high level information of STEP-NC data with explicit toolpaths [13]. Meanwhile, Ridwan et al. introduced a machining optimization framework based on STEP that enables feedrate optimization using process monitoring and control [14]. Strong focus of development and prototyping is still needed to properly demonstrate the mechanism and benefits of this advanced STEP-NC programming category.

One of the primary limitations for STEP-NC has been the lack of programming platforms based on ISO 14649 (ARM) implemented on existing industrial machining equipment. This is one of the key points, not only to demonstrate the capabilities of STEP-NC programming, but also to validate the new models and proposals for the standard. For the realization of this third category and provide an ARM-based implementation on existing machine-tools, the software application elaborated in Sect. 9.2 was proposed and developed. This application is showcased and detailed here because it is one of the most mature systems and has become a central piece in the demonstration of STEP-NC. Many of the future possibilities of advanced manufacturing based on STEP-NC has been developed and validated on this system and it serves as a platform on which future developments would be based.

## 9.2 STEP-NC Application Showcase: SPAIM

The use of high-level information to communicate with CNC controllers calls for a new organization of the different CAM and CNC modules (i.e. toolpaths generation module, workplan selection module, process parameterization module, etc.). In the case of milling, manufacturing a part consists of removing some material using a cutting tool driven by a NC controller. All the motions and actions such as axis commands are controlled by low-level information generated at the CNC level and all the information is totally transparent to the user. The electrical control and algorithms are not easy to be understood by humans. G-code programming can be seen as a relatively upper-level whose details can be understood by humans but



**Fig. 9.4** SPAIM platform at IRCCyN institute



still remains quite difficult to interpret. The low-level G-code information is directly translated by the CNC controller to generate elementary motions and actions on the machine-tool. In a direct contrast, STEP-NC programming is based on high-level information such as feature geometry or definition of the process data. The STEP-NC format is object-oriented and can still be understood by humans.

However, it is still necessary to communicate with the CNC controller and actuate the different parts of the machine-tool. This means that a translation from the high-level object-oriented information to well-adapted and accurate low-level information is necessary. This job, as explained, is carried out by the interpreter. Any STEP-NC interpreter is consequently machine-tool specific as it makes the link between the STEP-NC file and the data required to control the machine's axes. The structure of a STEP-NC interpreter can be built in several ways by using different technologies to compute the high-level information in STEP-NC files to lower-level information for machining. An example of this is seen in SPAIM.

SPAIM (STEP-NC Platform for Advanced and Intelligent Manufacturing) is a platform based on STEP-NC Interpreted Programming approach, which allows an implementation on most industrial CNC controllers such as the one at the IRCCyN laboratory in France (Fig. 9.4). Its implementation associates two main objectives. First it stands as a demonstrator to showcase the benefits of STEP-NC and secondly as a development platform for future STEP-NC research and validations.

STEP-NC object-oriented programming has helped to shift the toolpath generation to the shopfloor level. As a result, some intelligent and decision-making power can be transferred into the CNC controller. By extension, self-learning algorithms could be utilized to produce better quality parts by error compensating [15]. In the SPAIM implementation, the interpreter module is a key part of the controller because it translates STEP-NC manufacturing data into explicit toolpaths using manufacturing feature geometrical characteristics and programming parameterization of each manufacturing step.

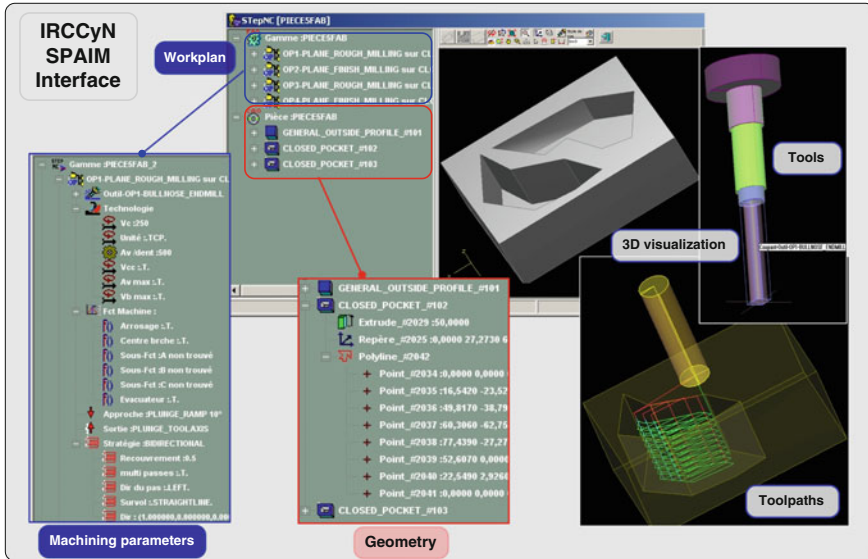


Fig. 9.5 STEP-NC human/machine interface on the NC controller

Although it generates explicit machining toolpaths automatically, the interpreter still calls for user validation before sending the toolpaths to the machine-tool. This stage is considered compulsory because it alerts the user of the expected movements that the machine-tool. Furthermore, the user can check whether proposed toolpath meet the manufacturing constraints. If not, the user can make modifications as necessary. After validation, corresponding output file (G-codes, for first level programming) is automatically executed by the controller.

This platform is composed of a Human Machine Interface (HMI) and several computation modules. The HMI on the NC Controller displays a 3D visualization of the manufacturing data (CAD model, toolpaths, cutting tools, etc.), the machining parameters, and a tree-view of the STEP-NC data (Fig. 9.5). The user can modify these data using the interface. After modification, CAD models and explicit toolpaths are automatically regenerated and the STEP-NC file is updated. Machining can then be executed directly from the interface following a visual validation of the toolpath, machining parameters and modifications.

Following this implementation, the platform can enable any CNC to read and handle STEP-NC data built according to the ISO 14649 standard. This has already been implemented and validated on a high-speed manufacturing machine-tool with parallel kinematics architecture. Designed by Fatronik, this machine called 'VERNE' is equipped with a Siemens Sinumerik 840D NC controller [16].

Another version of the platform has also been implemented for a 5-axis machine again equipped with a Siemens 840D controller and it is used for the laser cladding Additive Manufacturing process.

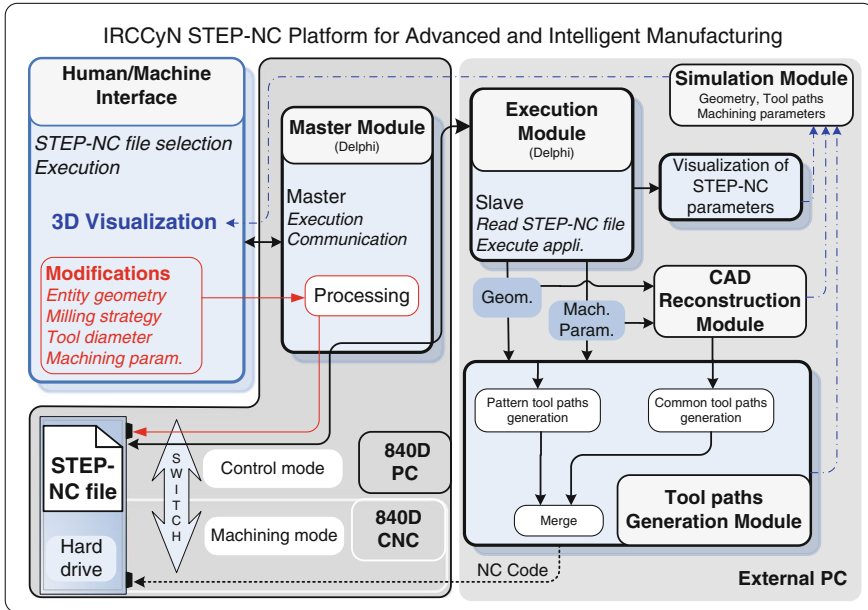


Fig. 9.6 Architecture of SPAIM

### 9.2.1 SPAIM Architecture

SPAIM is composed of several modules controlled by Delphi applications (Fig. 9.6). Some are installed directly on the CNC controller while others are on an external computer to reduce the computation load on the CNC. However, all the modules running on the external PC can be implemented in the CNC computer if its computing capacity allows it. The external PC can be seen as an extension of the capacities of the CNC hardware.

The main modules include:

- *Human Machine Interface*: the user can control the CNC platform using this module. It displays the results of analysis and computations made by the interpreter using data from the simulation module on the screen.
- *Master module*: implemented on the NC controller, this module is directly linked with the HMI and sends the orders to other modules at the user's request. For parameter modification, it locates and replaces the corresponding elements in the STEP-NC data.
- *Execution module*: distributes orders from the Master module via a local or an Internet network. This module reads and analyses the STEP-NC file through the master module and to send the requested information for processing to the toolpath generation module and to the simulation module.

- *CAD reconstruction module*: rebuilds the CAD geometry from the entity description in the STEP-NC data. This automatic tool sends corresponding commands to Delcam PowerSHAPE CAD software [17] for geometry reconstruction. The CAD model is then used by the toolpath generation module to generate common strategies. This module provides feedback from the STEP-NC data to the CAD model as well.
- *Toolpath generation module*: is divided into two components running in parallel. The first handles every common strategy defined in the ISO 14649 standard (e.g. contour parallel, bidirectional, etc.). It uses the toolpath generation module of a commercial software (i.e. Delcam PowerMILL [18]). The second, developed at IRCCyN, handles pattern strategies such as those used for trochoidal and plunge milling toolpaths. According to the manufacturing data, the execution module collects the corresponding toolpath generation module for each machining operation and merges the toolpaths results before sending the NC code to the controller for execution.
- *Simulation module*: sends back the results of the different computations (e.g. STEP-NC data analysis, 3D geometry and explicit toolpaths as a VRML file, etc.) to the HMI for visualization.

The toolpath generation module is based on a vendor component for the generation of common strategies. The SPAIM platform can benefit from the skills and performance of the CAM software. Moreover, it shows that even if the numerical chain is redistributed, all the current knowledge is still needed contrary to the belief of some CAD/CAM vendors. The strong interaction between STEP-NC, CAD/CAM and the CNC is what allows some of the benefits to be obtained.

### ***9.2.2 Benefits of a STEP-NC Enabled Controller***

As stated, STEP-NC interpreted programming make it possible to use the STEP-NC standard within existing machine-tools and NC controllers. SPAIM shows the feasibility of this and can be used as a prototype for future implementations. At this first level of the integration of STEP-NC, most of the advantages of STEP-NC at the CNC level are already evident (Fig. 9.7):

- (1) STEP-NC data can be read and executed directly on several machine-tools equipped with an interpreter without any modification to the data. This compatibility is enabled by the high-level description of geometry and process data without any specificity to a single machine-tool. All machine-tool information and functional models would be provided by the CNC platform. For example, the execution of the same STEP-NC file on different machine structures would use different spindle speeds, feedrates or toolpaths, because each NC controller would compute the most suitable and efficient parameterization, according to the target equipment.

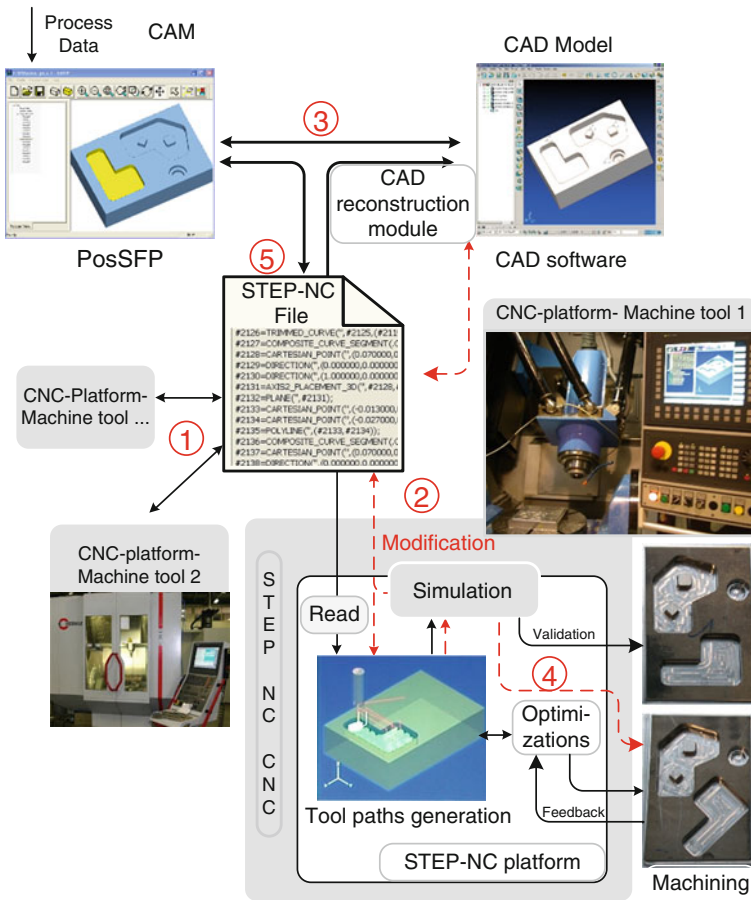
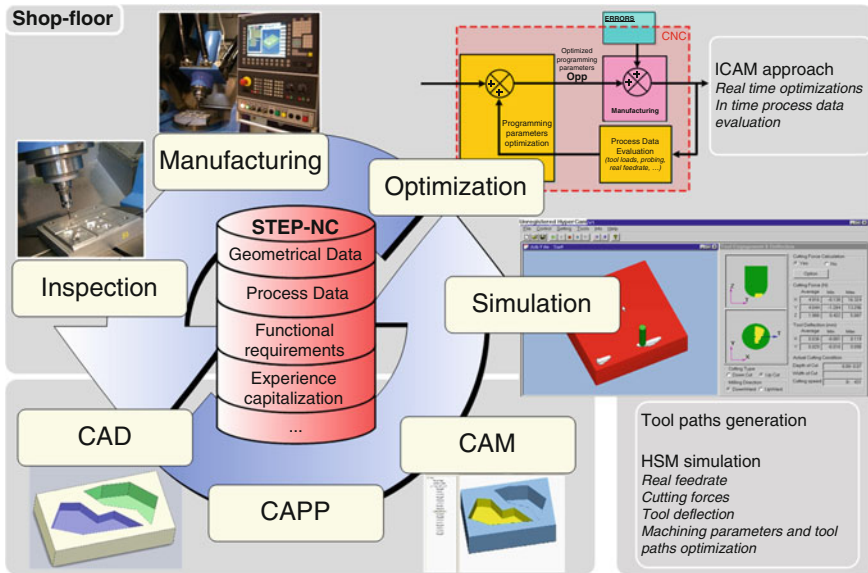


Fig. 9.7 STEP-NC platform numerical chain

- (2) Modifications of the geometry and the machining parameters can be achieved at the shopfloor level directly on the CNC HMI. These modifications automatically lead to regeneration of the toolpaths, geometry displays and STEP-NC model tree update. The corresponding STEP-NC data and CAD model are updated as well.
- (3) Feedback from CNC to CAD/CAM software is possible since the STEP-NC data is always up to date. Modifications can be done at shopfloor level during the first manufacturing phase of a part. This knowledge feedback enables process planning level to learn and improve the future manufacturing phases.
- (4) Optimizing the machining parameters and the toolpaths is easier at the CNC level. SPAIM allows new non-linear optimization based on the STEP-NC data.



**Fig. 9.8** Advanced programming in a comprehensive STEP-NC environment

- (5) STEP-NC data transfer file has a small size since it contains only high-level data. This therefore would reduce transfer time making it well suited for Internet-based collaborative manufacturing.

### 9.2.3 Advanced CNC Programming

The SPAIM platform opens the door for advanced programming methods. As a demonstrator and development platform validation of the ISO 14649 standard, it permits new simulation and optimization approaches to be implemented by taking into account high-level STEP-NC data. Optimization developments previously using g-codes such as a real feedrate simulation module and a tool deflection module can also be included in the STEP-NC framework with some adaptation. The implementation of these modules will be based on research works already done at IRCCyN on 3D solid simulation [19] and tool deflection compensation [20]. As illustrated in Fig. 9.8, SPAIM in its implementation makes each part of the manufacturing numerical chain interoperable. It presents a comprehensive environment dedicated to advanced programming.

Based on this integrated environment, implementation of Intelligent Computer Aided Manufacturing (ICAM) methods is possible and some have already been achieved. The ICAM method was developed to allow toolpath programming optimization based on real-time process information to produce better and more accurate parts. For example, toolpath regeneration based on on-machine inspection

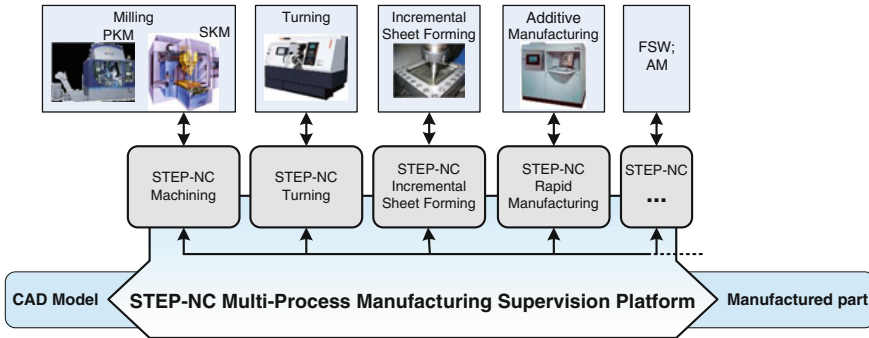


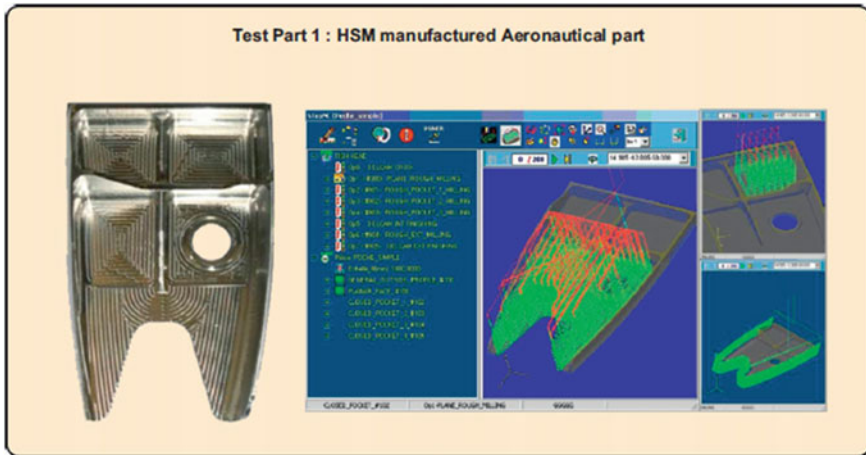
Fig. 9.9 STEP-NC multi-process hybrid manufacturing concept

has been carried out with G-codes with promising results. The real-time adaptations of machining parameters (i.e. feedrate) and toolpaths are possible by using process data evaluation from the CNC controller (motor amperage, delivered power, real feedrate, articular coordinates of the joints, etc.). The scope of future possibility can and should be expanded further and having such a platform at the center of development has many benefits most importantly in closing the loop between once isolated manufacturing systems and sharing of knowledge and experiences between systems and processes.

In fact, it is true to say that developments at IRCCyN on the STEP-NC standard and advanced NC programming methods with SPAIM are not limited to milling. Instead, they include other processes such as additive manufacturing, incremental sheet forming, turning and additive manufacturing processes amongst others. The development and integration of other processes such as wire electro discharge machining is also an important consideration [21]. The development of the STEP-compliant process models and platforms is essential to achieve the goal of interoperability as later discussed.

The long-term purpose of these efforts is aimed at the development of a comprehensive STEP-NC multi-process supervision platform shown in Fig. 9.9. In such an environment, the digital model of a part can reflect the modifications and updates concerning each manufacturing process. Moreover, interactions between two processes will be available and encouraged. In this sense, the manufacture of industrial parts which need a combination of several manufacturing processes will benefit by information sharing leading to part and process optimization. The use of the STEP-NC standard will ensure interoperable and bidirectional data flow between all stages of the product development process. As a result, STEP-NC multi-process supervision concept optimizes the whole design and manufacturing chain from a CAD model to the manufactured product.

All of the opportunities of STEP-NC promise to increase production flexibility, process optimization and efficiency by harnessing the linking foundation that STEP-NC enables as well as enjoying high knowledge and information transfer. Some implementations have been accomplished to help showcase these opportunities



**Fig. 9.10** Aerospace STEP-NC test part ‘The Fishhead’

(Sect. 9.3). However there are areas that remain scarcely explored that hold tremendous possibility to increase productivity, interoperability and flexibility beyond current industrial capabilities. Some researchers have shown a glimpse into these possibilities by defining frameworks and proposals as to how to best reach these possible new heights [9, 14, 22, 23, 24, 25].

## 9.3 Current Possibilities with STEP-NC

### 9.3.1 Flexibility in Milling (+ Other Processes)

Two of the key benefits of STEP-NC highlighted were the ability to read and execute the same part program on multiple machines by using the high-level non-specific data; as well as modify part geometry and machining parameters directly on the shopfloor. These benefits are important because of the increased flexibility they add in terms of part and program modification. Two separate workpieces can be used to describe these benefits. The first is the now famous aerospace testpart, the Fishhead, used around the world for demonstrating STEP-NC (Fig. 9.10). The second is a simple part with pockets, bosses and a hole with counterbore (Fig. 9.11).

These workpieces and the associated scenarios would be applicable to any current industrial CNC that have been made STEP-NC compatible with the addition of the SPAIM application. This would allow them to be capable of Indirect or Interpreted STEP-NC programming and benefit instantly of advanced flexibility.



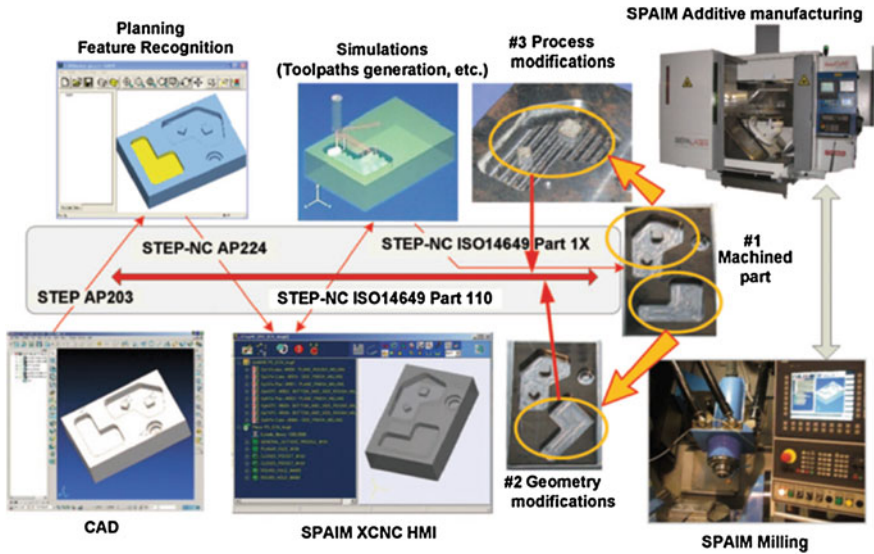


Fig. 9.11 Increasing flexibility through shopfloor modification using STEP-NC

The manufacture of the first part, the Fishhead shown in Fig. 9.10, is straightforward: 3-axis milling for roughing operations followed by 5-axis finishing operations. Using SPAIM, the manufacture of this part is completely flexible and adaptable to the current conditions on the shopfloor. This flexibility involves the following scenarios: (1) a required cutting tool is not available, damaged or broken. Solution: the operator finds a suitable replacement of similar characteristics but of a different size, he or she modifies the associated tool or adds a new one to the STEP-NC data; (2) an expert operator at the shopfloor recommends a last-minute modification to a particular milling strategy due to changed clamping locations, part reorientation, tool chatter etc. Solution: operator modifies the associated strategies in the STEP-NC data.

For each of these scenarios and modifications, the STEP-NC data is updated and the modifications archived for traceability. The new data is fed back to CAD/CAM/simulation modules to accommodate changes, the updates are then simulated, and new explicit toolpaths are generated for execution. All these activities are performed without leaving the machine’s HMI. Each of these scenarios were not originally envisioned in the process planning stage but are completely supported with the multidirectional STEP-NC programming.

In a series production environment, this flexibility is useful for the manufacture and qualification of the first parts used to validate the process. After the validation and the start of production, the ability to quickly and easily switch between similar CNC machines becomes very useful again since STEP-NC is not machine-specific.

The necessary modules used by SPAIM would be available within the CNC CPU environment. The advantage would therefore be seen through increasing production agility, reducing downtimes and delays as well as improving macro process planning around damaged machines for example.

For non-series production, these benefits of flexibility and adaptability are equally applicable, even essential, since variability and process/production demands could be more significant. This would therefore necessitate high production agility to meet these demands.

The manufacture of the second part, the pockets with bosses shown in Fig. 9.11, is also straightforward: pocket milling with an appropriate strategy to accommodate the bosses. The true power of STEP-NC is evident and essential when one considers the increased ability of a shopfloor or factory to adapt and optimize their operation even as scheduling priorities and machine maintenance occur. This second test part emphasizes how optimization and last-minute modifications could enhance flexibility and adaptation to current shopfloor conditions [25].

The steps and computation necessary to manufacture this part with SPAIM is shown in Fig. 9.11. From the part CAD model, the feature recognition module is executed and the generic STEP-NC data are created. For each feature, the manufacturing process parameters are selected according to the operator assisted by process simulation tools. An initial part (#1) was directly machined from the input of STEP-NC data into SPAIM. This initial part, conventionally machined, serves as a reference to which flexibility possibilities would be judged.

Using SPAIM's HMI, several tests were conducted with part geometry and machining parameters modifications directly on the machine on the shopfloor. As an added benefit, such modifications are not only limited to the shopfloor. Direct modification can also be attained from distributed and distant computers that share the SPAIM application modules. Two examples of modifications are presented in Fig. 9.11, the first deals with geometry changes and the second deals with manufacturing process change. For the first modified part, a rotation of a feature was performed (#2) showing possible geometry modifications on the shopfloor or at the final stages of production process planning. The new data is recorded and feedback to CAD and simulation modules thus taking into account the modified data. The toolpath generation module updates the strategies and the part is re-machined. For the second modified part (#3), the manufacturing operation to produce the two bosses was changed from milling to additive manufacturing (elaborated in Sect. 9.4.4). The platform, which is distributed on both milling and additive manufacturing machines tools, carried out the required operations automatically: toolpath generation, validation and control. A feedback to macro process planning data in CAD and inter-process simulation is also concurrently enabled from the modified STEP-NC data before re-machining the part. Such scenarios are common place in manufacturing and following the old G-code paradigm shown in Fig. 9.1 would require going back up the numerical chain to isolated CAD/CAM environments and making the necessary modifications. The resulting changes would

then be transferred to the CNC after post processing, all of which are time-consuming activities. With STEP-NC and the SPAIM application, the constant feedback and dialogue between the associate systems means that the previous limitations can be circumvented resulting in unparalleled flexibility and a work lose?: coming. The range of this application is certainly not limited to the milling process as already explained but it could be adapted to numerous other processes as well. As manufacturing conditions constantly change, higher level systems such as Process Planning can also benefit by the increased agility of the shopfloor and new optimized non-linear planning can be established as a result of this capability.

The illustration with the Fishhead and the ‘Pockets with bosses’ parts serve to emphasize how current industrial CNC machines can exploit the Indirect and Interpreted levels of STEP-NC. Manufacturers will instantly see increased flexibility and agility as they are able to modify part geometry, machining parameters, and even optimize processes and production with direct input from the shopfloor. With the constantly changing manufacturing conditions, manufacturers are searching for ways to be even more agile and flexible. These inherent benefits of STEP-NC directly provide a solution to meet these higher demands of the factory and the ubiquitous machine-tools that support it.

To understand how these capabilities can be obtained by manufacturers, a list of necessary components is detailed in [Sect. 9.7](#). Manufacturers are encouraged also to visit the IRCCyN laboratories in France for a first-hand look and demonstrations.

## 9.4 Advanced and Future Possibilities with STEP-NC

The many benefits of STEP-NC and the applications to exploit it are well documented as described above. The capability for higher flexibility and agility in a single process is unparalleled using STEP-NC and they serve as a way to allow the manufacturing industry to migrate to the STEP-NC framework in a step by step manner. However, from the Authors point of view, despite these very useful capabilities, it is the scarcely explored future possibilities of the STEP-NC paradigm that are even more important and in need of serious attention. These possibilities understandably represent a shift in the mindset of the manufacturing industry and the adoption of a different method of system interaction. This section elaborates on these possibilities that are seen as the next frontier in the evolution of manufacturing methods. The following topics are discussed:

1. Interoperability through hybrid manufacturing environments
2. Improved manufacturing supervision and traceability
3. Flexibility and efficiency with high knowledge and information transfer
4. Production optimization and simulation in multi-process manufacturing.

### **9.4.1 STEP-NC in: Manufacturing Interoperability**

It is known that in the current global context, a major problem lies in data portability between systems in a manufacturing data chain that is often distributed [10].

The ability of STEP-NC to support intelligent NC programming applications has been explained and demonstrated. It offers a vendor-neutral solution with numerous opportunities for improving manufacturing cohesiveness and production collaboration. A general consensus is that STEP-NC can stand as a valid solution to respond to current portability and interoperability needs. Numerous research projects have proposed and elaborated this idea: the Universal Manufacturing Platform developed at the University of Bath [23] and the UbiDM (Design and Manufacture via Ubiquitous Computing Technology) of Pohang University of Technology [24]. Despite current implementations of STEP-NC being limited to research and feasibility demonstrations, they represent a solid foundation for future developments that will be used to persuade the manufacturing industry and break the current roadblocks to wider adoption.

#### **9.4.1.1 SPAIM + IIMP**

Extending the capabilities of STEP-NC as an interoperable solution, two manufacturing platforms (SPAIM and the IIMP project from the University of Auckland) can be seen to fill this role adequately, either separately or integrated.

The framework of SPAIM, already explained, bridges the barrier between the typical systems used in the manufacturing numerical chain (CAD-CAM-CNC) with a particular focus at the shopfloor. Enabling interoperability amongst these systems, SPAIM represents a comprehensive STEP-NC multi-process supervision platform (Fig. 9.9).

Another system that has similar capabilities is the Intelligent and Interoperable Manufacturing Platform (IIMP) [26]. This system augments the capabilities of SPAIM by extending manufacturing efficiency and interoperability. Data portability between the heterogeneous proprietary formats of CAD/CAM/CNC systems and process interoperability is also realized in the IIMP. Although IIMP and SPAIM can be implemented individually, they collectively form a complete framework that covers the whole manufacturing data chain. By interacting with every link, this framework improves supervision and integration of the machining systems (Fig. 9.12).

The focus of the IIMP is to foster a collaborative environment among existing and future CAD/CAM/CNC systems. It relies on the high level information in STEP-NC to make interoperable links between the associated systems within the manufacturing data chain. Like SPAIM, it keeps commercial CAD/CAM/CNC data and software as vital well-established knowledge and information sources. IIMP does not try to link all systems and their data perfectly but provides a mechanism via STEP-NC that allows it to convey only the small subset of

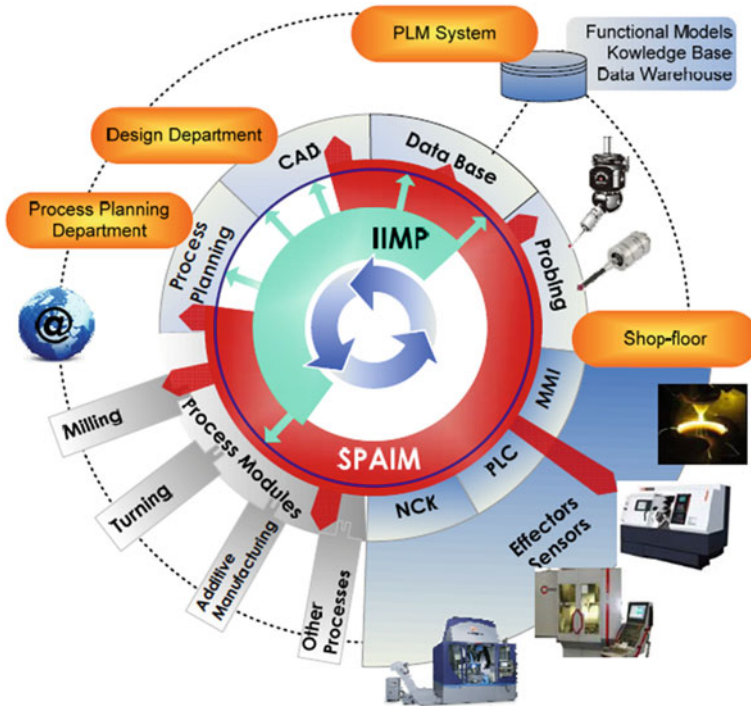


Fig. 9.12 Complementarities of SPAIM and IIMP in for advanced CNC manufacturing

accurately defined data between systems. This allows homogeneity at the platform level compared to the heterogeneity amongst the associated systems.

### IIMP Architecture

The architecture of the IIMP is composed of three main element groups: an orchestrator, an application module pool and an execution core (Fig. 9.13).

The orchestrator's role, the main decision center of the platform, is to act as the main HMI. It get user requests, generate a roadmap of their project and control information flow between modules and the execution core via the event-driven supervisor. Due to this central role, the orchestrator's capabilities directly impact the platform's efficiency. The roadmap generator has a direct link to the application module pool and it is used to translate the user's request into a list of events and tasks that each application modules will perform in a predefined order. To fulfill the user's request, the previously defined roadmap is followed. Template scenarios and roadmaps also exist as a way to transfer knowledge from previous requests and to optimize execution of repeated tasks and projects.

With the roadmap generated, the event-driven supervisor calls the application modules according to the roadmap and the resulting information is stored in the

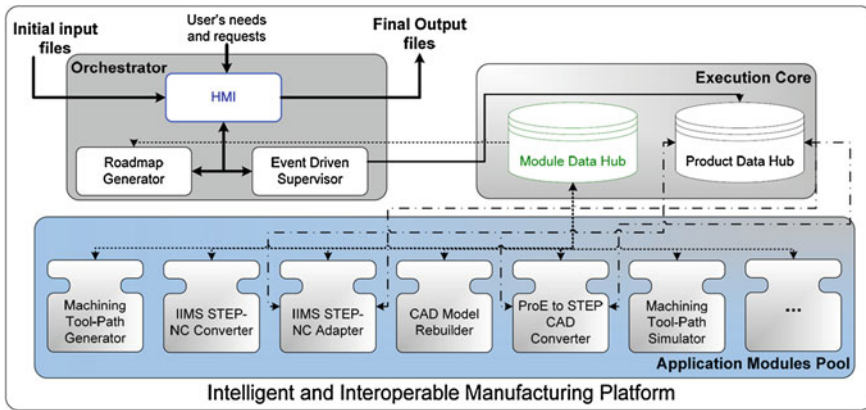


Fig. 9.13 Overview of the intelligent and interoperable manufacturing platform

execution core. The supervisor controls the information flow within the platform modules and the product data hub of the execution core. The execution core is responsible for run-time data exchange and storage by utilizing two hubs. The module data hub focuses on the application module pool and it stores all relevant data for the orchestrator to generate the roadmap and control the information flow during the process. The product data hub manages internal information during project execution. The application module pool consists of different standalone modules that perform specific tasks as directed by the supervisor, including toolpath generator, CAD file converter, CAM functions etc.

Two key features of this platform are flexibility and robustness which are obtainable through its modular design. It was designed to encapsulate existing software tools with an event- and data-driven layer used to control their input and output data. This results in no internal interactions between the modules and there is no restriction concerning the number and the nature of the modules [9]. The design (Fig. 9.14) follows the function block concept of the IEC-61499 standard for distributing industrial-process measurement and control systems [27].

The primary goal of the IIMP is to gather and synchronize heterogeneous data from CAD/CAM/CNC systems. It does have some limitations such as significant dependency on the orchestrator's efficiency as well as optimization trade-off for flexibility. However, it does offer a viable solution, complimentary to SPAIM, to the interoperability of distributed manufacturing with STEP-NC.

The issue of interoperability is an important one. It is a key motivation for the proposal of hybrid manufacturing environments, such as these, which are able to make existing software systems cohabitate with the new STEP-NC programming approaches. This allows continual development and permits step by step migration into the STEP-NC framework. Building on the knowledge obtained in implementing these systems represents a strong resource for future developments.

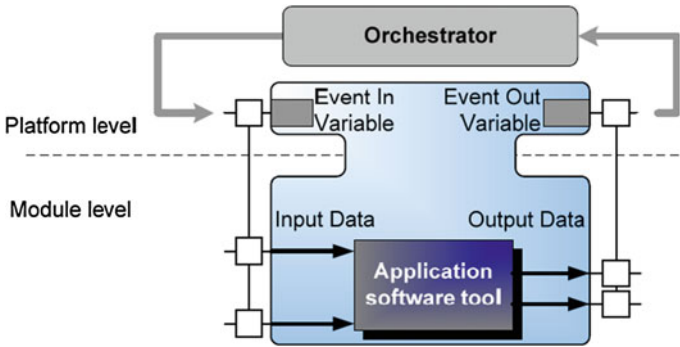


Fig. 9.14 Function Block concept for the application module pool

### 9.4.2 SPAIM + XMIS in a Digital Factory

Going even farther beyond the interoperability capabilities of the process-focused standard SPAIM and IIMP frameworks, in 2011, Laguionie et al. proposed the extension of STEP-NC from the process manufacturing-level to the higher enterprise-level systems so that a complete integration of the enterprise numerical chain could be possible [25]. This would see the integration of applications from enterprise level (PLM, Enterprise Resource Planning (ERP), etc.), plant level (Manufacturing Execution System, Quality, Control, etc.) and low-level process automation and controls (NC programming, monitoring, inspection, etc.). They proposed an eXtended Manufacturing Integrated System (XMIS) for feature-based manufacturing with STEP-NC, integrating information from design to manufacture.

#### XMIS Architecture

The goal of the XMIS is to control the manufacturing process from design to product, by integrating Manufacturing Engineering, Manufacturing Quality and Validations and Manufacturing Production data (Fig. 9.15). XMIS is based on several units managed by a Production Project Unit (PPU). A multi-directional collaboration between the units allows their integration in an extended manufacturing numerical system. Feedback from each layer of manufacturing is enabled for experience capitalization, process optimizations and process planning. It is consequently adaptable to the specific needs and evolutions of the company.

XMIS is composed of the following main units:

Production Project Unit (PPU) Stores and manages high-level project data for production. It makes the link with the PLM platform by including Design and Development, Analyst, Prototyping, Marketing, Service and Support, Supplying, Sales, etc. Relevant information is shared not only with different services within the company but also with external actors (manufacturing partners, customers, etc.). A main goal of the PPU is to control the visibility of data. While its objective is to convey the right information to the right place for a specific need.

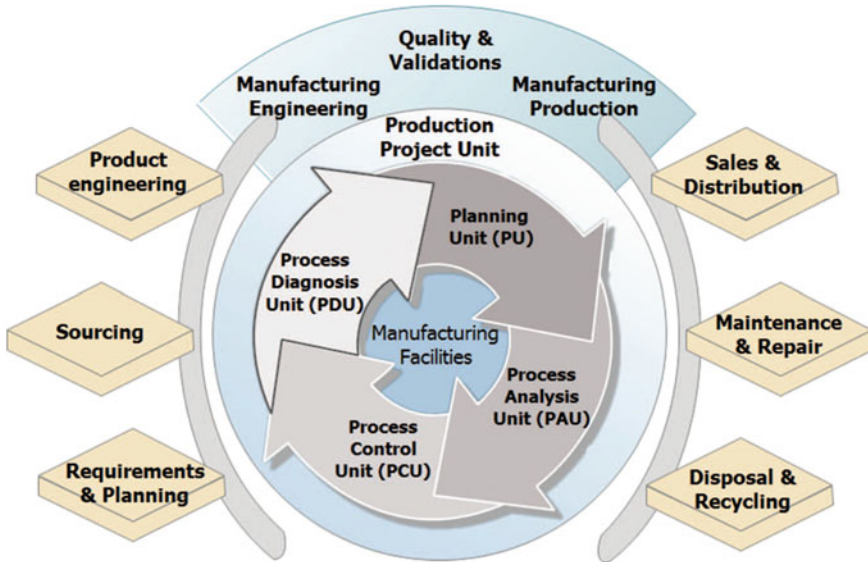


Fig. 9.15 The XMIS in a PLM environment

*Planning Unit (PU)* Treats the CAD model of a part for sales process planning. Several aspects are considered by linking with PPU. Manufacturing workplan and features are selected according to available processes, facilities, resources and manufacturing partners.

*Process Analysis Unit (PAU)* Runs part manufacturing simulation, optimization and verification. This includes operation sequencing, tool and machining conditions choices, toolpath (TP) generation, etc. This unit is linked with the manufacturing data warehouse and the manufacturing knowledge database in a bidirectional way. Such a link allows for the inherent benefit of optimization from experience and experience capitalization of the best solutions.

*Process Control Unit (PCU)* Executes process online control and monitoring. It performs online compensations, optimizations, diagnosis and adaptive control. Depending on the company structure, this unit can share applications at shopfloor level, on manufacturing facilities level and at production management level.

*Process Diagnosis Unit (PDU)* Runs measuring and analysis tools for part conformance, production quality and validation. It includes machining post-diagnosis and corrections. Shopfloor experience capitalization can be implemented and a feedback to other units is allowed, thanks to the Manufacturing Information Pipeline (MIP) directly linked with the PPU.

Each of these main units make up the XMIS and they each gather tools and modules that have to communicate together using industrial information and communication standards. The communication standards supported by XMIS are selected to facilitate interoperability. Consequently, to their limited range, proprietary vendor-specific pseudo-standards are used in a limited way when necessary.



Open communication standards such as XML and MTConnect are used to facilitate reliable communication and interconnectability between heterogeneous hardware and software applications. Meanwhile, the case to extensively use the STEP-NC data model in XMIS is an important consideration. Although it is a relatively young standard and in need of further development and implementation, STEP-NC represents a compromise that bridges the gap between specific and generic manufacturing data. It covers a very large range of manufacturing data to suite different environments while at the same time enabling data feedback from the shopfloor to CAx systems.

The STEP standard is also used within XMIS since it offers a unified standard to describe all the aspects of a product during its life cycle. Defined STEP application protocols are dedicated to different application domains from design to maintenance. It is used widely in the industry and integrated in most CAD/CAM systems for design data exchange. The XMIS concept proposes to combine STEP and STEP-NC for internal and external data exchange between modules and units from design to NC and inspection.

Within XMIS, two categories of STEP-NC data are identified: Generic STEP-NC data and Optimized STEP-NC data. These two categories are influenced by the scope (either Macro or Micro Process Planning) in which STEP-NC data is needed. Figure 9.16 describes the necessary inputs and units required to create both the Generic and Optimized STEP-NC data.

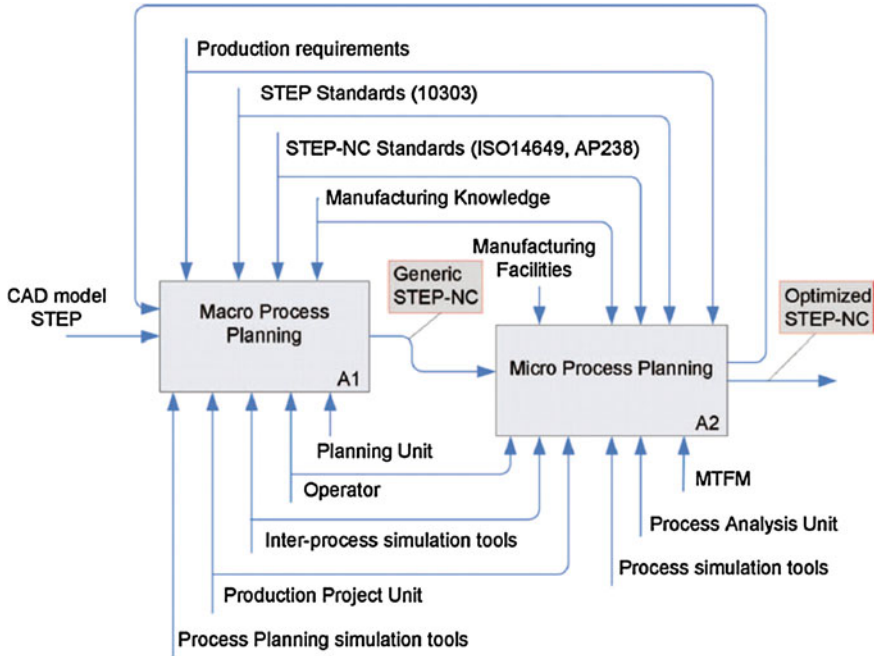
In Macro Process Planning (MacroPP), the focus is on process choices based on manufacturing features definition and work planning requirements. Performed by the PU, MacroPP treats parts manufacturing in a multi-process context.

With the help of inter-process and intra-process planning simulation tools in the PU, MacroPP gathers information to create manufacturing operations workplan by using Generic STEP-NC data. In this scope, Generic data only depends on process choices and therefore totally independent from any specific machine-tool or manufacturing equipment.

In Micro Process Planning (MicroPP), all the activities related to particular manufacturing resources and parameters are gathered together. Performed within the PAU with the help of specific intra-process and inter-process simulation tools, MicroPP is linked with the machine-tool functional models (MTFM) of the selected manufacturing facilities. All the information generated is optimized for a manufacturing resource within the PPU and supported by optimized STEP-NC data. Links to other departments of the company can be made and the optimized STEP-NC data can be directly executed on the CNC controller.

## **XMIS MIP**

Facilitating the exchange of information within XMIS is the MIP shown in a 3D representation in Fig. 9.17. The MIP makes a sharp departure from the traditional top-down data flow of the current numerical chain. Instead it uses multi-directional data exchange, supported by standards, enabling data exchange horizontally between modules (M<sub>ui</sub>) of the units, vertically between the units and radially between all modules and the PPU.



**Fig. 9.16** Generic and Optimized STEP-NC data in XMIS

The modules are integrated in the units as ‘plug and play’ while the Unit Interpreter links the computation results of the units with the PPU Databases. Local Unit Databases store the internal unit computation data in the native module standard and can be interpreted in vendor-neutral standards (STEP or STEP-NC) for inter-modules data exchange. The PPU manages databases shared between the different units (Manufacturing Knowledge Warehouse, TP database and MTFM). Data visibility is controlled at the PPU level for units to have access to the relevant information. PPU also manages visibility and security of manufacturing information exchange with all the other stakeholders services of the company (design, sales, supply chain, etc.) and partners (customers, collaborators, etc.). Machining facilities at the shopfloor are totally integrated in the numerical chain.

The MIP supports machining control and monitoring from the PCU. From the STEP-NC-specific data, PCU directly pilots the machining resources using a STEP-NC compliant controller. The use of current industrial controllers is also enabled by adapting TP and machining parameters for specific NC. Quality of the production is measured and validated by the Production Diagnosis Unit. Online and offline inspections are supported by ISO 14649 (Part 16) also called STEP-NC inspection. Sensor feedback from machine-tool to Production Control and Diagnosis Units can be supported by open communication standards. Experience capitalization and data management are allowed from design to machined part.

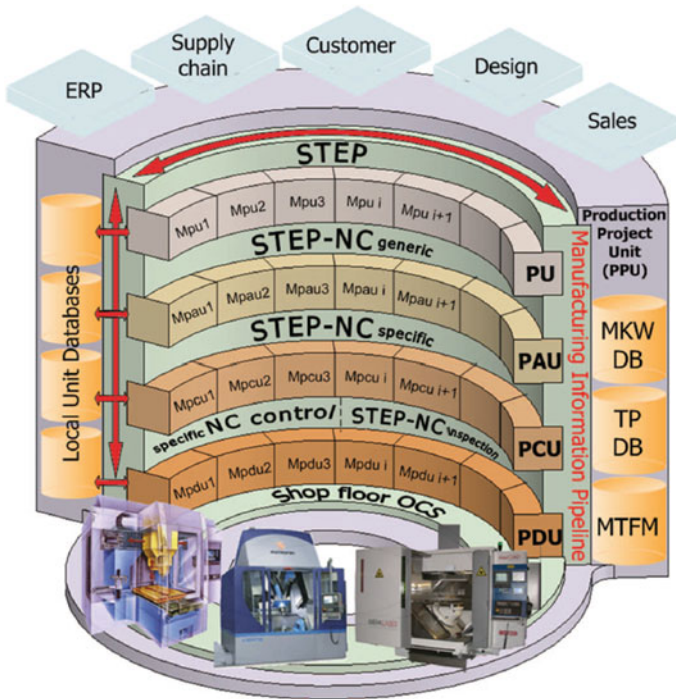


Fig. 9.17 XMIS units and the manufacturing information pipeline

Hence, the MIP supports common standards for improving interoperability at every stage of the manufacturing numerical chain.

Given the structure and architecture of the XMIS units and the MIP, they could be adapted to any enterprise according to their needs, requirements and capabilities. This therefore completely supports the ‘Design anywhere, Build anywhere, Support anywhere’ paradigm expected of manufacturers today in the evolution of the Digital Factory.

To demonstrate and validate this XMIS concept, the capabilities of the standard SPAIM application was extended (Fig. 9.18). A version exists that demonstrates parts of the XMIS system, tailored to the needs at IRCCyN. Some XMIS units are totally implemented and tested as part of the first demonstrator of the XMIS. Extended CAD (CAD + PU) and CNC (CNC + PDU, PCU, PAU) environments are defined and implemented to encompass milling and AM machines.

The concept of XMIS applied on industrial equipment is promising particularly since its design is modular with multi-directional high-level data exchange. Using both STEP and STEP-NC data to increase the interoperability between different vendors-specific applications will result in significant reduction in the information loss while at the same time improve knowledge capitalization. As technological

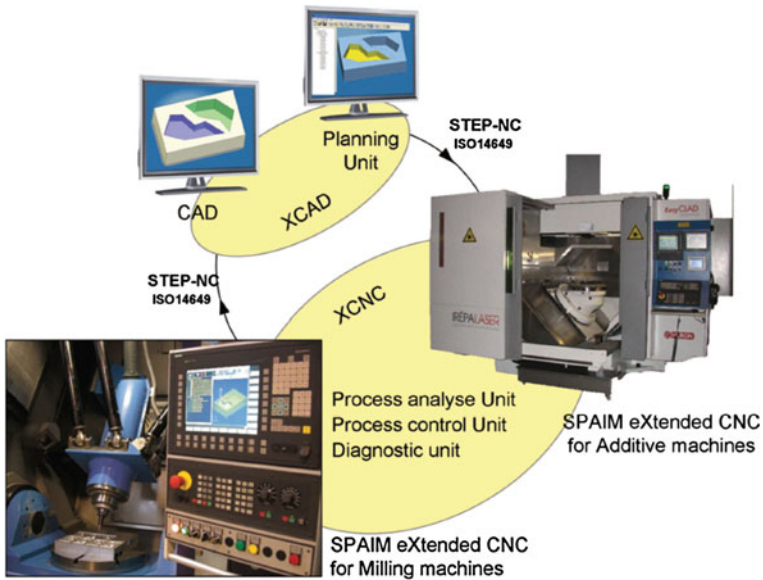


Fig. 9.18 SPAIM in XMIS

capabilities in manufacturing continue to improve, the implementation of XMIS will lead to the realization of feature-based manufacturing with STEP-NC, integrating information from design to manufacture.

### 9.4.3 STEP-NC in: Open-Architecture CNC (Open-NC)

Within the scope of the FP7 European Commission funded project, FoFdration, the development and extension of the STEP-NC framework into an open architecture CNC has been proposed and developed. Centered around a vision of the Next-Generation self-learning, intelligent and efficient CNC controller, the Consortium defined the FoFdration Smart Machine Controller Open Architecture (FSMC-OA) [28]. The resulting controller prototypes uses currently available tools and new developments to build an Open-NC platform for showcasing CNC functionalities that are expected even demanded, on future CNC machine-tool. It is also a way to adequately showcase the inherent open-ness, portability, scalability and interoperability of new applications and processes (Fig. 9.19).

Efforts within this project have seen the transformation of an industrial CNC (Cincinnati Milacron ‘Sabre’ milling machine with a NUM controller) into a unique multi-controller Integrated Test Platform (ITP). Alongside the original NUM-750 numeric controller, two separate CNCs have been added: the LinuxCNC (formerly EMC2) open architecture controller and the legacy nC-12 controller from Fidia.

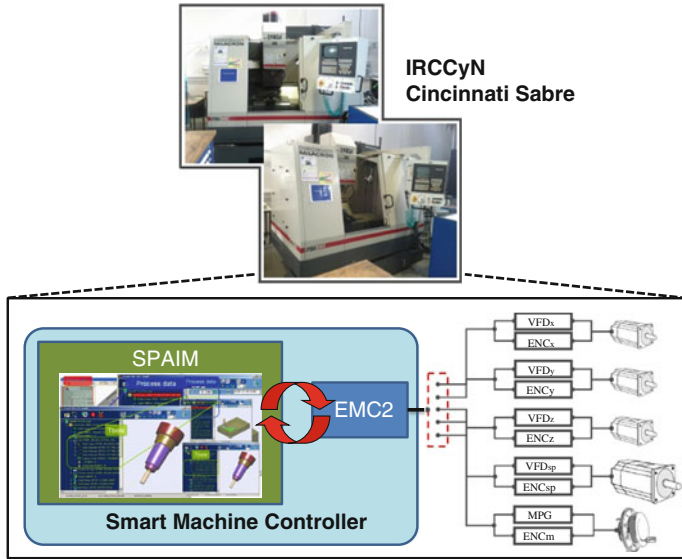


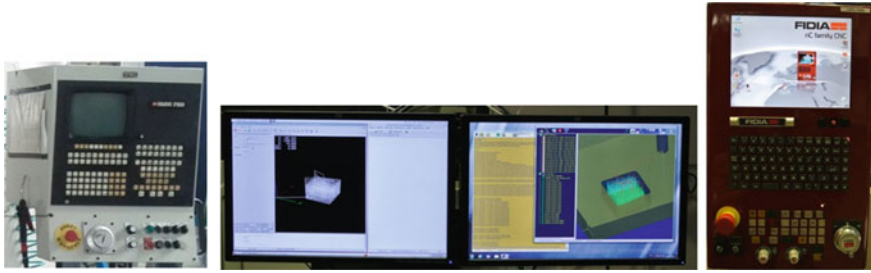
Fig. 9.19 IRCCyN’s Sabre machine-tool equipped with the Open-NC controller

They all co-exist within a single machine-tool, sharing the same physical configuration. This implementation allows selecting, via a physical signal switch, which controller actuates the Sabre’s axes (spindle, linear machine axes and handwheel).

The ITP was created to realize and demonstrate the evolutionary steps and benefits of STEP-NC. Furthermore, it serves as a unique all-in-one station to concretely illustrate how current legacy CNC controllers can, in the short term, benefit from and eventually migrate to the STEP-NC paradigm in the medium term. It also sets the stage for long term future developments towards the Next-Generation self-learning, intelligent and efficient CNC controllers. Using the three physical controllers mentioned, the ITP can be operated within 4 different CNC controller environments: NUM, LinuxCNC-ISO, LinuxCNC-STEP-NC and Fidia (Fig. 9.20).

*NUM:* The NUM controller is capable of conventional 3-axis CNC machine control and milling similar to any industrial vertical milling machine. The simple controls and wide use makes this controller a good candidate for conventional milling activities.

*LinuxCNC (ISO + STEP-NC)* The LinuxCNC controller possesses 2 separate operational modes thanks to its Linux OS and Windows OS on-board computers. It can be used as a conventional ISO-code driven CNC controller providing state-of-the-art 3-axes control for the Sabre’s axes using its Linux based real-time kernel. The Windows extension PC means that a host of software from CAD, CAM to algorithms and in-house developed software can be used with this CNC. This is an important point because testing new developments and algorithms is either not possible or difficult to do on propriety CNCs. Therefore having such an



**Fig. 9.20** HMIs on ITP: NUM-750, LinuxCNC-ISO, LinuxCNC-STEP-NC and Fidia nC-12

open architecture CNC removes this limiting barrier. The LinuxCNC can also be used as a STEP-NC compliant controller with the addition of the SPAIM platform for advanced STEP-NC programming and control.

*Fidia nC-12 CNC* The Fidia controller is also a conventional CNC platform based on Windows OS with Fidia-developed servo drives, PLC and other I/O peripherals. It provides state of the art CNC functionalities and advanced control. As it operates within the Windows “de facto” standard environment, implementation of third party software such as CAD/CAM is possible. This fact also makes it a good candidate to demonstrate STEP-NC compliancy again with the addition of the SPAIM platform.

The ITP focuses on addressing 3 main points:

- (1) STEP-NC programming evolution in the three levels: Use the unique multi CNC controller environment provided by the ITP to demonstrate the first (Indirect) and second (Interpreted) STEP-NC evolution stages while continuing the evolution of the third stage (Adaptive) in STEP-NC advanced programming.
- (2) Real-time process data: Combining the flexibility of STEP-NC provided by SPAIM with algorithms and the ICAM methodology developed for toolpath programming optimization based on real-time process to produce better and more accurate parts.
- (3) NC-Interpolation and trajectory planning: Develop smoother and more accurately controlled toolpaths driven by optimization.

The IPT (Fig. 9.21) is seen as an important tangible demonstrator for manufacturers to understand how they can begin to benefit with STEP-NC. It will also serve to show the steps necessary to make a legacy CNC machine-tool STEP-NC compliant. Manufacturers will be encouraged to see and make use of some of the many benefits of STEP-NC today on their existing machine-tools and NC controllers allowing them to realize higher flexibility and process optimization.

Researchers can also benefit from this unique machine-tool with multiple controller environments. It is well known that no two machine-tools or CNC controllers, even from the same manufacturer, perform equally. This makes it difficult to perform substantial comparison between different controllers and



Fig. 9.21 The FoFdration integrated test platform

machines. Therefore, having a single machine configuration with multiple controllers permits a more appropriate comparison that is not hindered by large machine to machine variation. The ITP can allow this to happen and can be used for tests and research activities ranging from comparing machine and controller performance, to STEP-NC developments and milling process optimizations. All of which are underpinned by extending the flexibility and efficiency of manufacturing tools with high knowledge and information transfer.

### 9.4.4 STEP-NC in: Optimization and Simulation for Multi-process

#### Optimization

Extending the STEP-NC standard to other processes is an important step in making this paradigm shift widespread. Amongst the existing processes already defined, the ISO 14649 part 13 [29] focuses on wire Electro Discharge Machining (EDM) [30] and other works are done to build a STEP-NC data model for Additive Manufacturing (AM) [31]. By supporting all these different kind of process data, STEP-NC offers the ability to integrate, supervise, link and make interoperable all these processes using a unified integrated manufacturing numerical chain. Each process has its own specificities that result, if following the current G-code numerical chain, in the isolation of their implementation. Information available in G-code files cannot be exploited as it is too low level: only a small quantity of the data linked to the machining process is included. As STEP-NC supports a large field of object-oriented data from CAD to CNC, it allows integrating multiple

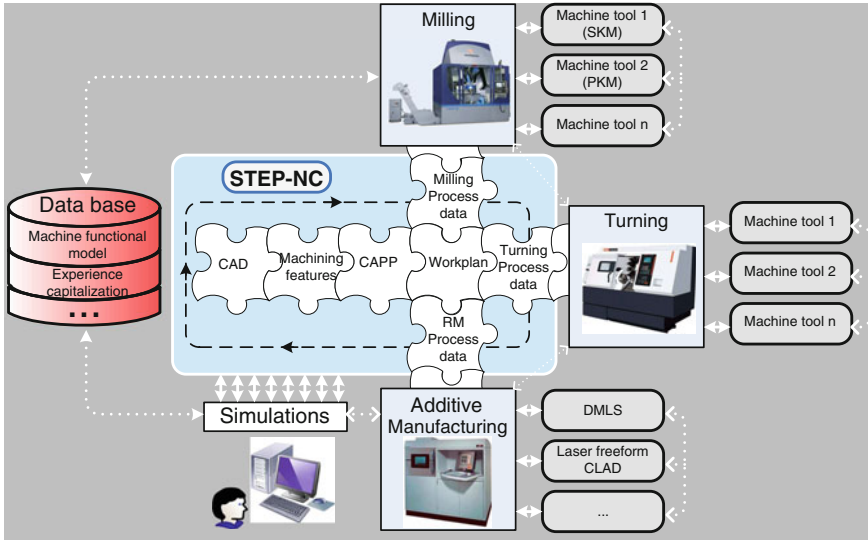


Fig. 9.22 STEP-NC multi-process numerical chain

processes through a common standard for several machining processes. To support such a configuration, the innovative concept described by Fig. 9.9 is essential for comprehensive multi-process supervision. The part model can reflect the modifications and updates concerning every manufacturing process. When process dependencies are necessary, STEP-NC is an appropriate standard to unify, support and link various process data. Whereas the current numerical chain only relies on expert users’ practices, STEP-NC offers new possibilities to integrate a comprehensive data support. It opens the way for a strong communication hub between experts by using a common well-adapted language. It would also provide data support for computational simulation and optimization of the multi-process workplan. However, multi-process approach involves more and more complex scenarios (Fig. 9.22) for which large efforts and continued research are needed to developed and validate a totally integrated numerical chain.

With the interaction of multiple processes and experts, decisions can be made not only by considering individual processes but by integrating the requirements of the whole manufacturing numerical chain. This interaction has an implicit benefit of allowing extensive part and process optimization throughout the factory.

Additive Manufacturing

A novel process that lends itself quite nicely to the STEP-NC framework as well as to the multi-process capabilities is Additive Manufacturing. This process has been defined in STEP-NC by researchers at IRCCyN and it opens a new range of manufacturing possibilities. Since material is added instead of removed, compared to other conventional processes, the types and complexity of new part creation that can be produced is greatly increased. Even more useful is the possibility of part and



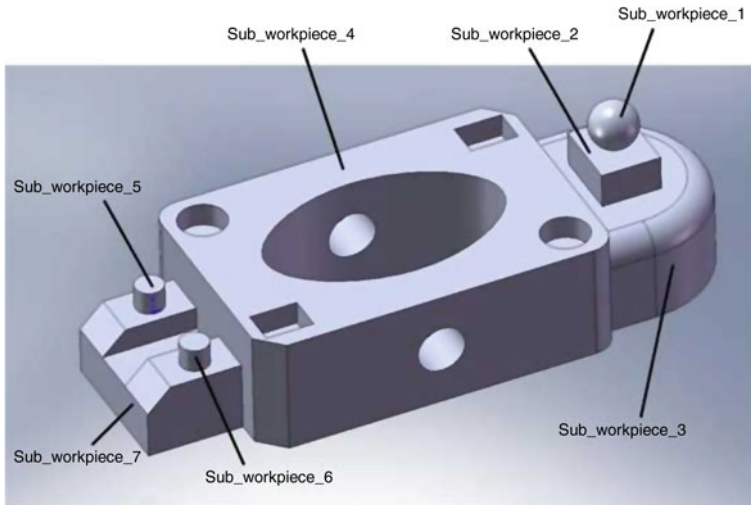


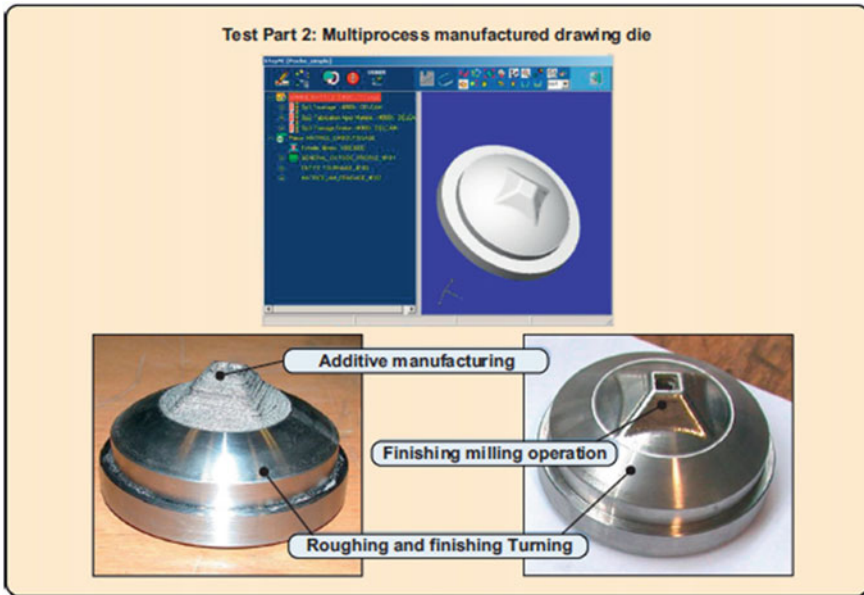
Fig. 9.23 Example part for additive manufacturing

manufacturing hybridization by combining multiple processes (Fig. 9.23). In the figure, multiple features that would have been milled, for example, are redefined as AM sub-workpieces to be created using a multi-process workplan. The main advantage is therefore overall production simplicity. Furthermore, some features, like conformal cooling channels for example, are more easily fabricated with AM than conventional methods, so again production capabilities can be extended. The addition of AM capitalizes from the use and reuse of previously defined manufacturing features for other processes in STEP-NC. For features not currently defined, the researchers proposed the addition of new AM related features. An illustration of this strength of AM in multi-process STEP-NC is shown below.

### Experimental Validation Study

To emphasize this idea of optimization (process and production), an experimental study was conducted to validate the efficiency of such a multi-process manufacturing approach. The test part in Fig. 9.24 is used. It is a drawing die for which functional requirements calls for hardened steel at the top of the die.

Although this part is geometrically symmetric and would be more efficient in a turning process, it would need to be machined by milling to accommodate for the pyramidal shape at the top. Further processing such as thermal treatment would be required to fulfill the functional requirement. With the multi-process capability of STEP-NC within the SPAIM application, a different processing scheme presents itself and the results are more efficient in terms of meeting production requirements as well as utilizing each process efficiently. In this new scheme, three manufacturing processes were used: additive manufacturing (CLAD), turning and milling.



**Fig. 9.24** Drawing die in multi-process manufacturing

A STEP-NC enabled CNC controller possesses workplans (micro process plan) made of multi-process workingsteps. In this environment, the effectiveness of a multi-process manufacturing application like SPAIM is evident (Fig. 9.24). This scheme presents several situations that multi-process optimization (performance, execution and general) is a natural conclusion and opens up new avenues for production that were once considered too complex or even impossible.

**Performance optimization:** tailoring production for performance and functional requirements such as strength, surface and hardness characteristics though multi-material parts i.e. adding hardened steel via AM to a different substrate metal.

**Execution optimization:** creating new possibilities for part fabrication, this optimization allows production execution sequences to be flexible and gives macro and micro process planning more agility. One benefit is being able to decouple manufactured features from any specific process e.g. creating bosses from AM instead of by milling which would have resulted in excessive material waste. This optimization opens a whole new direction in the way parts are produced, especially in the advent of novel fabrication methods like AM.

**General optimization:** reducing manufacturing difficulty and complexity by splitting or distributing production to other specialized processes. An example is combining turning and AM instead of straight milling as was done with the drawing die test part.

The manufacture of this drawing die, for example, satisfies all three optimizations at once: simplicity, tailoring and flexibility. With a conventional

manufacture of this part, there is little cohesion between the part and the processes used to fabricate it. This means that if multiple processes were used to manufacture this part without STEP-NC, each associated process would have its individual process execution plan and they would be totally independent of each other. This lack of cohesion makes it very difficult to share information and knowledge between the experts and the processes. However, in STEP-NC, there are very cohesive links which are magnified and allow more holistic and optimized part production. There is a single process plan even for the multi-process manufacture of this part. Features that are not involved in a particular process can be ‘frozen’ or hidden from view and ‘unfrozen’ when needed. Knowledge is shared between processes in this case which leads to improved optimization execution. Although the added extra process could be considered an increase in complexity or production time, the benefits of being able to create a unique multi-material component for targeted applications outweighs those concerns depending on the sector.

Briefly explained in [Sect. 9.3.1](#), multi-process optimization was also applied to the manufacture of the simple part with pockets shown in [Fig. 9.11](#). In this case, the relative manufacturing difficulty is reduced by extending production to multiple processes, each with their own inherent advantages: straight pocket milling with simpler toolpath for area clearance and AM to add the bosses using similar or different materials compared to the base substrate material.

Embracing STEP-NC in a multi-process environment has significant advantages especially in the new possibilities it creates for part fabrication. Sharing information between processes leads to performance, execution and overall optimizations that are difficult or impossible to achieve by conventional methods.

### Simulation

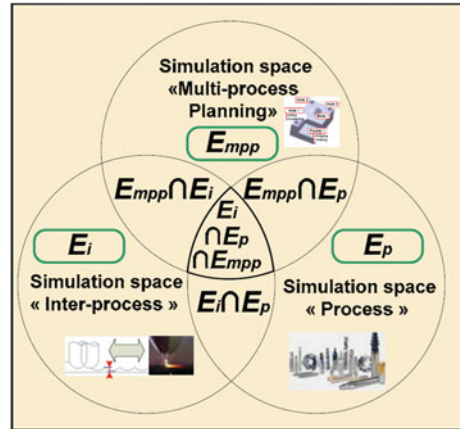
In order to progress the advance possibilities of STEP-NC use, Laguionie et al. defined a proposal on how simulation could be approached in a multi-process environment [22]. Three main simulation spaces ([Fig. 9.25](#)) are identified and defined as well as their interaction zones. A simulation space is defined as a boundless but structured set where simulations can be performed. The figure shows schematized imaginary boundaries representing knowledge limits however the boundaries can be infinitely extended. These simulation spaces can be distinguished by their respective objectives. They are complementary, interdependent and closely linked.

The main input of the manufacturing numerical chain is the CAD model of the part. This CAD model contains data for part geometry, tolerances, material, etc. These characteristics must be respected and are a theoretical objective for manufacturing process. However, every process has its own constraints that must be taken into account when deciding the manufacturing workplan of the part.

### Main Simulation Space

- (1) From CAD to manufacturing features and process plan: the Multi-process planning simulation space (Empp)

**Fig. 9.25** Simulation in a multi-process environment



Discretizing a CAD model into manufacturing features means more than recognizing usual geometrical entities. A lot of research works have been done on manufacturing feature extraction [32]. It appears that the solution can hardly be separated from the manufacturing process involved. In other words, the choice of manufacturing feature geometry is linked to a process. Thus, a simple part can lead to several feature decompositions depending on the process choice in a multi-process environment, as the process is not fixed. In some cases, indecision of a suitable process to machine a particular feature can lead to further simulations. This is the exploration field for the  $E_{mpp}$  which objective is to find the optimal machining features and process plan.

### (2) Process simulation space ( $E_p$ )

The workplan is composed of several workingsteps associated with the manufacturing features. The process simulation space ( $E_p$ ) gathers all the computations and choices concerning process data and machining parameters selection.  $E_p$  is involved for high level manufacturing data selection in STEP-NC but also integrates shopfloor simulation in the interpreter. This shopfloor simulation can be performed offline or online. Typically, an offline simulation will be privileged in the STEP-NC interpreter for tool paths programming, tool paths optimizations, machining parameters adaptation, etc. Real time optimizations, simulations from sensors feedback in the CNC could also be done online and are part of  $E_p$ .

### (3) Inter-process simulation space ( $E_i$ )

Relations between processes take a central position in a totally integrated multi-process context. The Inter-process relationship and optimizations are fully integrated in the numerical chain and are a large consideration of the  $E_i$  space, which is new here. In the traditional G-code based numerical chain, a large part of the simulations concern the process ( $E_p$ ), manufacturing experts choose the well adapted workplan based on their experience and knowledge ( $E_{mpp}$ ), but inter-process

simulation ( $Ei$ ) is hardly possible due to different data standards and expert communication languages. In the STEP-NC numerical chain,  $Ei$  is the missing link between  $Empp$  and  $Ep$  when several processes are required. Relationship and communication channels between the different processes are enabled with a direct effect on the machining features, the workplan and the process data.

### Interactions and Optimized Solutions

#### (1) Process and process planning interaction

The intersection between  $Empp$  and  $Ep$  spaces is the place for process simulations in close link with process planning, and reciprocally.  $Empp \cap Ep$  includes the bidirectional relationship between CAPP and each selected process. The relationship between machining entities and process parameters are important. For example, in pocket milling, the tool diameter is limited by the corner radius for finishing operations. Similarly, characteristics of a process can also lead to the selection of a feature over another. For instance, using a step drill to drill a hole and the associated counter bore results in forcing the two features to be merged in one workingstep.

#### (2) Process planning and inter-process relations interaction

The constraints of each process must be taken into account when creating machining features and workplan. These interactions are simulated in  $Ep \cap Ei$ . For example, the thermal effects of Direct Laser Manufacturing (DLM) with powder injection [33] constrain the selected features to be manufactured very early in the workplan, before the finishing operations of high speed milling. Conversely, machining features sometimes constrain the choice of the process sequence and consequently of the inter-process relations. For example, if a flat section must be milled after a turning operation, inter-process interactions are ordered by process planning.

#### (3) Process and inter-process interaction

The process simulation space  $Ep$ , directly linked with the machine-tool functional model, can provide the simulation results associated with a selected process. The results can have consequences on the other processes machining parameters. For example, the finishing cutting conditions in a milling process (feedrate, cutting speed, etc.) can be adapted due to thermal effects after DLM machining a feature. A multidirectional data exchange enables a workingstep associated with a process to be performed in an intelligent multi-process context.

#### (4) A global, optimized solution in the comprehensive $Empp \cap Ep \cap Ei$ space

The goal to reach is the simulation of a totally integrated manufacturing environment that considers at the same time, relevant multi-process planning, inter-process relationships and process attributes simulations. However, simulation in  $Empp \cap Ep \cap Ei$  involves complex reasoning and simulation methods. Processing the three presented simulation spaces at the same time is certainly not the best way to initiate a solution. Future technological advances could allow this.

The main emphasis here is that with the different simulation spaces identified, this provides a roadmap on which the different experts along the numerical chain can rely to help them make choices, again not only by considering their specific area but by considering and integrating the requirements of the whole manufacturing numerical chain. It also allows STEP-NC developers to have complex situations that they must create solutions for. This is a challenge and one that should be developed further by exploring the limits of the STEP-NC data exchange standard supporting multi-process planning and process data.

## 9.5 Discussion

One aspect of STEP-NC that is often overlooked is how to enter into the STEP-NC paradigm in the first place. Usually the entry point is a CAD part file supported with manufacturing and resource information. Starting from this CAD part file, the conversion to STEP-NC is currently a semi-automatic activity. Due to the feature-based design of STEP-NC, there is an inherent need to be able to recognize the features described. With the help of manufacturing feature recognition, based on ISO 10303-AP224, some software systems have been developed to extract features using CAD input files (e.g. STEP or parasolids files) [34]. The process is currently semi-automatic because some features like holes are regular enough that they could be extracted automatically. However for more complex features (even some types of pockets or slots), a higher level of human interaction is required to identify those features. The variability of opinion about features within manufacturing and the ambiguous nature of some features make it very difficult to have complete automation. At present, no system exists that is capable of autonomous feature recognition of manufacturing features. Part of the problem has to do with the fact that not all manufacturing features are standardized and not all of the intent behind the design of such features is transmittable without human interaction/input. Even when the design intent could be transmitted, that does not always provide useful information for manufacturing.

For simple features like pockets and holes, SPAIM as a STEP-NC enabled application uses its bidirectional link with CAD to be able to create a part for manufacture directly in STEP-NC format on the HMI. This method is similar to the wizard-like capability of some industrial CNC machines that allow 'Conversational or Shopfloor Programming' via software embedded on the controller's computer. The extent of this method is currently limited but represents another starting point to enter the STEP-NC world.

It should be emphasized that the adoption of STEP-NC does not mean that the current seemingly isolated departments (design, planning etc.) of an enterprise will remain isolated or become obsolete. The exact opposite is true. Adopting STEP-NC necessitates and facilitates the integration of these departments. The full flexibility and optimization capabilities offered by STEP-NC are in fact realizable with such integration. The bidirectional data flow between systems in the STEP-NC

numerical chain is not limited only to the systems. The experts that are behind those systems for designing, planning, programming and operating etc. are equally essential. They make up a substantial knowledge source that process and multi-process optimization relies on.

Therefore, the need for human knowledge and experiences in the STEP-NC numerical chain should not be diminished or overlooked. Although many aspects can be automated such as rules-based tools for simulation, cutting tool and machining parameters selection, the need for human experience to optimize process planning for example is essential. Decision making about what strategies are available to manufacture a particular feature, the type of supporting manufacturing resources available are all essential aspects for which a human STEP-NC programmer is currently required.

To develop such a capacity, the programmer(s) is required to have substantial knowledge about manufacturing systems not only limited to CAM or CNC. Current CAD/CAM/CNC programmers and operators collectively already possess the requisite knowledge base. The challenge is to teach them how to think in the non-linear manner that is required by STEP-NC and its applications. This step is not unlike any training to learn and use new machines or software systems.

The leap forward is not as substantial as is often expected and one of the opportunities for the promoters of STEP-NC (the Authors included) is to be able to effectively communicate this both in print and with tangible demonstrations.

It should not be forgotten that STEP-NC and the framework, described in this chapter, are adaptable. Depending on the needs and capabilities of an enterprise, the extent of STEP-NC use and control can be scaled. In a serial production environment for example, the need for shopfloor modifications during production execution is extremely limited and could be completely removed as a capability. In this case, experts would not be required directly on the shopfloor to manipulate STEP-NC but they are expected at higher levels such as planning and validation. Considering this, the cohesive link that STEP-NC enables for systems in the numerical chain (design, planning and eventual optimization of the manufacturing process) can be performed with the efficient use of enterprise resources.

In a non-serial production environment, the needs would again change depending on the capabilities of the enterprise. The conclusion is however still the same: only the aspects of STEP-NC that provide direct benefits and added-value to an enterprise would be used. The very large range of manufacturing data within STEP-NC can be totally tailored to suite different environments.

## 9.6 Future Focus

For the future, several areas including the advanced possibilities with STEP-NC discussed in this chapter, will require continued development and implementation. Some areas of focus should include:

- improving the capacity of current systems by making them more comprehensive, automated and autonomous.
- implementing and validating the advanced possibilities discussed, particularly the interoperability aspects.
- defining and creating studies that highlight the economic impact as well as quantitative advantages of STEP-NC and STEP-NC systems.
- properly defining industrial needs and qualitative objectives for manufacturing programming.
- consideration of the human element in STEP-NC particularly as it relates to enabling modifications and optimizations as well as improving decision-making and validation at every step of a project.

For all of this to happen, the promoters of STEP-NC will be required to effectively and convincingly communicate positive results to the industry by providing more tangible industry-related demonstrations. Only with sufficient and diverse support from both industrial and academic will this future of advance manufacturing with STEP-NC be realized.

## 9.7 Adopting STEP-NC (Q&A)

One very important question that is often asked is “What is needed in terms of software and hardware for an enterprise to migrate a machine to STEP-NC?”

As a first prototype used to demonstrate the STEP-NC paradigm, the development of SPAIM required a few constraints. The most important of these is software, which coincidentally are currently used in the conventional CAD-CAM chain.

For the moment, if an enterprise wants to adopt STEP-NC based on the SPAIM application, the following items will be needed:

- CAD software: Delcam PowerShape
- CAM software: Delcam PowerMill (with post processor: Delcam DuctPost)
- and of course, STEP-NC programming and training support from IRCCyN.

Initially SPAIM was implemented only on a Siemens 840D controller, however with continued development, it has been implemented on other controllers.

The benefits highlighted in this chapter are currently available through SPAIM on the following legacy CNC controllers: Siemens 840D, Num 750, Fidia nC-12, and Heidenhain. They are also available on the Open-NC LinuxCNC-based open architecture controller. For all the legacy controllers mentioned, no hardware modification is required which makes implementation a quick process. This is a positive point for easing manufacturers concerns about implementation time.



### Question and Answer

Q: Could other CAD/CAM systems be used with SPAIM?

A: Currently, no since SPAIM was initially developed using Delcam products. It is by no means limited to those products however. To be able to use other CAD/CAM software systems with SPAIM, more development effort would be required to make the adaptation and automation links.

Q: Could any machine's CNC be made STEP-NC compatible?

A: Yes. Currently available on 4 major legacy industrial CNCs.

Q: Could an old machine-tool be hardware-upgraded or retrofitted and be made STEP-NC compliant?

A: Yes. Having performed an upgrade of the 3-axis Sabre machine-tool by adding LinuxCNC and Fidia controllers for example, the know-how, components and efforts necessary for this adaptation are known and available to be used for future implementations. The addition of SPAIM will allow it to be STEP-NC enabled.

## 9.8 Conclusions

The benefits of STEP-NC and the applications that support it have been discussed. There are many opportunities for the manufacturing industry to adopt the paradigm shift however continued demonstrations and industry targeted use cases will still need to continue to facilitate a step by step migration to the cohesive possibilities offered by STEP-NC.

With more implementation and validation of complex, integrated manufacturing environments with STEP-NC, the manufacturing industry will be capable of realizing the following benefits: interoperability through hybrid manufacturing environments, improved manufacturing supervision and traceability, flexibility and efficiency with high knowledge and information transfer as well as production optimization and simulation in multi-process manufacturing. All of these benefits will increase systems interoperability and portability, production efficiency and above all flexibility and agility within an enterprise.

Currently, STEP-NC and its implemented applications are advanced enough to allow current industrial CNCs to quickly be made STEP-NC compatible with the addition of the SPAIM application for example. Immediately, they will be capable of Indirect or Interpreted STEP-NC programming and benefit with flexibility and agility. Manufacturers will be able to make modifications to toolpath and strategies, part geometry, tools selection, process plans, optimize individual processes and realize new capabilities with multi-process, all with direct links to the shop-floor. This is undeniably a beneficial capability as manufacturers struggle to meet the demands of a quickly changing competitive industry. STEP-NC at its current state will allow them to be as adaptive and agile as possible as they plan, optimize

and react to the conditions on the shopfloor. Additionally, the future potential of STEP-NC enabled manufacturing is promising and more than capable of addressing the eventual needs that will arise.

## References

1. ISO\_6983-1(1982) Numerical control of machines—Program format and definition of address words—Part 1: data format for positioning, line motion and contouring control systems. In: International Standard Organization
2. Pratt MJ (2001) Introduction to ISO 10303—the STEP standard for product data exchange. *J Comput Inf Sci Eng* 1:102–103
3. Newman ST, Nassehi A, Xu XW, Rosso RSU Jr, Wang L, Yusof Y et al (2008) Strategic advantages of interoperability for global manufacturing using CNC technology. *Robot Comput Integr Manuf* 24:699–708
4. ISO\_14649-10 (2004) ISO 14649 part 10—international standard organization. Industrial automation systems and integration—physical device control—data model for computerized numerical controllers—part 10, general process data. ISO TC 184/SC1/WG7/FDIS
5. Weck M, Wolf J, Kiritsis D (2001) The STEP compliant NC programming interface evaluation and improvements on the modern interface. In: IMS project forum. Verita, Ascona
6. STEP Tools, ST-Machine STEP-NC for CAM-CNC
7. Newman ST, Allen RD, Rosso JRSU (2003) CAD/CAM solutions for STEP-compliant CNC manufacture. *Int J Comput Integr Manuf* 16:590
8. Wang H, Xu X, Tedford JD (2007) An adaptable CNC system based on STEP-NC and function blocks. *Int J Prod Res* 45:3809–3829
9. Minhath M, Vyatkin V, Xu X, Wong S, Al-Bayaa Z (2009) A novel open CNC architecture based on STEP-NC data model and IEC 61499 function blocks. *Robot Comput Integr Manuf* 25:560–569
10. Suh SH, Lee BE, Chung DH, Cheon SU (2003) Architecture and implementation of a shop-floor programming system for STEP-compliant CNC. *C-A Design* 35:1069–1083
11. Xu XW (2006) Realization of STEP-NC enabled machining. *Robot Comput Integr Manuf* 22:144–153
12. Shin S-J, Suh S-H, Stroud I (2007) Reincarnation of G-code based part programs into STEP-NC for turning applications. *Comput Aided Des* 39:1–16
13. Xu L (2009) STEP-NC in support of machining process optimization. In: Xu X, Nee AYC (eds) *Advanced design and manufacturing based on STEP*. Springer, London, pp 169–196
14. Ridwan F, Xu X, Liu G (2012) A framework for machining optimization based on STEP-NC. *J Intell Manuf* 23(3):423–441
15. Kumar S, Nassehi A, Newman ST, Allen RD, Tiwari MK (2007) Process control in CNC manufacturing for discrete components: a STEP-NC compliant framework. *Robot Comput Integr Manuf* 23(6):667–676
16. Martin YS, Gimenez M, Rauch M, Hascoet JY (2006) VERNE—a new 5-axes hybrid architecture machining centre. 5th Chemnitzer Parallelkinematik Seminar. Chemnitz (Germany), April 25–26, pp 657–676
17. Delcam. PowerSHAPE—CAD Design and Modelling Software. <http://www.powershape.com>. Accessed 2008
18. Delcam. PowerMILL—Your Total Manufacturing Solution. [www.powermill.com](http://www.powermill.com). Accessed 2008
19. Dugas A, Lee JJ, Terrier M, Hascoet JY (2003) Development of a machining simulator considering machine behaviour. *Proc Inst Mech Eng B J Eng Manuf* 217(9):1333–1339

20. Seo TI, Dépincé P, Hascoet JY (1997) Paths compensation for tool deflection in end milling. *Ifac-Ims Intelligent Manufacturing Systems*. Seoul (Korea), July 21–23
21. Sokolov A, Richard J, Nguyen VK, Stroud I, Maeder W, Xirouchakis P (2006) Algorithms and an extended STEP-NC-compliant data model for wire electro discharge machining based on 3D representations. *Int J Comput Integr Manuf* 19(6):603–613
22. Laguionie R, Rauch M, Hascoet J-Y (2009) Simulation and optimization in a multi-process environment using STEP-NC. *IEEE ICCA 2009*:2384–2391
23. Newman ST, Nassehi A (2007) Universal manufacturing platform for CNC machining. *CIRP Ann Manuf Technol* 56(1):459–462
24. Suh SH, Shin SJ, Yoon JS, Um JM (2008) UbiDM: a new paradigm for product design and manufacturing via ubiquitous computing technology. *Int J Comput Integr Manuf* 21(5):540–549
25. Laguionie R, Rauch M, Hascoët JY, Suh SH (2011) An eXtended manufacturing integrated system for feature-based manufacturing with STEP-NC. *Int J Comput Integr Manuf* 24(9):785–799
26. Rauch M, Laguionie R, Hascoët JY, Xu X (2009) Enhancing CNC manufacturing interoperability with STEP-NC. *J Mach Eng* 9(4):26–37
27. IEC 61499 (2005) Function blocks for industrial-process measurement and control systems—part 1: architecture
28. Hentz JB, Nguyen VK, Mader W, Panarese D, Gunnink JW, Gontarz A, Stavropoulos P, Hamilton K, Hascoet JY (2012) An enabling digital foundation towards smart machining. *CIRP ICME 2012*
29. ISO14649-13 (2003) ISO14649 part13—industrial automation systems and integration—physical device control-data model for computerized numerical controllers—part 13, process data for wire-EDM. *TC 184/SC 1*
30. Sokolov A, Richard J, Nguyen VK, Stroud I, Maeder W, Xirouchakis P (2006) Algorithms and an extended STEP-NC-compliant data model for wire electro discharge machining based on 3D representations. *Int J Comput Integr Manuf* 19:603–613
31. Bonnard R, Mognol P, Hascoet JY (2008) Rapid prototyping project description in STEP-NC model. In: *Proceedings of the sixth CIRP international seminar on intelligent computation in manufacturing engineering*, Naples, Italy
32. Babic B, Nestic N, Miljkovic Z (2008) A review of automated feature recognition with rule-based pattern recognition. *Comput Ind* 59:321–337
33. Barnes S, Timms N, Bryden B, Pashby I (2003) High power diode laser cladding. *J Mater process Technol* 138:411–416
34. Xu X et al (2006) STEP-compliant process planning and manufacturing. *Int J Comput Integr Manuf* 19:491–494

# Chapter 10

## Optimum CNC Free-form Surface Machining Through Design of Experiments in CAM Software

N. A. Fountas, N. M. Vaxevanidis, C. I. Stergiou  
and R. Benhadj-Djilali

**Abstract** Sculptured surface machining (SSM) is an operation widely applied to several industrial fields such as aerospace, automotive and mold/die. The number of the parameters and strategies involved to program such machining operations can be enormously large owing to surface complexity and advanced design features. To properly reduce the number of parameters, design of experiments (DOE) methodology along with statistical analysis can be adopted. In this paper DOE and respective analysis were used to conduct machining experiments with the use of a computer aided manufacturing (CAM) software. Major goal is to investigate which of process parameters are worthy of optimization through intelligent systems. Two scenarios were considered to machine a sculptured part; one involving 3-axis roughing/3-axis finish machining experiments and one involving 3-axis roughing/5-axis finish machining experiments. Roughing operation was common for both scenarios. The problem was subjected to discrete technological constraints to reflect the actual industrial status. For each machining phase, two quality objectives reflecting productivity and part quality were determined. Roughing experiments were conducted to minimize machining time ( $t_{mr}$ ) and remaining volume ( $v_r$ ); whilst finishing experiments were targeted to minimize machining time ( $t_{mf}$ ) and surface deviation ( $s_{dev}$ ) between the designed and the machined 3D model. Quality characteristics were properly weighted to formulate a single objective criterion for both machining phases. Results indicated that DOE applied to CAM software, enables numerical control (NC) programmers to have a clear understanding about

---

N. A. Fountas · N. M. Vaxevanidis (✉)  
Department of Mechanical Engineering Educators, School of Pedagogical and Technological Education (ASPETE), N. Heraklion, GR 14121, Athens, Greece  
e-mail: vaxeve@ath.forthnet.grvaxeve@aspete.gr

C. I. Stergiou  
Department of Mechanical Engineering, Technological Educational Institute (TEI) of Piraeus, 250 Thivon Str, GR 12244, Egaleo, Greece

N. A. Fountas · R. Benhadj-Djilali  
Faculty of Science, Engineering and Computing, Kingston University, Penrhyn Road, Kingston Upon Thames KT1 2EE, UK

the influence of process parameters for sculptured surface machining operations generating thus efficient tool paths to improve productivity, part quality and process efficiency. Practically the work contributes to machining improvement through the proposition of machining experimentation methods using safe and useful platforms such as CAM systems; the application of techniques to avoid problem oversimplification mainly when large number of machining parameters should be exploited and the evaluation of quality criteria which allow their assessment directly from CAM software.

## **10.1 Introduction**

### ***10.1.1 Sculptured Surface Machining***

Many of today's modern products are designed with free form surfaces either to attract consumers or to meet special functional requirements like those identified to mould/die, aircraft and automotive industries. Sculptured parts are machined on 3-or 5-axis CNC machine tools with the aid of a CAM systems to model machining operations need to be performed. The final product of CAM systems is the generated NC code translated from CL (cutter location) data. The NC code drives machine tools to perform cutting operations so as to efficiently remove the unnecessary material from the rough stock producing the final product. Thus; the main problem of Sculptured surface machining [1] is to:

- Produce the optimum sequence of machining operations to machine a sculptured part.
- Produce the best sequence of NC blocks for each machining operation; and
- Select the optimum cutting parameters and conditions for each NC block.

### ***10.1.2 Aims and Objectives***

The research presented in this chapter is mainly concentrated to the application of DOE methodology for determining machining parameters affecting both roughing and finishing phases, to formulate the best machining sequence for optimized sculptured surface machining. The ultimate goal is to use most important CAM parameters identified from experiments in order to subject them to a loop of evaluations which will be conducted by an intelligent optimization platform until arriving at optimal solutions. Another goal is to introduce such an experimental approach in practice, i.e. in machine shops and industries to facilitate machining modelling by accelerating earlier planning stages for preparing and evaluating machining operations. Such a methodology would require proper process planning and modelling analysis of its studied tasks, as well as suitable determination of

quality characteristics for their control and optimization, regarding the process results. Since the study concerns both roughing and finishing stages three optimization targets are assessed; the material's volume after roughing, the machining time and the surface deviation which is introduced as a novel parameter to reflect surface quality in finishing. The overall study may substantially contribute to the scope of interest on optimal Sculptured Surface Machining technology, through:

- The proposition, the development and implementation of a new optimization strategy based on evolutionary techniques as practically viable;
- The efficient and robust processing of large numbers of optimization parameters;
- The proper treatment of multi objective optimization concept along with its alternative regions of optimal solutions;
- The interaction of software modules with practical tools and computer aided systems used in industrial practices such as CAM software.

## 10.2 Literature Review

It is a common industrial task to properly define major machining parameters and their value ranges. The optimum definition of machining parameters has questioned a large number of researchers worldwide. Specifying optimum machining parameters for a manufacturing process is now imperative to areas where production cost is of paramount importance. Indicative machining parameters for roughing and finishing that usually treated as independent variables for machining optimization; are:

- Cutting tool geometry [2];
- Spindle speed of the CNC machine tool [3];
- Feed rate [4];
- Cutting speed [5].

For sufficient exploration of optimum machining parameters, quality objectives should be determined to reflect problem's responses. Regarding the machining phase, quality objectives may vary considering the outcomes of both the machined part accuracy and machining efficiency as well. Research works presented so far specify one (single objective optimization) or more quality objectives (multi objective optimization). Such quality objectives are the ones bulleted below:

- Minimum machining time (including part, or tool setup time, cutting time, rapid traverse time, time needed for tool changes, etc. [6];
- Maximum material removal rate (MRR) [7];
- Minimum tool deflection [8, 9];
- Surface roughness [6, 10];
- Minimum cutting force components [11];
- Maximum tool life [12].

Resources needed to machine parts by material removal, impose the determination of technological constraints in order to ensure feasibility of manufacturing processes. Several constraints are set with variations regarding the scope of optimization. Such constraints are the following:

- Maximum available spindle power, maximum cutting force, maximum allowable cutting load and maximum allowable tool deviation [13].
- Maximum spindle torque and maximum tooling feeds [14].
- Ranges of cutting parameters [15, 16].
- Cutting tool geometry [16].
- Tool wear and yielded heat [17, 18].
- Geometrical and/or technological configurations of machine tools, stability of CNC machine tools, etc., [19].

Roughing and finishing constitute the most important machining phases for milling operations. During the roughing process a high material removal rate should be achieved. The outcome of rouging operation is a part geometry closed to the ideally designed one. Finishing operation aims at removing the remained material resulting thus; to the final product which should meet predetermined requirements. Roughing is characterized by its ability to increase productivity, whereas high surface quality and dimensional/geometrical accuracy should be the main characteristics of finishing. As a machining sequence; both phases play key roles to a successive manufacturing process.

To optimize roughing and finishing operations, algorithms for geometrical calculations have been utilized. Such algorithms can be found in different optimization methods like “contour map” [20], “Convex Hull Boxes” [21] and geometrical feature recognition methods [22] for roughing and “Z-Map” [1], 3D point cloud data [23] and surface subdivision methods [3] for finishing.

Tool paths for roughing and finishing have also been investigated to facilitate machining optimization studies. The most often used tool paths are summarized in [24]. The work in [25] proposed a method for achieving the largest possible cutting width in terms of optimum machining direction in each surface point, whilst the work in [26] proposed a 5-axis tool path generation approach which leads to a constant “scallop” height, between parallel planes. Yang et al. [23] developed a NURBS fitting algorithm to generate tool paths with fewer control points.

Many optimization algorithms appeared in the literature utilize mathematical equations as fitness functions. Chang et al. [27] adopt process models for single and multi pass machining operations. As a “fitness function” the processing time “ $t_{pr}$ ” to machine a part is considered. “ $t_{pr}$ ” in their work, is calculated by Eq. 10.1:

$$t_{pr} = t_m + t_h + t_t \left( \frac{t_m}{t} \right) \quad (10.1)$$

where,

- $t_{pr}$  the total machining time;
- $t_m$  the material handling time;
- $t_h$  the idle times for part clamping, orienting, fixturing and removal;
- $t_t$  the tool change time (idle time);
- $t$  the tool life, calculated by ‘Taylor’ equations and “ $(t_m/t)^{-1}$ ” the number of parts that can be produced before the tool requires changing.

Equation 10.1 is adopted by many researchers. However, its major drawback is that provides approximated results, at least in the case of cutting tool life estimation. The rest components of processing time can be accurately calculated except of the term ‘ $t_h$ ’ which may vary regarding the part’s size and its probably special fixturing alternatives on the machine table.

According to most research works, spindle speed, feed rate and cutting depth are the investigated parameters, regardless of the part geometry; whilst quality objectives are treated separately. Hence, one should expect to arrive at “partial” or case oriented solutions. As machining parameter ranges for values are wide, their combinations are theoretically infinite. To find the “elite” of parameter combinations, several techniques are implemented to indicate only parameter combinations that satisfy quality objectives. Such techniques are:

- Taguchi’s design of experiments [28].
- Artificial neural networks, which operate as predictive functions after suitable training [29].
- Graphical techniques [30].

Machining optimization problems are perplexing due to the non linear dependence of the parameters involved [31]. A wide range of methodologies have been developed so far. An early and popular approach is to find the roots of the first derivative for a cost function if such can be obtained, [32, 33]. The cost function includes one or more quality targets. The roots correspond to local extrema of a function, while the second derivative checks whether the local extreme is of a minimum or maximum. To utilize this approach a function’s second derivative must exist whilst its first derivative should be continuous. This is rare since such optimization problems are complex. To address such difficulties, researchers have used different techniques that enable the calculus when it comes to non continuous fitness functions. Some of the major strategies have been proposed so far, are integer programming, [34]; dynamic programming, [35] and geometric programming, [36]; artificial intelligence (AI) approaches such as genetic evolutionary algorithms, GAs EAs [36]; simulated annealing [37] and particle swarm optimization [38]; hybrid optimization systems such as GA SA [39]; optimization using “Tribes”, [40]; heuristics and GAs, [41]; genetic algorithms coupled with artificial neural networks (GAs ANNs), [42]; GA quadratic programming [43]; simulated annealing and ANNs [44], etc.



Most of the research work been proposed so far aims at proposing mathematical relations among studied parameters in order to express deterministic objectives. More over, the process parameters investigated for optimization are few. Specifically, the main cutting conditions (cutting speed, feed rate and depth of cut) are usually processed; whilst others which are just as important are neglected. Such approaches tend to downgrade sophisticated optimization strategies and introduce them as impractical or inefficient despite the fact that some of their elements could facilitate machining processes. Nevertheless, mach of the strategies presented for machining optimization lead to partial problem solving and raise questions about the very essence of their solution quality. Therefore, another approach able to allow the problem's investigation in its whole may be introduced for sculptured surface machining optimization.

### **10.3 Introduction of CAM Systems to Sculptured Surface Machining Optimization**

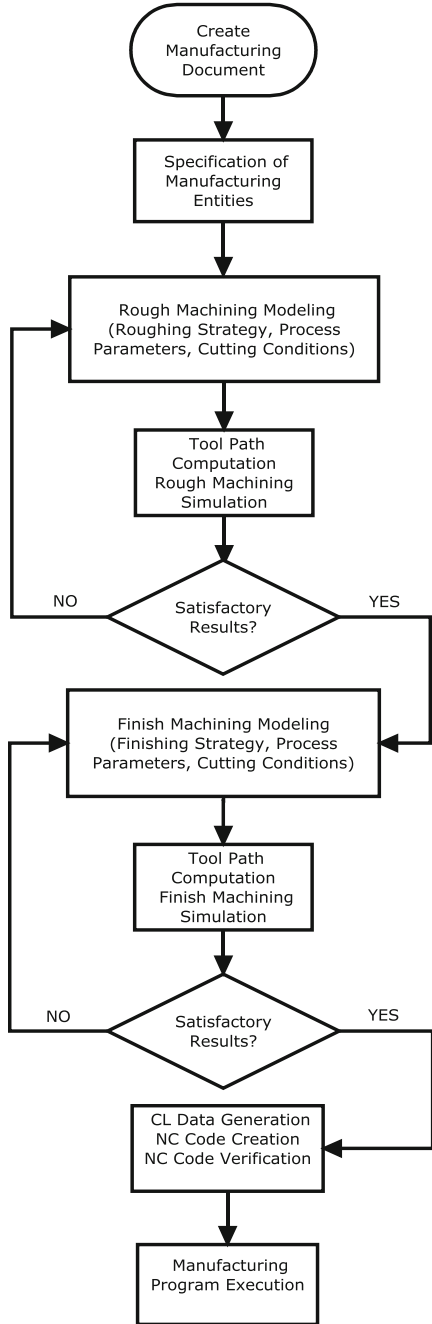
In the case of sculptured surface machining, a vast number of process parameters appear due to the complexity of parts' surfaces. CAM software provides the safest way of creating several machining scenarios and assessing generated results; while it constitutes also a very useful tool when minimization of production cost is of major importance. CAM software provides all the necessary entities to model machining operations and build a manufacturing program to be executed by a CNC machine tool avoiding thus problem oversimplification. The basic process planning flowchart when using CAM software is presented in Fig. 10.1.

CAM systems aim at facilitating all kinds of industrial tasks and they offer a large suite of tools/options yielding thus the risk of the direct impediment to their use. The number of parameters of sculptured models may be large enough and vary depending on the complexity of their surfaces. Therefore, minimization of process parameters is imperative and should be conducted through statistical techniques. To achieve successful parameter elimination, it is important to determine objectives associated to productivity and quality. Design of experiments and analysis of variance (ANOVA) are low cost tools for identifying the most influential parameters on a machining problem's response and finding optimum or near optimum values of sculptured surface machining strategy parameters.

### **10.4 Design of Experiments for Sculptured Surface Machining in CAM Software**

Taguchi's DOE methodology was implemented to design machining experiments in CAM software for a designed sculptured geometry. The major goal was to identify the optimum machining program in terms of process productivity and

**Fig. 10.1** Traditional approach of machining modeling operations in CAM software



resulted part quality. For this scope, three sets of experimental programming scenarios were established. The first set involves roughing experiments; assessed by an equally weighted combination among two quality objectives specified; remaining volume,  $v_r$  ( $\text{mm}^3$ ) and roughing time,  $t_{mr}$  (min). The second machining set involves finishing experiments based on the 3-axis milling technology and subjected again to an equally weighted combination among quality objectives namely surface deviation,  $s_{dev}$  (mm) and finishing time,  $t_{mf}$  (min). The third set of machining experiments involves finishing operations based on the 5-axis milling technology and subjected to an equally weighted combination among the same quality objectives as determined for the second set. Finally optimum programming sequences for the two scenarios are formulated by coupling each of the best finishing operations (3- and 5-axis) to the best rouging scenario. Surface deviation is introduced in the study as a novel quality objective in terms of surface quality. In order for the problem to be practically viable and reflect actual industrial state, technological constraints were also taken into account. The following sections present and explain all the basics involved to the design of machining experiments with the use of CAM software. Further on, the DOE methodology is described and the analysis of results is shown.

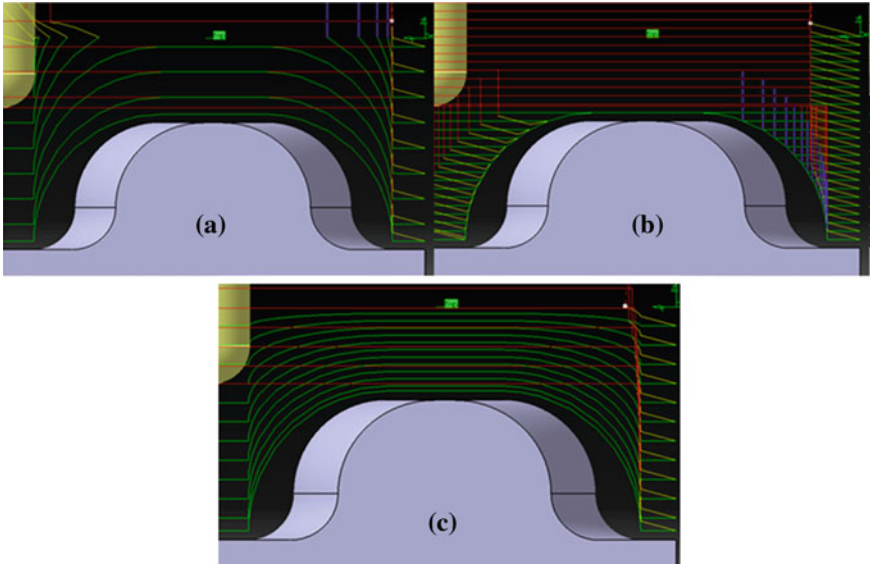
### ***10.4.1 Identification of Strategies, Parameters and Quality Objectives***

Sweeping strategies were adopted to model both roughing and finishing processes. During roughing, sweeping strategies maintain the material's uniformity so as to approach the final shape. Sweeping involves machining by vertical planes; whilst in each vertical plane additional tool paths which remove the scallops formulated by the axial levels are also incorporated. In the case of finishing operations, sweeping strategies follow the sculptured part's curvature to produce a finished surface according to dimensional specifications and predetermined tolerances.

#### **10.4.1.1 Sweeping Machining Strategy for Roughing and Related Process Parameters**

To prepare the roughing process for machining experiments three different tool path types under sweeping strategy were identified. These tool path types are presented below:

- **Z offset:** Roughing process is achieved by applying peripheral milling on a sculptured part's regions, generated by offsetting the original roughing "slice" (Fig. 10.2a).



**Fig. 10.2** Tool path styles for rough machining by sweeping strategies: **a** Z Offset; **b** Z Plane; **c** Z Progressive

- **Z plane:** The part is machined plane by plane. The planes are perpendicular to the tool axis (Z axis). The respective tool path type performs linear continuous movements in the XY plane, (Fig. 10.2b).
- **Z progressive:** The part is machined by interpolating the tool path between the part and the top of a theoretical rough stock (Fig. 10.2c).

The parameters related to sweep roughing strategy are as follows:

**Radial cutting depth (Stepover):** The distance joining two successive passes in XY plane. The parameter can be adjusted by determining the total number of paths, specifying a cutting depth ratio related to the tool's diameter ( $\% \emptyset$ ), or by defining the maximum scallop height to be left between passes.

**Axial cutting depth (Stepdown):** The distance joining two successive passes in the Z level. This parameter can be adjusted by determining the number of levels in Z direction considering the total depth of cut, or by calculating cutting passes through the tool height ratio, or by assigning specific arithmetic values for the cutting passes.

**Feeds and speeds:** Feedrate in (mm/rev) and spindle speed in (rpm). The choice of spindle speed determines the cutting speed which is equivalent to the surface speed of the cutting tool. This not only depends on the spindle speed but also on the cutter diameter (the higher the spindle speed and the larger the cutter diameter, the higher the cutting speed). Feed is the movement of the milling cutter in the machining direction.

**Cutting tool:** Different types of cutting tools are existed regarding their dimensional and geometrical configurations. The tools are selected according the part's material properties and special features. In the particular case of roughing operations, flat end mills are usually programmed to remove the material from the raw stock.

#### 10.4.1.2 Sweeping Machining Strategy for Finishing and Related Process Parameters

Sculptured parts are machined on 3-, or 5-axis machine tools. In 3-axis machine tools, ball end cutters are used for finishing by approximating the surface via small line segments (facets). In 5-axis machine tools, finishing is performed with ball end, corner radius, or flat end mills. The machining strategies and their process parameters are presented below for the 3-axis milling mode and the 5-axis milling mode for finishing:

##### 3-axis finishing

The tool path strategies selected for 3-axis finishing are as follows:

**Zig zag:** The tool path alternates directions during successive passes (Fig. 10.3a).

**One way next:** This strategy produces a tool path that always follows the same direction during successive passes and moves diagonally from the end of a pass to the beginning of the next one (Fig. 10.3b).

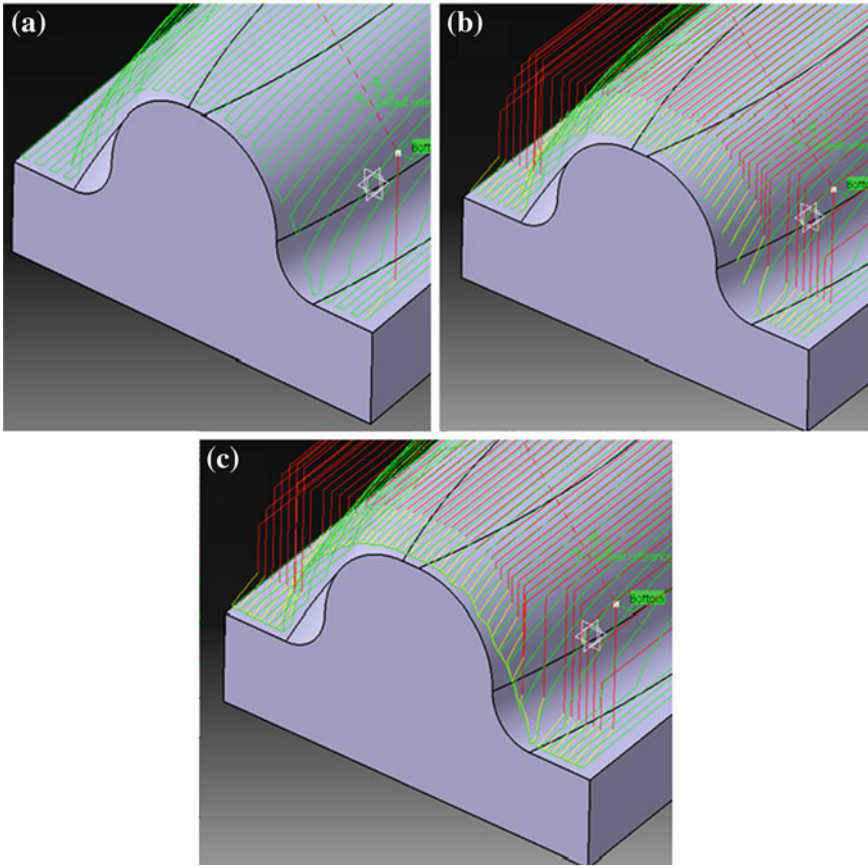
**One way same:** The tool path always has the same direction during successive passes and returns to the first point in each pass before moving on to the first point in the next pass (Fig. 10.3c)

**Radial cutting depth (Stepover):** The distance joining two successive passes in XY plane. The parameter can be adjusted by determining the total number of paths, specifying a cutting depth ratio related to the tool's diameter ( $\% \varnothing$ ), or by defining the maximum scallop height to be left between passes.

**Feeds and speeds:** Feed rate in (mm/rev) and spindle speed in (rpm). The choice of spindle speed determines the cutting speed which is equivalent to the surface speed of the cutting tool. This not only depends on the spindle speed but also on the cutter diameter (the higher the spindle speed and the larger the cutter diameter, the higher the cutting speed). Feed is the movement of the milling cutter in the machining direction.

**Cutting tool:** In 3-axis finishing, the tool axis is fixed thus; possible degrees of freedom are restricted. Since these limitations are existed, tools having spherical geometries (ball end mills) are used to finish sculptured parts due to their advantage of moving over all surface regions and machining any point around the curvature.

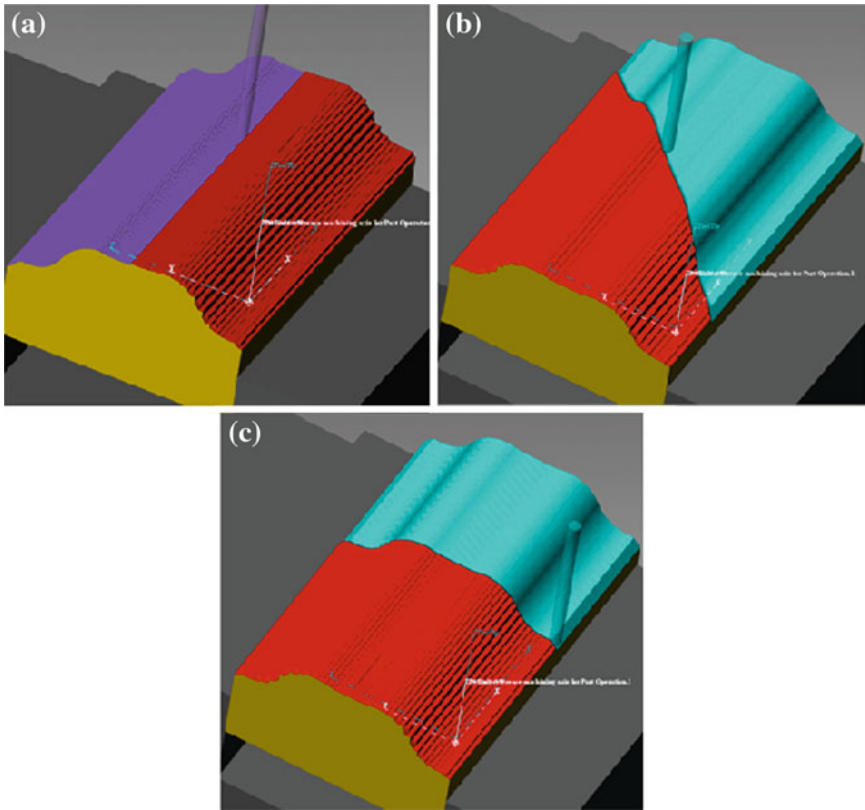
**Feed direction:** The direction that the tool follows to cut the material regarding feed (Fig. 10.4).



**Fig. 10.3** Tool path styles for 3-axis finish machining by sweeping strategies: **a** Zig zag; **b** One way next; **c** One way same

### 5-axis finishing

In 5-axis finishing, the strategy of tool axis orientation is determined regarding the geometrical features of the tool and the part to be machined. The tool can be rotated by an angle between its vertical axis and the surface (*lead* angle). This angle is maintained by the tool along the milling direction. In addition, the tool can be rotated by a side angle between its vertical axis and the surface (*tilt* angle). These angles can be constantly or variably adjusted to ensure that the tool's non cutting tip does not contact the surface. *Lead* and *tilt* angles are computed for each point regarding the normal vector of the sculptured surface and are shown in Fig. 10.5a and b respectively.



**Fig. 10.4** Feed directions for the cutting tool: **a** 0°; **b** 45°; **c** 90°

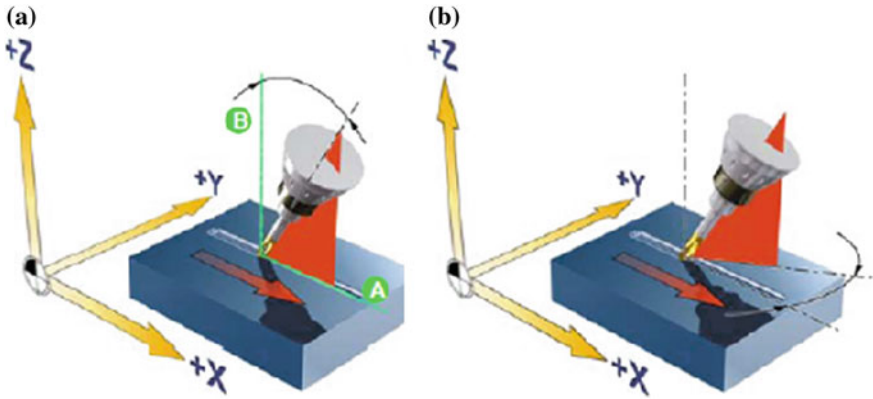
Tool path types are in such a way applied to avoid collisions between the tool's rear side and machined part. The values of these parameters were not considered as an optimization issue at least for the time being; these were determined according the literature and industrial considerations. The tool path types applied to model 5-axis finishing operations for the experiments are as follows:

**Fixed lead and tilt angles for milling:** Two constant values are specified for these angles (Fig. 10.6a)

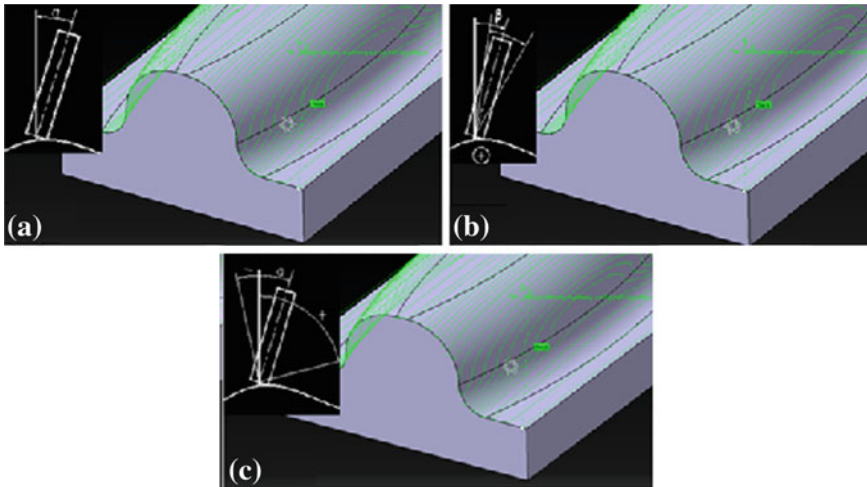
**Fixed lead and variable tilt angles for milling:** a value range for *Tilt* angle and a constant value for *lead* angle are specified (Fig. 10.6b).

**Variable lead and fixed tilt angles for milling:** a value range for *lead* angle and a constant value for *tilt* angle are determined (Fig. 10.6c).

Same determinations to those of 3-axis finishing were applied to the rest of process parameters (speeds and feeds, feed direction and stepover).



**Fig. 10.5** 5-axis machining tool axis orientations: **a** lead angle; **b** tilt angle

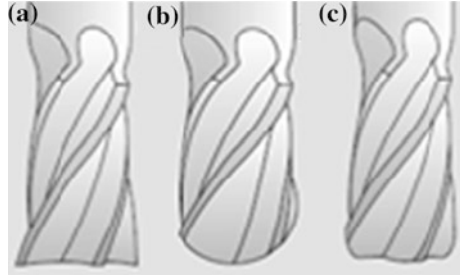


**Fig. 10.6** Tool path styles for 5-axis finish machining by sweeping strategies regarding tool axis orientation: **a** Fixed lead and tilt mode; **b** Fixed lead and variable tilt mode; **c** Variable lead and fixed tilt mode

**Tool type:** As in 5-axis machining more freedom degrees are existed, all types of cutting tools can be implemented. Hence; to assess the different types of end mills applied to 5-axis sculptured surface machining, three tool geometries of constant diameter ( $\varnothing 8$ ), were tested; a flat end mill; (Fig. 10.7a), a ball end mill; (Fig. 10.7b) and corner radius mill (Fig. 10.7c). Since the tool geometry in this case was different, tool orientation issues to prevent collisions were carefully investigated.



**Fig. 10.7** Types of typical HSS end mills: **a** Flat end mill; **b** Ball end mill; **c** Corner radius mill



### 10.4.1.3 Quality Objectives and Objective Function Formulation

It has already been mentioned that three quality objectives were identified to formulate the problem's responses regarding the machining phase. Definitions for these quality objectives are described as follows:

*Machining time:* The time needed to remove the material from a part. Its magnitude is automatically calculated on CNC machine controls and CAM systems after the tool path calculation. Through its calculus, it is easy to observe where productivity is decelerated and how it can be improved.

*Remaining volume* on the model after roughing: The uncut material of a roughed or semi-finished part that remains to be removed at the finishing stage. Lower remaining volume amount leads to quicker finishing operations, lower chip loads and cutting forces thus; drastically reducing overall Machining Time while maintaining high part quality.

*Surface deviation:* The maximum allowed deviation from the mean area computed between the ideally designed and the machining simulated surface. Mean area is computed using Eq. 10.2 as follows:

$$s_{dev} = \left| \frac{A_f - \bar{A}}{\bar{A}} \right| \quad (10.2)$$

where,

$s_{dev}$  the surface deviation in (mm);

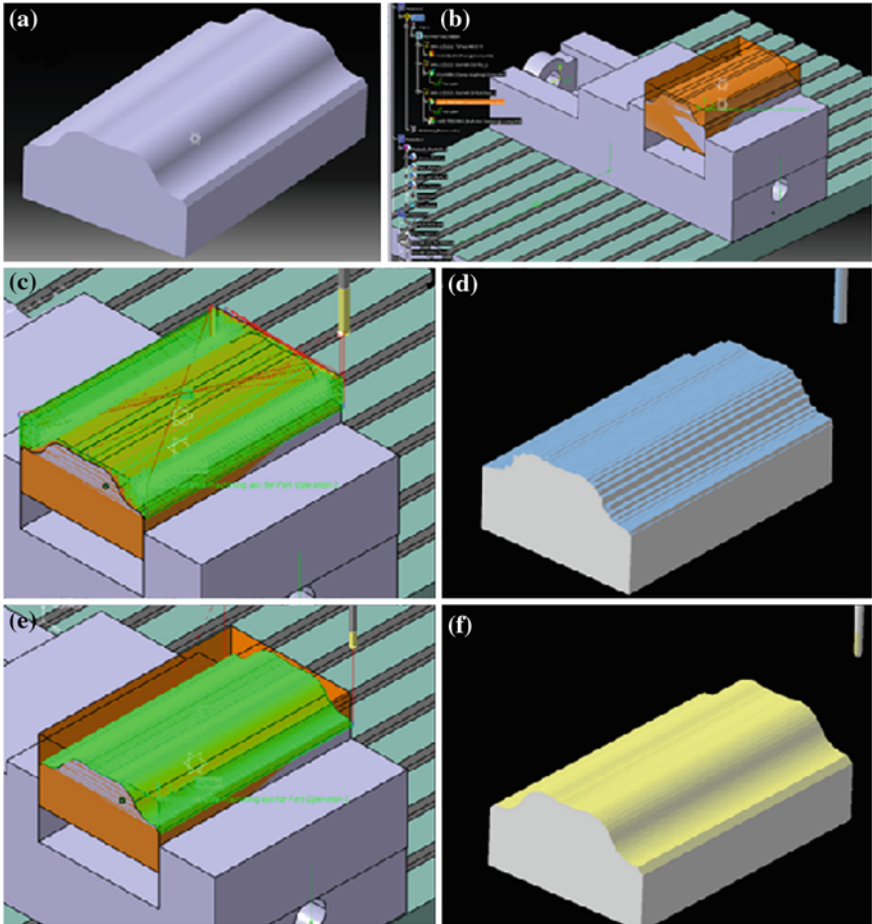
$A_f$  the area of the finished model's surface;

$\bar{A}$  the mean area:  $\bar{A} = \frac{A_f + A_d}{2}$ , ( $A_f$ : finish machined surface;  $A_d$ : designed surface).

The objective functions were formulated as follows:

- For roughing operation

$$QC(v_r, t_{mr}, w_1, w_2) = w_1 \times v_r + w_2 \times t_{mr} \quad (10.3)$$



**Fig. 10.8** Machining modelling stages: **a** Sculptured part; **b** Setup; **c** Rough machining tool path; **d** Resulting roughed model; **e** Finish machining tool path; **f** Resulting finished model

where,

- $QC$  the combination of the measured objectives;
- $v_r$  the remained volume (normalized);
- $t_{mr}$  the normalized rough machining time;
- $w_1, w_2$ , the weight coefficients of the above criteria ( $w_1 = 50\%$  for remaining volume and  $w_2 = 50\%$  for rough machining time).

- For finishing operations (both milling technologies)

$$QC(s_{dev}, t_{mf}, w_3, w_4) = w_3 \times s_{dev} + w_4 \times t_{mf} \tag{10.4}$$

**Table 10.1** Experimental design of rough and finish machining operations

Roughing				
a/a	Factors	Level 1	Level 2	Level 3
1	Machining strategy	“Z offset”	“Z plane”	“Z progressive”
2	Tool diameter, $\varnothing$ (mm)	8	10	12
3	Spindle speed, $\eta$ (rpm)	1800	2200	2600
4	Feed rate, $f$ (mm/min)	300	450	600
5	Stepdown, $a_p$ (% $\varnothing$ )	20 %	35 %	50 %
6	Stepover, $a_e$ (% $\varnothing$ )	35 %	37.5 %	40 %
3-axis finishing				
1	Feed direction (deg)	0°	45°	90°
2	Machining strategy	“Zig zag”	“1 way next”	“1 way same”
3	Tool diameter, $\varnothing$ (mm)	6	7	8
4	Spindle speed, $\eta$ (rpm)	2800	3400	4000
5	Feed rate, $f$ (mm/min)	150	220	290
6	Stepover, $a_e$ (% $\varnothing$ )	5.625 %	8.125 %	10.625 %
5-axis finishing				
1	Feed direction (deg)	0°	45°	90°
2	Machining strategy	“Fixed L&T”	“VL&FT”	“FL&VT”
3	Tool type	Ball end	Flat End	Corner Radius
4	Spindle speed, $\eta$ (rpm)	2800	3400	4000
5	Feed rate (mm/min)	150	220	290
6	Stepover, $a_e$ (% $\varnothing$ )	5.625 %	8.125 %	10.625 %

where,

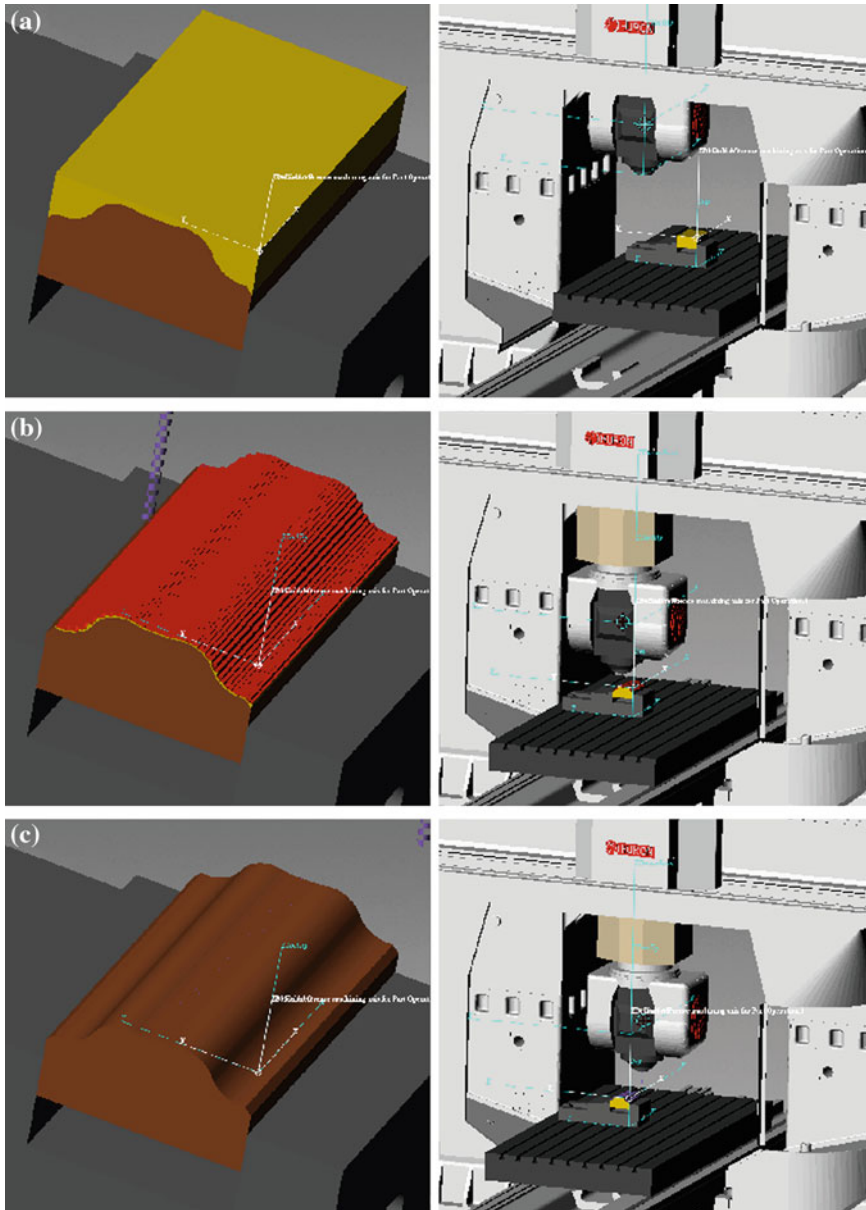
$QC$  the combination of the measured objectives;

$s_{dev}$  the surface deviation (normalized);

$t_{mf}$  the normalized finish machining time;

$w_3, w_4$ , the weight coefficients of the above criteria ( $w_3 = 50$  % for surface deviation and  $w_4 = 50$  % for finish machining time).

Equations 10.3 and 10.4 actually turn multi objective optimization into a single objective one to facilitate computations and simplify the problem’s perplexity. It is essential to say that weight coefficients determine the influential importance for quality objectives and may be specified according the user’s preferences. Hence, quality objectives may be weighted differently thus; leading to different results in a problem’s overall response. Thereby, equality among quality objective weight coefficients was preferred to obtain more clear depiction of results it the current study.



**Fig. 10.9** Machining simulations in CAM software: **a** machine setup; **b** roughing process simulation; **c** finishing process simulation

**Table 10.2** Rough machining results for machining time,  $t_{mr}$  (min) and remaining volume,  $v_r$  ( $\text{mm}^3$ )

No	Machining strategy	Tool ( $\emptyset$ )	$\eta$ (rpm)	$f$ (mm/min)	$\alpha_p$ ( $\% \emptyset$ )	$\alpha_e$ ( $\% \emptyset$ )	$t_{mr}$ (min)	$v_r$ ( $\text{mm}^3$ )
1	Z offset	8	1800	300	20	35	455	61941.792
2	Z offset	8	2200	450	35	37.5	166.23	63161.373
3	Z offset	8	2600	600	50	40	76.35	64501.283
4	Z offset	10	1800	450	35	40	98.55	69966.775
5	Z offset	10	2200	600	50	35	57.53	66447.247
6	Z offset	10	2600	300	20	37.5	254.2	68213.631
7	Z offset	12	1800	600	50	37.5	50.18	73376.137
8	Z offset	12	2200	300	20	40	189.33	75966.469
9	Z offset	12	2600	450	35	35	92.6	71028.856
10	Z plane	8	1800	450	50	37.5	141.06	63200.537
11	Z plane	8	2200	600	20	40	199.06	64603.863
12	Z plane	8	2600	300	35	35	288.25	61992.238
13	Z plane	10	1800	600	20	35	153.01	66569.789
14	Z plane	10	2200	300	35	37.5	185.43	68329.643
15	Z plane	10	2600	450	50	40	87.22	70030.827
16	Z plane	12	1800	300	35	40	125.43	76171.811
17	Z plane	12	2200	450	50	35	73.43	71077.830
18	Z plane	12	2600	600	20	37.5	102.33	73564.057
19	Z prg	8	1800	600	35	40	568.23	64488.924
20	Z prg	8	2200	300	50	35	895.38	61927.185
21	Z prg	8	2600	450	20	37.5	1440.31	63137.108
22	Z prg	10	1800	300	50	37.5	550.05	68200.859
23	Z prg	10	2200	450	20	40	851.9	69922.149
24	Z prg	10	2600	600	35	35	423.06	66416.966
25	Z prg	12	1800	450	20	35	687.3	71015.700
26	Z prg	12	2200	600	35	37.5	287.03	73336.818
27	Z prg	12	2600	300	50	40	380.15	75936.300

Since quality characteristics are of different magnitudes, an inherent bias may appear owing to objective values' order and results can be quite disorienting to the problem's response. Therefore, given that the volume is in  $\text{mm}^3$ , machining time in min and surface deviation in mm, a proper normalization of the results was conducted [45].

#### 10.4.1.4 Technological Constraints

In order to select the proper tools and meaningful machining parameter ranges, the configurations and properties of the CNC machine tool were taken into account. For the CNC machine selected to simulate the experiments in the CAM environment, the following constraints were specified:

**Table 10.3** 3-axis finish machining results for machining time,  $t_{mf}$  (min) and surface deviation,  $s_{dev}$  (mm) based on optimum roughing as the first process

No	Machining strategy	Tool ( $\emptyset$ )	$\eta$ (rpm)	$f$ (mm/min)	$FD$ (deg)	$\alpha_e$ ( $\% \emptyset$ )	$t_{mf}$ (min)	$s_{dev}$ (mm)
1	Zig zag	6	2800	150	0	5.625	729.32	0.00065
2	Zig zag	6	3400	220	45	8.125	374.38	0.00044
3	Zig zag	6	4000	290	90	10.625	233.18	0.00036
4	Zig zag	7	2800	220	45	10.625	247.25	0.00045
5	Zig zag	7	3400	290	90	5.625	375.21	0.00036
6	Zig zag	7	4000	150	0	8.125	435.01	0.00069
7	Zig zag	8	2800	290	90	8.125	228.23	0.00040
8	Zig zag	8	3400	150	0	10.625	293.13	0.00076
9	Zig zag	8	4000	220	45	5.625	409.13	0.00035
10	1 W next	6	2800	220	90	8.125	397.25	0.00042
11	1 W next	6	3400	290	0	10.625	199.4	0.00087
12	1 W next	6	4000	150	45	5.625	792.4	0.00040
13	1 W next	7	2800	290	0	5.625	325.23	0.00062
14	1 W next	7	3400	150	45	8.125	472.31	0.00044
15	1 W next	7	4000	220	90	10.625	262.13	0.00050
16	1 W next	8	2800	150	45	10.625	317.36	0.00052
17	1 W next	8	3400	220	90	5.625	434.26	0.00035
18	1 W next	8	4000	290	0	8.125	198.16	0.00069
19	1 W same	6	2800	290	45	10.625	217.33	0.00055
20	1 W same	6	3400	150	90	5.625	842.5	0.00038
21	1 W same	6	4000	220	0	8.125	343.5	0.00075
22	1 W same	7	2800	150	90	8.125	502.11	0.00039
23	1 W same	7	3400	220	0	10.625	227.26	0.00078
24	1 W same	7	4000	290	45	5.625	353.23	0.00036
25	1 W same	8	2800	220	0	5.625	376.28	0.00062
26	1 W same	8	3400	290	45	8.125	215.04	0.00041
27	1 W same	8	4000	150	90	10.625	337.5	0.00040

- Maximum available motor power:  $P = 18$  kW
- Maximum available spindle speed:  $n = 10000$  rpm.
- Maximum available feed rate velocity:  $V_f = 4000$  mm/min (40 m/min).
- Machine tool rapid traverses:  $V_r = 4000$  mm/min.

The power demand is used as the major constraint and is defined applying Eq. 10.5:

$$P_C = \frac{a_p \times a_e \times v_f}{60 \times 10^6 \times \eta} \times k_c, \quad (10.5)$$

where,

$P_C$  the demanded cutting power from the machine tool in kW;

$a_p$  cutting depth (in mm);

$a_e$  radial cutting depth (in mm);

**Table 10.4** Five axis finish machining results for machining time,  $t_{mf}$  (min) and surface deviation,  $s_{dev}$  (mm) based on optimum roughing as the first process

No	Machining strategy	Tool type	$n$ (rpm)	$f$ (mm/min)	$FD$ (deg)	$\alpha_e$ ( $\% \emptyset$ )	$t_{mf}$ (min)	$s_{dev}$ (mm)
1	Fixed L&T	Ball	2800	150	0	5.625	619	0.00084
2	Fixed L&T	Ball	3400	220	45	8.125	348.45	0.00039
3	Fixed L&T	Ball	4000	290	90	10.625	327.7	0.00077
4	Fixed L&T	Flat	2800	220	45	10.625	270.08	0.00043
5	Fixed L&T	Flat	3400	290	90	5.625	330.8	0.00021
6	Fixed L&T	Flat	4000	150	0	8.125	430.26	0.00091
7	Fixed L&T	Radius	2800	290	90	8.125	226.13	0.00040
8	Fixed L&T	Radius	3400	150	0	10.625	328.28	0.00019
9	Fixed L&T	Radius	4000	220	45	5.625	507.31	0.00040
10	VL&FT	Ball	2800	220	90	8.125	320.16	0.00015
11	VL&FT	Ball	3400	290	0	10.625	230.11	0.00066
12	VL&FT	Ball	4000	150	45	5.625	738.16	0.00087
13	VL&FT	Flat	2800	290	0	5.625	324.48	0.00085
14	VL&FT	Flat	3400	150	45	8.125	517.18	0.00036
15	VL&FT	Flat	4000	220	90	10.625	230.11	0.00031
16	VL&FT	Radius	2800	150	45	10.625	333.56	0.00031
17	VL&FT	Radius	3400	220	90	5.625	430.26	0.00058
18	VL&FT	Radius	4000	290	0	8.125	222.55	0.00039
19	FL&VT	Ball	2800	290	45	10.625	202.05	0.00097
20	FL&VT	Ball	3400	150	90	5.625	295.26	0.00018
21	FL&VT	Ball	4000	220	0	8.125	293.35	0.00017
22	FL&VT	Flat	2800	150	90	8.125	447.3	0.00036
23	FL&VT	Flat	3400	220	0	10.625	223.56	0.00104
24	FL&VT	Flat	4000	290	45	5.625	386.41	0.00132
25	FL&VT	Radius	2800	220	0	5.625	422.03	0.00030
26	FL&VT	Radius	3400	290	45	8.125	266.3	0.00013
27	FL&VT	Radius	4000	150	90	10.625	344.26	0.00023

$v_f$  cutting speed (m/min);

$\eta$  spindle motor efficiency;

$k_c$  cutting force per  $\text{mm}^2$  depending on the material type.

## 10.5 Experimental

### 10.5.1 Machining Setup Preparation with CAM Software

The machining setup of a test sculptured part was prepared in commercial CAM system. Figure 10.8 illustrates the stages involved to the initial part setup and its machining processes.

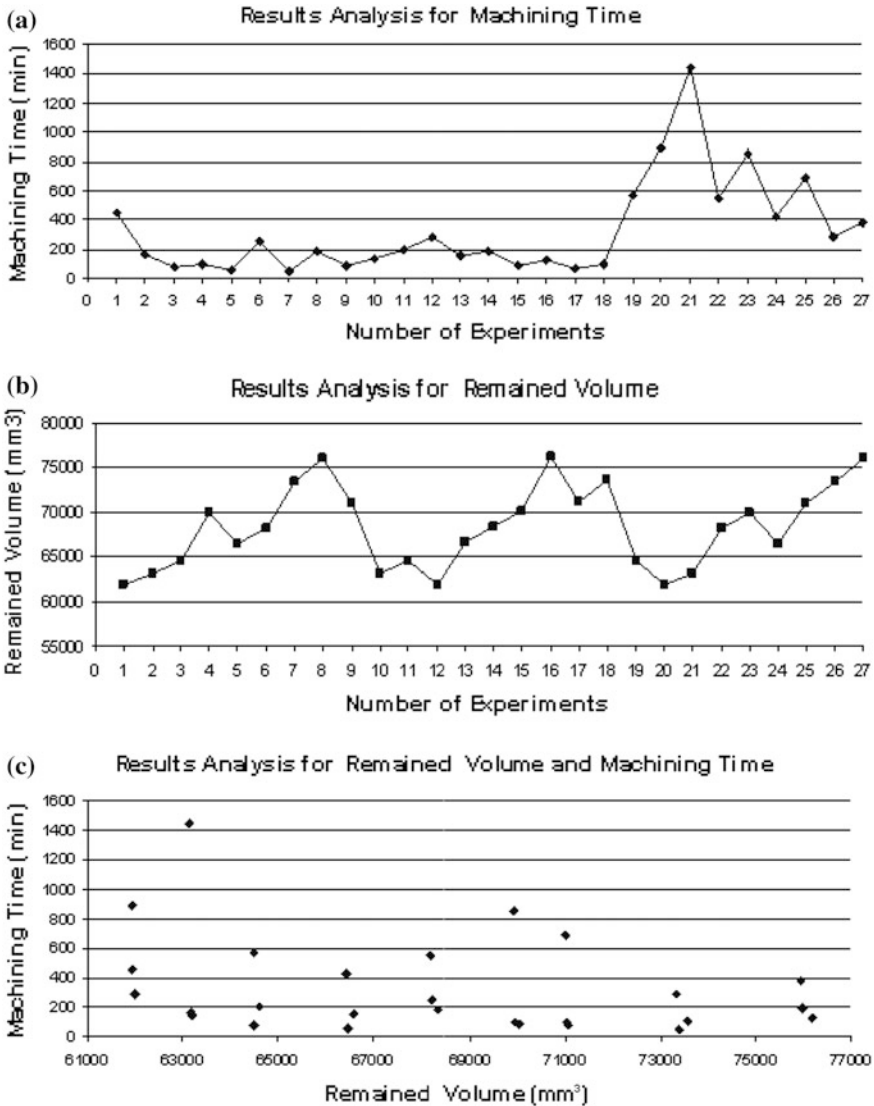


Fig. 10.10 Results analysis for roughing: a rough machining time; b remaining volume and c both quality objectives

### 10.5.2 Design of Experiments

Fractional factorial approach was adopted to design the machining experiments since this approach investigates only the fraction that combines all possible repetitions without losing important information. The machining parameters involved





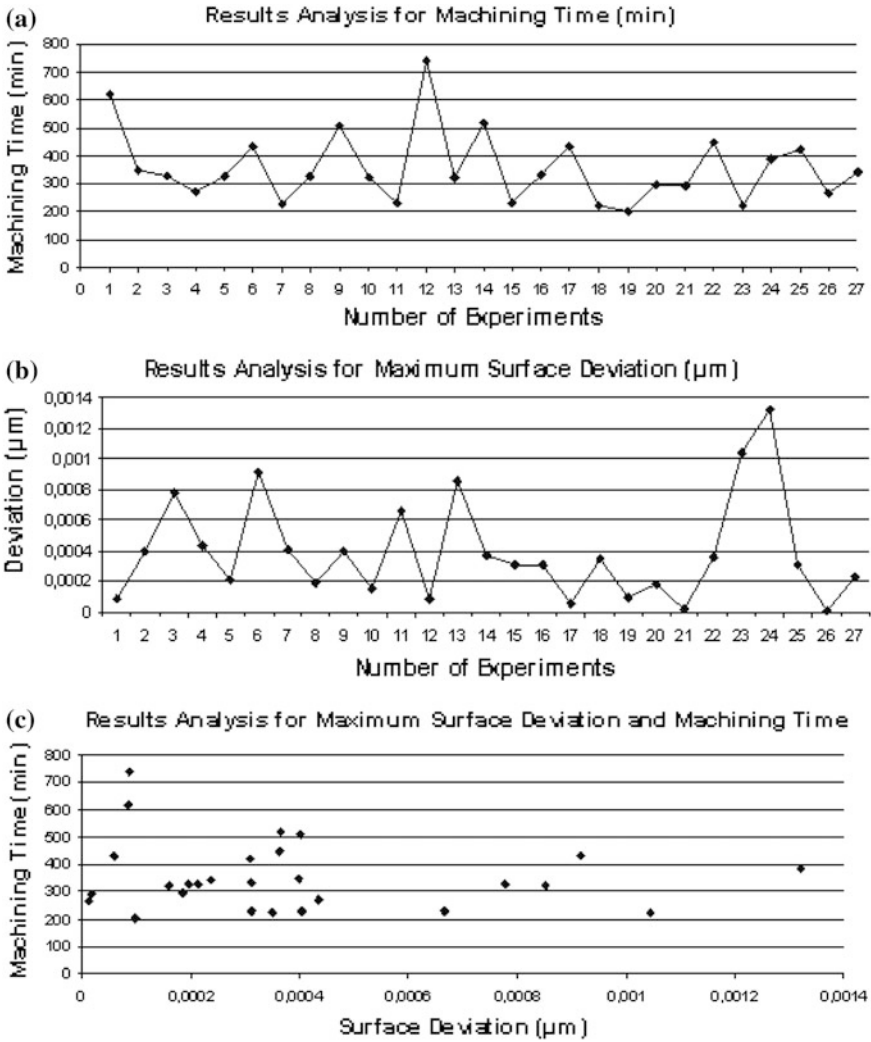


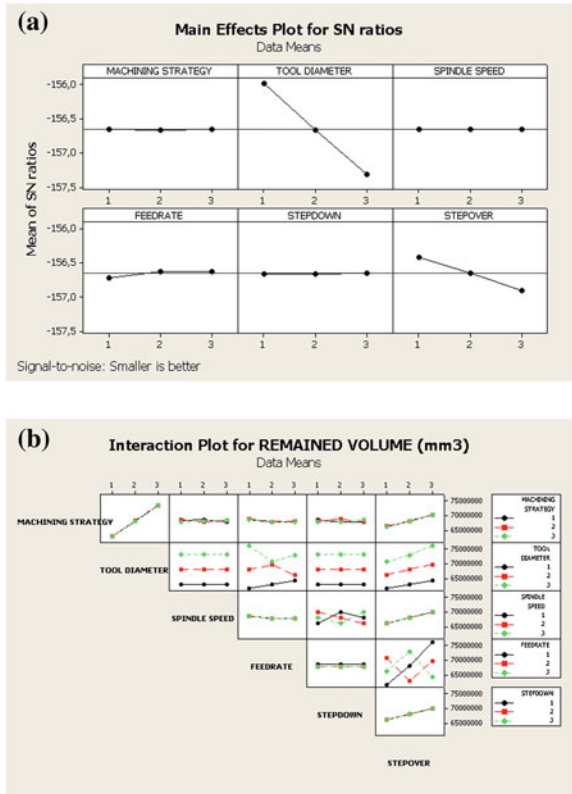
Fig. 10.12 Results analysis for 5-axis finishing: a finish machining time, b surface deviation and c both quality objectives

### 10.5.3 Machining Simulation and Experimental Data Collection

#### 10.5.3.1 Roughing and Finishing Experiments

The results obtained after the experimental machining simulations were utilized to compute the combinatorial objective function by applying the formulas presented

**Fig. 10.13** **a** Main effects plot; **b** Interaction plot for rough machining parameters subjected to remaining volume



above (Eqs. 10.3 and 10.4). Typical machining simulations for roughing (Fig. 10.9a) and finishing operations (Fig. 10.9b) are illustrated below.

### 10.5.3.2 Experimental Results

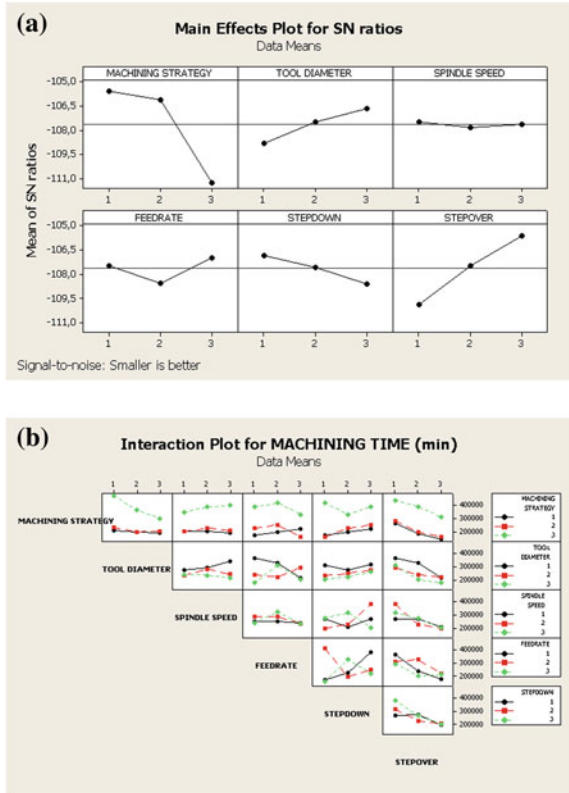
The results obtained after the machining simulations are presented in Tables 10.2, 10.3 and 10.4 for roughing, 3-axis finishing and 5-axis finishing respectively.

### 10.5.3.3 Experimental Analysis of Results for Machining Experiments

#### Roughing operation

Analysis for roughing operation indicated that unlike *Z progressive* strategy which is time consuming in its attempt to maintain volume uniformity, *Z offset* and *Z plane* strategies produce tool paths that tend to shorten cycle times. As for the

**Fig. 10.14** **a** Main effects plot; **b** Interaction plot for rough machining parameters subjected to machining time for roughing

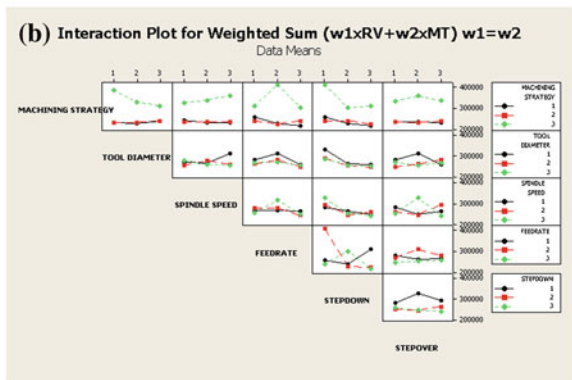
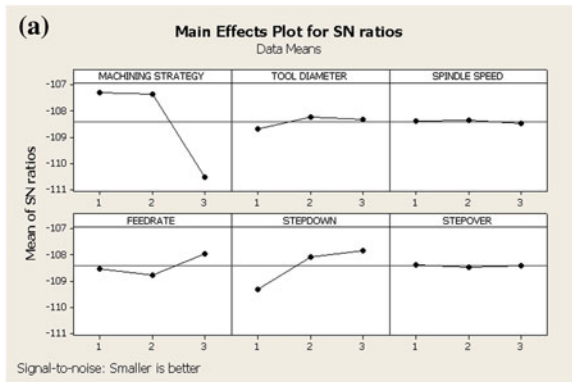


remaining volume, the rest of the parameters are responsible for its variations. However variations concerning remaining volume are quite uniform for all three machining strategies. Figure 10.10a illustrates the variations of rough machining time,  $t_{mr}$  which rough machining experiments yielded; whilst Fig. 10.10b illustrates the variations of remaining volume,  $v_r$ . Finally, Fig. 10.10c depicts the efficiency of these experiments to both quality objectives. The closest are the dots in the diagram's axes origin; the greater the efficiency is. More detailed analysis is given from further statistical analysis followed.

3-axis finishing operation

Analysis for 3-axis finishing operation indicated that those machining experiments having the lowest level of feed rate, resulted to longer machining times than the rest of the experiments (Fig. 10.11a). A first assumption is that feed rate is the

**Fig. 10.15** **a** Main effects plot for rough machining parameters subjected to an equally formulated combination among the objectives (remaining volume and rough machining time)

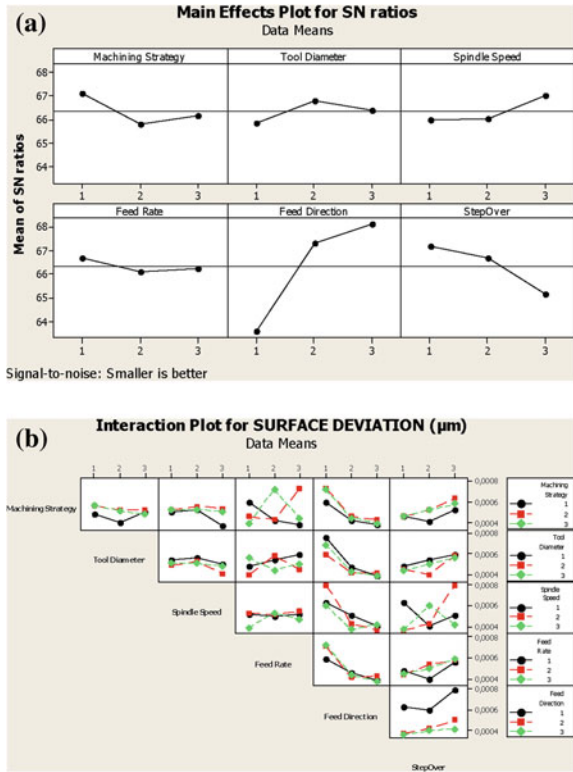


dominant parameter affecting machining time. As far as surface deviation is concerned, the results obtained are uniformly distributed with few experimental runs as exception that lowered surface deviation even more (Fig. 10.11b). Figure 10.11c depicts the efficiency of 3-axis finishing experiments to both quality objectives. Most of the experimental runs tend to approach the axes origin, indicating hence a good efficiency.

### 5-axis finishing operation

Analysis for 5-axis finishing operation indicated that variations of machining time are as uniform as those observed to 3-axis finishing runs with the difference that lower values were obtained (Fig. 10.12a). As for the surface deviation, the main

**Fig. 10.16** **a** Main effects plot; **b** Interaction plot for 3-axis finish machining parameters subjected to surface deviation

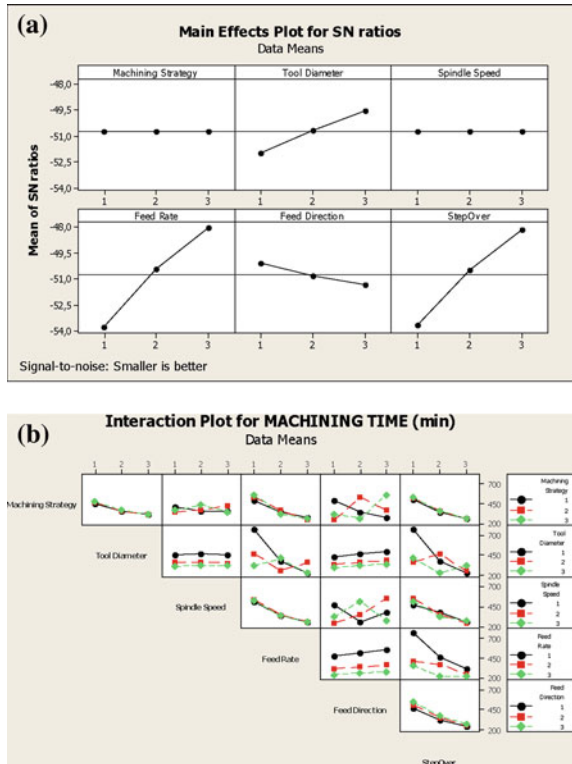


observation is that less uniformity than the one noticed in 3-axis finishing experiments appeared among results (Fig. 10.12b). Nevertheless the results of 5-axis finishing produced finer surface quality (referring to surface deviation). Figure 10.12c depicts the efficiency of 5-axis finishing experiments to both quality objectives. It is clearly observed that more solutions that those obtained from 3-axis finishing tend to approach the axes origin (more clustered solutions).

### 10.5.3.4 Statistical Analysis of Results for Machining Experiments

Further statistical analysis was conducted to study the influence of machining parameters to quality objectives. Statistical analysis involved the generation of main effects plots for S/N ratios and interaction plots among parameters to the equally weighted expression (Figs. 10.13, 10.14, 10.15, 10.16, 10.17, 10.18, 10.19, 10.20 and 10.21).

**Fig. 10.17** **a** Main effects plot; **b** Interaction plot for 3-axis finish machining parameters subjected to machining time for 3-axis finishing



Analysis of results for machining experiments (Category 1)

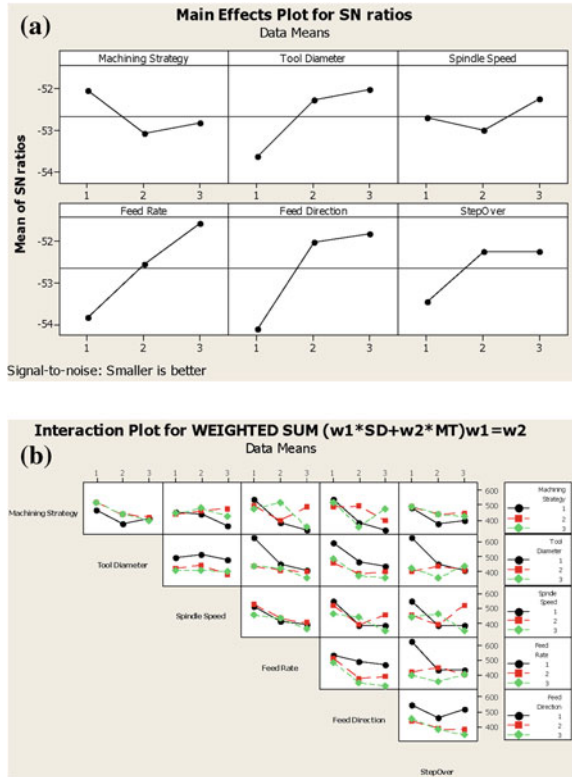
- Roughing phase (quality objective: Remaining volume,  $v_r$ )
- Roughing phase (quality objective: Rough machining time,  $t_{mr}$ )
- Roughing phase ( $WS = v_r \times 0.5 + t_{mr} \times 0.5$ )
- 3-axis finishing phase (quality objective: Surface deviation,  $s_{dev}$ )
- 3-axis finishing phase (quality objective: Finish machining time,  $t_{mf}$ )
- 3-axis finishing phase ( $WS = s_{dev} \times 0.5 + t_{mf} \times 0.5$ ).

Analysis of results for Machining Experiments (Category 2)

- 5-axis finishing phase (Quality objective: Surface deviation,  $s_{dev}$ )
- 5-axis finishing phase (Quality objective: Finish machining time,  $t_{mf}$ )
- 5-axis finishing phase ( $WS = s_{dev} \times 0.5 + t_{mf} \times 0.5$ ).

Even though ANOVA analysis of process parameters referred to each of the quality objectives was conducted, only results referred to the combinatorial expression of quality objectives are illustrated in the work since this was of main interest. ANOVA results are illustrated in Tables 10.5, 10.6 and 10.7 for roughing, 3-axis and 5-axis finishing parameters respectively.

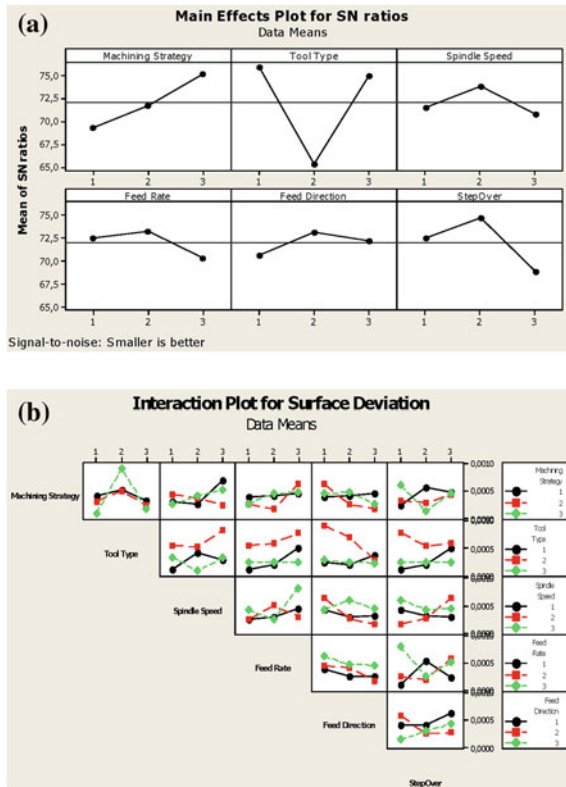
**Fig. 10.18** **a** Main effects plot; **b** Interaction plot for 3-axis finish machining parameters subjected to an equally formulated combination among the objectives (machining time and surface deviation)



According to statistical analysis of results obtained, machining strategy is the most important aspect to consider when modeling rough machining operations. Besides, proper rough machining strategy leads to shorter cycle times itself. It is mentioned that several tool path strategies may be available for their applications to sculptured parts but each of these may yield different variations to responses. Radial depth of cut and feed rate are also of great importance to minimize overall response. Feed rate does not influence remaining volume but as a dominant parameter to rough machining time affects also their relation. Tool’s nominal diameter plays important role to the resulting remained volume on the part’s surface. Smaller tools facilitate this quality objective since they can penetrate to the material more efficiently and leave more uniform volume that will maintain constant chip loads and relatively low cutting forces to finishing. On the other hand, smaller tools in diameter result to longer cycle times. The combinatorial criterion formulated for this process indicated that a small cutter under the proper roughing strategy may satisfy both criteria. Strong interactions were noticed between to almost all machining parameters. Spindle speed with feed rate yielded strong interaction effects to rough machining response.



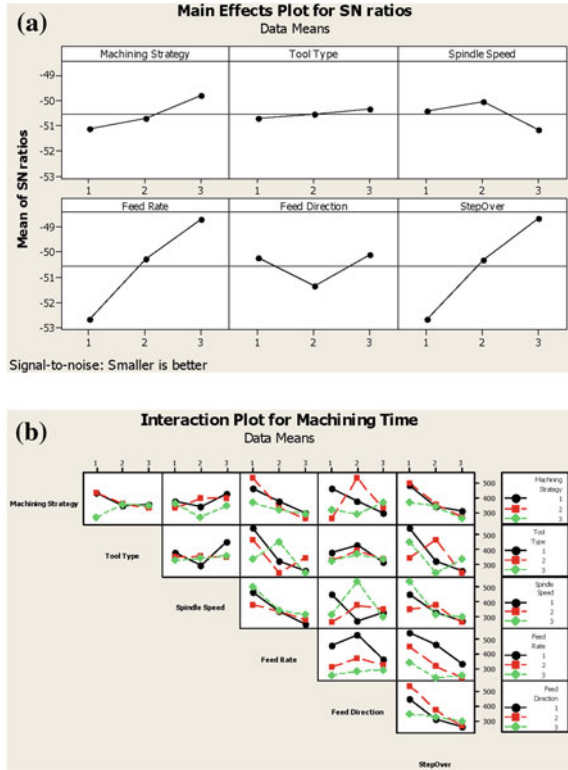
**Fig. 10.19** **a** Main effects plot; **b** Interaction plot for 5-axis finish machining parameters subjected to surface deviation



For 3-axis finishing operations results indicated that feed direction greatly affects the combinational response among finish machining time and surface deviation. However, feed direction may be constrained in advance regarding the part's geometrical features. Should this parameter has to be adjusted, one has to check the appropriate feed angle otherwise the cutter may be subjected to large trajectory variations regarding the sculptured contour; thus increasing cycle times or even to collisions among tool and part or tool and fixture. Feed rate and tool diameter follow next in terms of influence. In general, low feed rate values increase finish machining time whilst the opposite occurs when higher values are determined as it is observed in actual machining experiments. As far as tool diameter is concerned, smaller cutters produce finer surface finish hence may be generally preferred. The strategy applied along with small values for radial cutting depth to the tool; improve surface characteristics in 3-axis finishing. Concerning interactions among parameters in 3-axis finishing, noticeable ones occurred mainly among machining strategy and tool diameter.

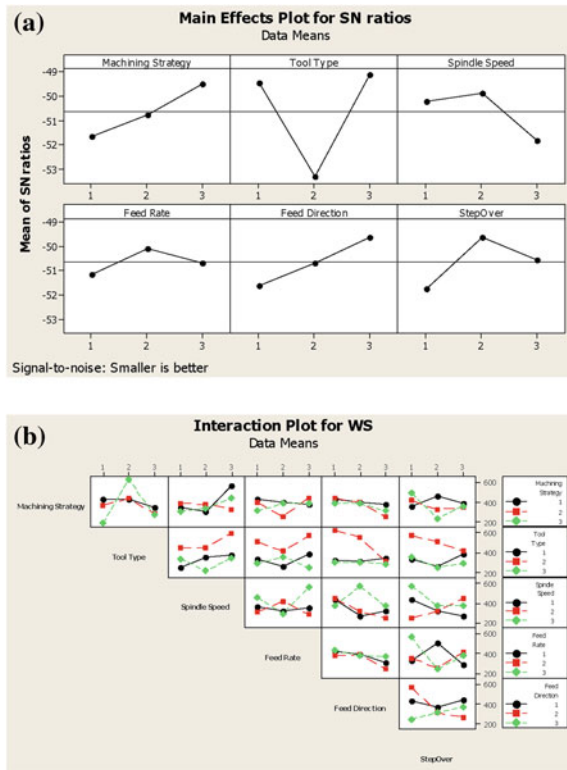
For 5-axis finishing the type of the cutter was found as the most influential to the outcome of both machining time and surface deviation. Results indicated that corner radius and flat end cutters produce finer surfaces than ball end mills.

**Fig. 10.20** **a** Main effects plot; **b** Interaction plot for 5-axis finish machining parameters subjected to machining time for 5-axis finishing



A machining strategy with suitable *lead* and *tilt* angles along with a proper choice of tool type (flat end or corner radius) are also responsible for a good surface finish. In general, different resulting outputs in terms of quality objectives may be obtained from process parameters' settings determined under different machining strategies. Especially for 5-axis machining operations, optimum regions to determine *lead* and *tilt* angles should be found to obtain more clear and concise results in terms of 5-axis machining strategies influence when using CAM systems. Spindle speed and feed rate are also of major importance to these objectives. Assuming that three different types of milling cutters were tested under three different 5-axis “*lead and tilt*” tool paths, one comes to verify that specifications for *lead/tilt* angles are adjusted mainly with the tool type as a reference. This is verified also from the interactions among machining strategy and tool type. It is of great importance to mention that specifications for *lead/tilt* angle values should satisfy issues deal with collisions detection to machining setups.

**Fig. 10.21** **a** Main effects plot for 5-axis finish machining parameters subjected to an equally formulated combination among the objectives (machining time and surface deviation)



**10.5.3.5 Formulation of Optimum Machining Sequences**

Machining programs were formulated for the two scenarios and optimum ones were highlighted. Figures 10.22 and 10.23 depict these machining programs for 3- and 5-axis machining scenarios respectively, whilst Tables 10.8 and 10.9 illustrate their parameter values. A comparison diagram in terms of process efficiency for 3- and 5-axis machining programs is illustrated in Fig. 10.24.

Optimum machining programs from each category were used to virtually machine the test sculptured model in *CG Tech's VERICUT*<sup>®</sup> [46] so as to verify their efficiency. As standard modules of the software for NC verification, *Auto-DIFF*<sup>®</sup> and *X Caliper*<sup>®</sup> were utilized to analyze the resulted models. The verification analysis was based mainly to excess material error. Excess material error, occurred lower during the analysis of the 5-axis machined model, than the one obtained by the 3-axis machined one. Figure 10.25a illustrates the resulting model machined in 3-axis; whereas Fig. 10.25b depicts the resulting model machined in 5-axis mode.

**Table 10.5** ANOVA results for Eq. 10.2 (roughing operation)

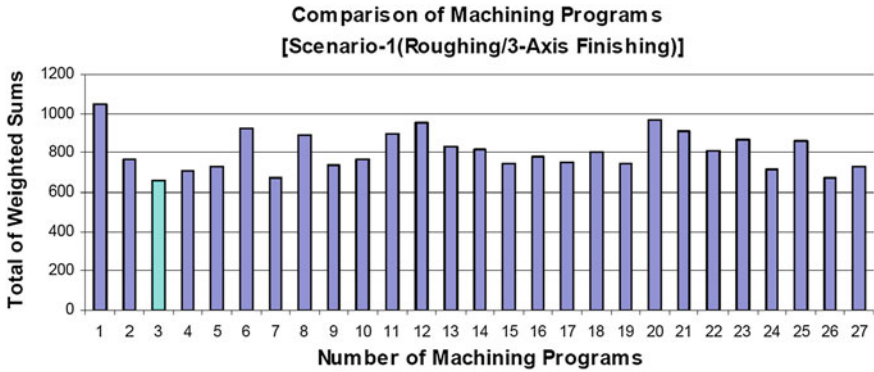
Source	DF	SS	MS	F	P
<i>Strategy</i>	2	405491	202745	19.96	0.000
Error	24	243736	10156		
Total	26	649226			
S = 100.8 R-Sq = 62.46 % R-Sq(adj) = 59.33 % Pooled StDev = 100.8					
<i>Tool diameter</i>	2	27988	13994	0.54	0.589
Error	24	621238	25885		
Total	26	649226			
S = 160.9 R-Sq = 4.31 % R-Sq(adj) = 0.00 % Pooled StDev = 160.9					
<i>Spindle speed</i>	2	1500	750	0.03	0.973
Error	24	647727	26989		
Total	26	649226			
S = 164.3 R-Sq = 0.23 % R-Sq(adj) = 0.00 % Pooled StDev = 164.3					
<i>Feed rate</i>	2	47547	23774	0.95	0.401
Error	24	601679	25070		
Total	26	649226			
S = 158.3 R-Sq = 7.32 % R-Sq(adj) = 0.00 % Pooled StDev = 158.3					
<i>Stepdown</i>	2	78719	39360	1.66	0.212
Error	24	570507	23771		
Total	26	649226			
S = 154.2 R-Sq = 12.13 % R-Sq(adj) = 4.80 % Pooled StDev = 154.2					
<i>Stepover</i>	2	2567	1284	0.05	0.954
Error	24	646659	26944		
Total	26	649226			
S = 164.1 R-Sq = 0.40 % R-Sq(adj) = 0.00 % Pooled StDev = 164.1					

**Table 10.6** ANOVA results for Eq. 10.3 (3-axis finishing operation)

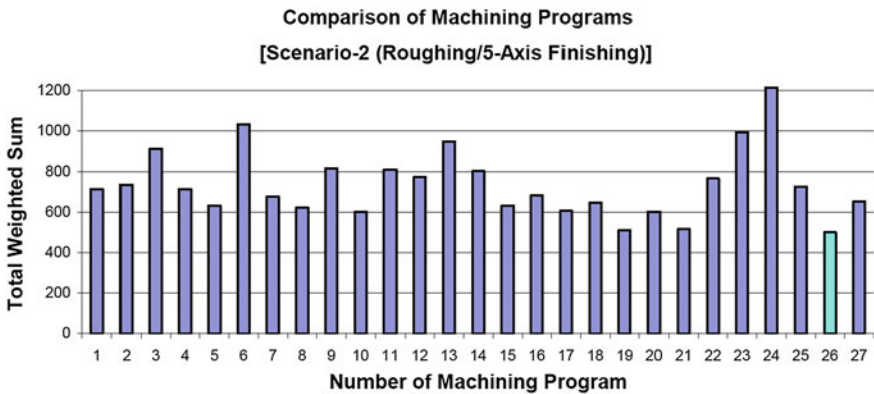
Source	DF	SS	MS	F	P
<i>Machining strategy</i>	2	8562	4281	0.43	0.655
Error	24	238945	9956		
Total	26	247507			
S = 99.78 R-Sq = 3.46 % R-Sq(adj) = 0.00 % Pooled StDev = 99.78					
<i>Tool diameter</i>	2	45416	22708	2.70	0.088
Error	24	202091	8420		
Total	26	247507			
S = 91.76 R-Sq = 18.35 % R-Sq(adj) = 11.55 % Pooled StDev = 91.76					
<i>Spindle speed</i>	2	6516	3258	0.32	0.726
Error	24	240990	10041		
Total	26	247507			
S = 100.2 R-Sq = 2.63 % R-Sq(adj) = 0.00 % Pooled StDev = 100.2					
<i>Feed rate</i>	2	60638	30319	3.89	0.034
Error	24	186869	7786		
Total	26	247507			
S = 88.24 R-Sq = 24.50 % R-Sq(adj) = 18.21 % Pooled StDev = 88.24					
<i>Feed direction</i>	2	73771	36885	5.10	0.014
Error	24	173736	7239		
Total	26	247507			
S = 85.08 R-Sq = 29.81 % R-Sq(adj) = 23.96 % Pooled StDev = 85.08					
<i>Stepover</i>	2	27926	13963	1.53	0.238
Error	24	219580	9149		
Total	26	247507			
S = 95.65 R-Sq = 11.28 % R-Sq(adj) = 3.89 % Pooled StDev = 95.65					

**Table 10.7** ANOVA results for Eq. 10.3 (5-axis finishing operation)

Source	DF	SS	MS	F	P
<i>Machining strategy</i>	2	9734	4867	0.16	0.854
Error	24	735165	30632		
Total	26	744898			
S = 175.0 R-Sq = 1.31 % R-Sq(adj) = 0.00 % Pooled StDev = 175.0					
<i>Tool type</i>	2	214105	107052	4.84	0.017
Error	24	530794	22116		
Total	26	744898			
S = 148.7 R-Sq = 28.74 % R-Sq(adj) = 22.80 % Pooled StDev = 148.7					
<i>Spindle speed</i>	2	57186	28593	1.00	0.383
Error	24	687712	28655		
Total	26	744898			
S = 169.3 R-Sq = 7.68 % R-Sq(adj) = 0.00 % Pooled StDev = 169.3					
<i>Feed rate</i>	2	14813	7407	0.24	0.786
Error	24	730085	30420		
Total	26	744898			
S = 174.4 R-Sq = 1.99 % R-Sq(adj) = 0.00 % Pooled StDev = 174.4					
<i>Feed direction</i>	2	51486	25743	0.89	0.423
Error	24	693413	28892		
Total	26	744898			
S = 170.0 R-Sq = 6.91 % R-Sq(adj) = 0.00 % Pooled StDev = 170.0					
<i>Stepover</i>	2	32379	16190	0.55	0.587
Error	24	712519	29688		
Total	26	744898			
S = 172.3 R-Sq = 4.35 % R-Sq(adj) = 0.00 % Pooled StDev = 172.3					



**Fig. 10.22** Efficiency of machining programs (category 1) in terms of the total weighted sum calculated (3rd experiment indicated as best)



**Fig. 10.23** Efficiency of machining programs (category 2) in terms of the total weighted sum calculated (26th experiment indicated as best)

### 10.6 Verification Experiment: Actual Machining Process of an Industrial Part

The methodology presented, was applied also to a machine an actual industrial part from automotive industry. The part was machined in a 3-axis CNC machine tool to give a general view in terms of machining efficiency and surface quality when the

**Table 10.8** Best machining sequence's results for program scenario 1

Machining program 1						
Stage: Roughing						
Parameters	Tool $\emptyset$	$\alpha_c$ (% $\emptyset$ )	Strategy	$\alpha_p$ (% $\emptyset$ )	$f$ (mm/min)	$\eta$ (rpm)
Levels	8 mm	40 (3.2 mm)	Z-offset	50 (4.0 mm)	600	2600
Quality objective	<i>Remaining volume</i>					
	64501.283 mm <sup>3</sup>					
Stage:	Finishing					
Parameters	Tool $\emptyset$	$f$ (mm/min)	Strategy	$\alpha_c$ (% $\emptyset$ )	$FD$ (deg)	$\eta$ (rpm)
Levels	6 mm	290	Sweeping zig zag	10.625 (0.64 mm)	90	4000
Quality objective	<i>Finish machining time</i>					
	233.18 min					
Totals:	Program machining time (min)	309.53	Remained volume (mm <sup>3</sup> )	64501.283	Surface deviation (mm)	0.0030



**Table 10.9** Best machining sequence's results for program scenario 2  
Machining program 2

Stage:		Roughing				
Parameters	Tool $\emptyset$	$\alpha_e$ (% $\emptyset$ )	Strategy	$\alpha_p$ (% $\emptyset$ )	$f$ (mm/min)	$\eta$ (rpm)
Level	8 mm	40 (3.2 mm)	Z offset	50 (4.0 mm)	600	2600
Quality objective	Remaining volume 64501.283 mm <sup>3</sup>					
Stage:	Finishing					
Parameters	Tool type	$f$ (mm/min)	Strategy	$\alpha_e$ (% $\emptyset$ )	$FD$ (deg)	$\eta$ (rpm)
Levels	Radius	290	FL&VT	8.125 (0.65 mm)	45	3400
Quality objective	Finish machining time 233.18 min					
Totals:	Program machining time (min)	342.65	Remained volume (mm <sup>3</sup> )	64501.283	Surface deviation (mm)	0.00013

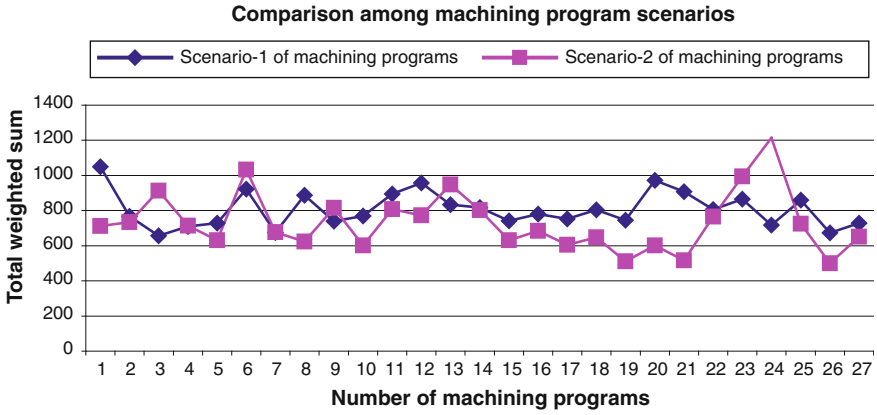


Fig. 10.24 Comparison among machining programs formulated for both categories

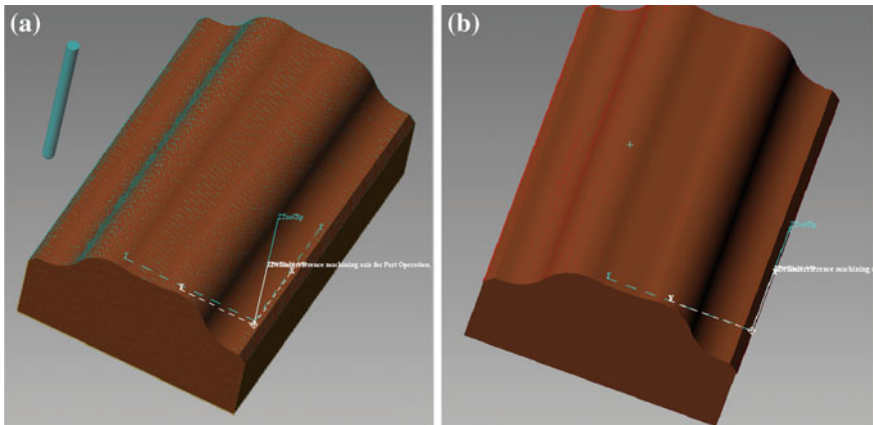
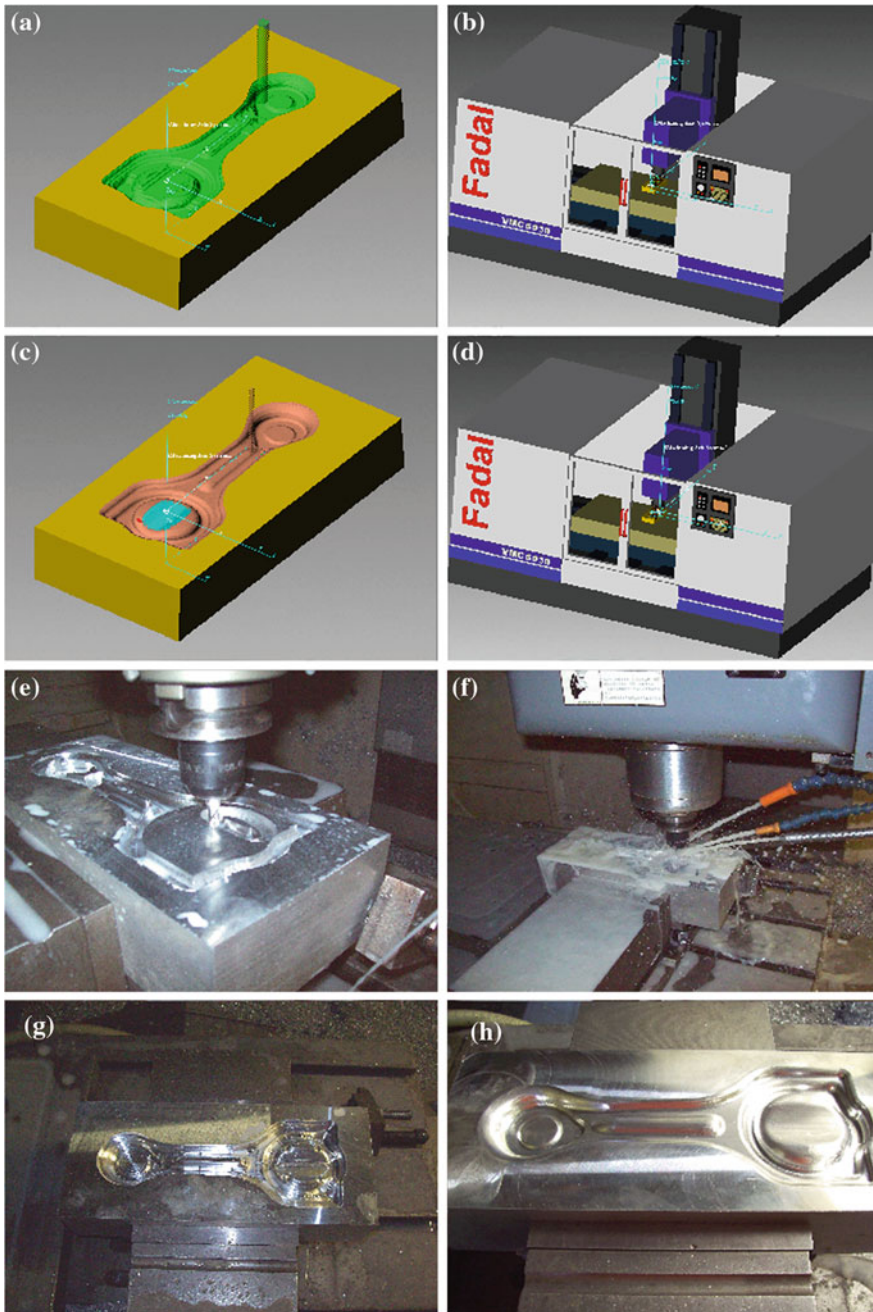


Fig. 10.25 Comparison analysis for the test sculptured part performed in VERICUT®: a 3-axis machining technology for finishing; b 5-axis machining technology for finishing

proposed approach is utilized. Both machining time and surface quality improved against ordinary practices. In Fig. 10.26a–d the machining process of the part simulated in VERICUT® [46] is depicted whilst Fig. 10.26e–h depict the actual machining operation along with the resulted parts.



**Fig. 10.26** Machining process of the Al-6061-T6 block: **a** Roughing simulation in *VERICUT*<sup>®</sup>; **b** *VERICUT*<sup>®</sup> machine tool setup; **c** finishing simulation in *VERICUT*<sup>®</sup>; **d** finishing simulation in *VERICUT*<sup>®</sup>; **e** actual roughing operation; **f** running of finishing process; **g** resulting rough part; **h** resulting finish part

## 10.7 Conclusions and Future Perspectives

This chapter proposed a practical methodology of suitably determining machining strategies and cutting parameters to machine sculptured surfaces with the use of CAM systems. 3- and 5-axis machining setups were modelled and performed using Taguchi's DOE strategy. The main goal was to formulate the best manufacturing program (either 3- or 5-axis) for a sculptured part in terms of productivity and quality. Three quality criteria were determined; the total machining time (roughing and finishing) the volume remained after roughing and the surface deviation between the designed and the machined model. Selections for strategies and machining parameters were subjected to certain technological constraints in order to better represent actual industrial activities. To obtain results and to evaluate the whole process, statistical techniques and ANOVA were conducted.

The study's basis, falls under the assumption that in actual industrial machining operations, a series of processes are always required to machine parts. Although the study was conducted according to the "two stage scheme" to build, execute and verify manufacturing program scenarios, more complex manufacturing tool paths can also be investigated under the same experimentation concept by applying DOE approach to each machining phase; thus optimizing the entire NC program. In general, DOE approach applied to CAM software produces quite similar results to those obtained when conducting actual machining experiments. This indicates that tool path generation can be efficiently optimized regarding the quality objectives. Even though comparisons among the two NC machining technologies applied to sculptured surface machining in terms of efficiency indicated that 5-axis machining produces better results than 3-axis machining; the results obtained for 3-axis machining may also meet industrial demands thus it is strongly recommended when large investments need to be avoided.

As a further extension, DOE approach may be applied to build optimization meta-models to be later used as fitness functions for intelligent optimization systems. Within the broader field of designing experiments, response surface methodology (RSM) can be also applied to obtain and compare results. As a general outcome, it is evident that DOE approach indeed provides a low cost technique for near-optimum results as it facilitates NC programming and process planning stages for sculptured surface machining.

## References

1. Choi BK, Jerard RB (1998) Sculptured surface machining: theory and applications. Kluwer Academic Publishers, Boston
2. Yusoff AR, Hassan MF, Mansor MH (2012) Multiobjective optimization of milling tool geometry for chatter suppression and productivity improvement. *Adv Mater Res* 445(21):21–26
3. Zhang X, Zhu L, Zhang D, Ding H, Xiong Y (2012) Numerical robust optimization of spindle speed for milling process with uncertainties. *Int J Mach Tools Manuf* 61:9–19

4. Bagci E (2011) Monitoring and analysis of MRR-based feedrate optimization approach and effects of cutting conditions using acoustic sound pressure level in free-form surface milling. *Sci Res Essays* 6(2):256–277
5. Boujelbene M, Moisan A, Bouzid W, Torbaty S (2007) Variation cutting speed on the five axis milling. *J Achiev Mater Manuf Eng* 21(2):7–14
6. Corso LL, Zeilmann RP, Nicola GL, Missell FP, Gomes HM (2012) Using optimization procedures to minimize machining time while maintaining surface quality. *Int J Adv Manuf Technol*. doi:[10.1007/s00170-012-4288-8](https://doi.org/10.1007/s00170-012-4288-8)
7. Rakesh KP, Prazapati HR (2012) Parametric analysis of surface roughness (SR) and material removal rate (MRR) of hardened steel on CNC turning using ANOVA analysis: a review. *Int J Eng Sci Technol* 4(7):3111–3117
8. Saffar RJ, Razfar MR, Salimi AH, Khani MM (2009) Optimization of machining parameters to minimize tool deflection in the end milling operation using genetic algorithm. *World Appl Sci J* 6(1):64–69
9. Wu Q, Zhang YD, Gao XS, Fang L (2012) Investigation of tool deflection of solid carbide end mill in cutting process. *Appl Mech Mater* 163:95–99
10. Amin N-AKM, Sulaiman SA, Zainun SN, Arif MD (2012) Surface roughness optimization in end milling of stainless steel AISI 304 with uncoated WC-Co insert under magnetic field. *Adv Mater Res* 576:119–122
11. Abainia S, Bey M, Moussaoui N, Gouasmia S (2012) Prediction of milling forces by integrating a geometric and a mechanistic model. In: *Proceedings of the world congress on engineering-WCE 3*, London, U.K
12. Wojciechowski S, Twardowski P (2012) Tool life and process dynamics in high speed ball end milling of hardened steel. In: *5th CIRP conference on high performance cutting—Procedia CIRP* 1:289–294
13. Milfelner M, Kopac J, Cus F, Zuperl U (2006) Intelligent system for machining and optimization of 3D sculptured surfaces with ball-end milling. *J Achiev Mater Manuf Eng* 14(1–2):171–177
14. Gajate A, Bustillo A, Haber RE (2012) Transductive neurofuzzy-based torque control of a milling process: results of a case study. *Int J Innovat Comput Inf Cont* 8(5b):3495–3510
15. Ozkelik B, Oktem H, Kurtaran H (2005) Optimum surface roughness in end milling Inconel 718 by coupling neural network model and genetic algorithm. *Int J Adv Manuf Technol* 27:234–241
16. Tandon V, El-Mounayri H, Kishawy H (2002) NC end milling optimization using evolutionary computation. *Int J Mach Tools Manuf* 42:595–605
17. Arokiadass R, Palaniradja K, Alagumoorthi N (2012) Tool flank wear model and parametric optimization in end milling of metal matrix composite using carbide tool: response surface methodology approach. *Int J Ind Eng Comput* 3:511–518
18. Mustafa A (2011) Determination and optimization of the machining parameters of the welded areas in moulds. *Sci Res Essays* 6(2):485–492
19. Erturk A, Ozguven H, Budak E (2006) Analytical modeling of spindle –tool dynamics on machine tools using Timoshenko beam model and receptance coupling for the prediction of tool point FRF. *Int J Mach Tools Manuf* 46(15):1901–1912
20. Chuang C-M, Yau H-TA (2005) New approach to z-level contour machining of triangulated surface models using fillet endmills. *Comp Aided Des* 37:1039–1051
21. Chuang SH, Pan CC (1998) Rough cut tool path planning for B-spline surfaces using convex hull boxes. *Int J Adv Manuf Technol* 14(2):85–92
22. Xu X, Hinduja S (1998) Recognition of rough machining features in 2 1/2 components. *Comp Aided Des* 30(7):503–516
23. Yang X, Chen ZC (2005) A new high precision fitting approach for NURBS tool paths generation. *DETC2005: ASME Conf September 24–28:255*
24. Kim BH, Choi B (2000) Guide surface based tool path generation in 3-axis milling: an extension of guide plane method. *Comp Aided Des* 32:191–199

25. Quinsat Y, Sabourin L (2007) Optimal selection of machining direction for three-axis milling of sculptured parts. *Int J Adv Manuf Technol* 33(7–8):684–692
26. Tournier C, Lartigue, C (2008) 5-axis Iso-scallop Tool Paths along Parallel Planes. *Comp Aided Des & Appl* 5(1–4):278–286
27. Chang T-C, Wysk RA, Wang H-P (2006) *Computer-aided manufacturing*, 3rd edition, Pearson Prentice Hall.
28. Taguchi G, Chowdhury S, Wy Y (2005) *Taguchi's quality engineering handbook*. Wiley, Hoboken
29. Lin JC, Tai CC (1999) Accuracy optimization for mould surface profile milling. *Int J Adv Manuf Technol* 15:15–25
30. Cakir MC, Gürarda A (2000) Optimization of machining conditions for multi-tool milling operations. *Int J Prod Res* 38(15):3537–3552
31. Liu Y, Wang C (1999) A modified genetic algorithm based optimisation of milling parameters. *Int J Adv Manuf Technol* 15(11):796–799
32. Duffuaa SO, Shuaib AN, Alam M (1993) Evaluation of optimization methods for machining economic models. *Comput Oper Res* 20(2):227–237
33. Wang J (1998) Computer-aided economic optimization of the end-milling operations. *Int J Prod Econ* 54:307–320
34. Özen S, Bayhan GM (2011) Optimization of depth of cut in multi-pass machining using hopfield type neural networks. In: *Proceedings of the 2011 international conference on industrial engineering and operations management Kuala Lumpur, Malaysia*
35. Wu PH, Li YW, Chu CH (2008) Tool path planning for 5-axis flank milling based on dynamic programming techniques: advances in geometric modeling and processing. *Lect Notes Comput Sci* 4975:570–577
36. Deepak SSK (2012) A geometric programming based model for cost minimization of turning process with experimental validation. *Int J Eng Sci Emerg Technol* 3(1):81–89
37. Qiao LH, Hu QW, Zhang HW (2012) Machining sequencing optimization based on simulated annealing algorithm. *Adv Mater Res* 472–475:1632
38. Mohankumar G, Ganesan K, Kumar R (2011) Optimization of machining parameters in turning process using genetic algorithm and particle swarm optimization with experimental verification. *Int J Eng Sci Technol* 3(2):1091–1102
39. Wang ZG, Rahman M, Wong YS, Sun J (2005) Optimization of multi-pass milling using parallel genetic algorithm and parallel genetic simulated annealing. *Int J Mach Tools Manuf* 45:1726–1734
40. Onwubolu GC (2006) Performance-based optimization of multi-pass face milling operations using Tribes. *Int J Mach Tools Manuf* 46:717–727
41. Akturk MS, Ghosh JB, Kayan RK (2007) Scheduling with tool changes to minimize total completion time under controllable machining conditions. *Comput Oper Res* 34:2130–2146
42. Tansel IN, Ozelik B, Bao WY, Chen P, Rincon D, Yang SY, Yenilmez A (2006) Selection of optimal cutting conditions by using GONNS. *Int J Mach Tools Manuf* 46:26–35
43. Belloufi A, Assas M, Rezgui I (2012) Optimization of Cutting Conditions in Multi-Pass Turning Using Hybrid Genetic Algorithm-Sequential Quadratic Programming. *J Appl Mech Eng* 1(1):1–5
44. Juan H, Yu SF, Lee BY (2003) The optimal cutting-parameter selection of production cost in HSM for SKD61 tool steels. *Int J Mach Tools Manuf* 43:679–686
45. Fountas N, Krimpenis A, Vaxevanidis NM, Davim JP (2012) Single and multi-objective optimization methodologies in CNC machining. In: Davim JP (ed) *Statistical and computational methods in manufacturing*. Springer, Heidelberg
46. CG Tech VERICUT® documentation and user's manual.

# Chapter 11

## Modeling of Micromachining

Angelos P. Markopoulos and Dimitrios E. Manolakos

**Abstract** Micromachining emerges as a key technology for the future of manufacturing due to its many applications in miniaturization but also for macro scale components with features in the micro world, e.g. microgrooves or microholes. However, phenomena in the microscopic level are quite different from the ones encountered in traditional machining and thus the analysis of micromachining raises difficulties. Since the use of experiments is costly, difficult to perform and it is hard to measure parameters at this level, modeling is considered the best alternative for performing the required analysis. In this chapter a review of the modeling methods used in micromachining is provided. Most of the chapter is dedicated to mechanical micromachining and its modeling and simulation via finite elements; this is because this kind of micromachining exhibits the most differences in the microworld with size effect and minimum chip thickness being topics of great interest and because this method has proven to be the favorite for many research groups worldwide. The chapter closes with a brief discussion on other modeling methods and micromachining processes and a wide reference list for all the topics is included.

### 11.1 Introduction

The demand for components that possess features in the micro- or nanometer regime has increased steadily over the past years. This is mainly due to the fact that such components find use in a wide range of applications. The IT related components is the leading and perhaps most important sector that incorporates these parts, while other sectors such as health and biomedicine, automotive

---

A. P. Markopoulos (✉) · D. E. Manolakos  
Laboratory of Manufacturing Technology, National Technical University of Athens,  
9, Iroon Polytechniou Avenue, 15780 Athens, Greece  
e-mail: amark@mail.ntua.gr

industry and telecommunications follow closely. The applications are microfluidics, pumps and valves, micronozzles, optical components, micromolds and microholes on various materials just to name some. At the same time, the integration of microsystems in more applications of the aforementioned areas of interest has given a boost to micro and nanomanufacturing and thus has intensified the research pertaining to MEMS and NEMS, advanced technology, increased performance and decreased cost.

For the production of microcomponents advanced manufacturing techniques are required. In the electronics industry, for the production of integrated circuits and computer chips, manufacturing methods are used to transfer patterns on wafers. This can be performed both by additive (deposition) and subtractive (etching) methods. Silicon based products are mainly produced with lithographic processes, the main one being photolithography. Over the last years Next Generation Lithographies (NGL) are developed in order to fulfill the quest for smaller size and yet larger capacity. NGL include X-ray Lithography, Extreme UV Lithography, LIGA, SCALPEL, Ion Beam Lithography and other manufacturing methods for silicon materials. Technological developments in this area are driven by the needs of electronics industry, achieving fast growth. However, most of these processes are planar or 2½ D and have limited application to materials other than silicon.

Technologies for processing other materials from a few  $\mu\text{m}$  to a few hundreds of  $\mu\text{m}$ , namely metals, ceramics and polymers, are in use. These micromachining processes are abrasive, conventional or non-conventional machining processes. By the term machining, processes that shape parts by removing unwanted material, are described. Unwanted material is carried away from the workpiece usually in the form of a chip; evaporation or ablation may take place in some machining operations. The more narrow term cutting is used to describe the formation of a chip via the interaction of a tool in the form of a wedge with the surface of the workpiece, given that there is a relative movement between them. These machining operations include turning, milling, drilling and boring among others and are usually referred as traditional machining processes and in the micro level as microcutting or mechanical micromachining processes. Abrasive processes such as grinding are also part of cutting processes of great importance in contemporary industry. Other non-traditional machining operations that may or may not include physical contact between cutting tool and workpiece or may not have a cutting tool in the same sense as traditional processes or utilize thermal or chemical energy for removing material from workpiece, are ultrasonic machining, water jet machining, electrodischarge machining, laser machining and electrochemical machining among others.

Today, these manufacturing methods are at the forefront of industrial integration and their applications have reached a high level of production maturity. Some of them have the prefix “micro” in their names to declare that they are processes following the same principles as the original macroscopic ones, but particularly designed as microprocesses, e.g. micromilling, microEDM etc. Micromachining processes are more advantageous compared to the other categories since it is possible to machine a variety of materials in complex shapes, with excellent



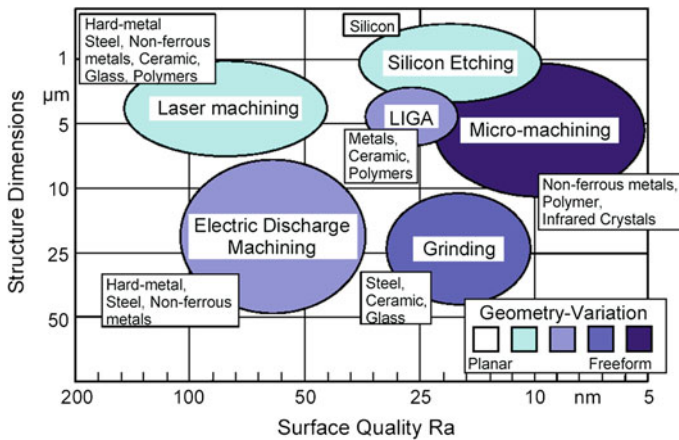


Fig. 11.1 Capabilities comparison of various microfabrication techniques [1]

surface finish and tight tolerances. In Fig. 11.1 the structure dimensions versus the surface quality attained by various microfabrication techniques is presented. In the same figure the geometry variation achieved as well as the materials that can be processed by each technique are depicted. The obvious advantages of mechanical micromachining processes place them in a predominant position for the construction of components used in various technology sectors. It is worth noting, however, that micromachining may also refer to ultraprecision process performed on large scale components with some features similar to those encountered in miniaturized parts; some of the topics discussed hereafter have their origins in ultraprecision machining. The following paragraphs focus on, but are not limited to, mechanical micromachining; mechanical micromachining has been extensively studied theoretically and experimentally and a vast number of modeling and simulation papers have been published. Reviews of micromachining and lithography-based processes can be found in [1–7]. Within these references, information on almost all micro and nanoprocesses utilized in modern practice can be found. Furthermore, discussion on the advantages and disadvantages of the processes, applications and special features is provided.

Although most non-conventional micromachining processes use the same principles in the macro and the micro realm, the same cannot be said for mechanical micromachining processes. The prefix “micro” in this case is of importance since a downscaling of cutting to mechanical micromachining is rather complex. The small dimensions of workpieces, cutting tools, depths of cut etc, bring up a number of issues that may play no significant role in traditional machining but are of grave importance in micromachining. The underlying mechanisms of chip formation are influenced by features known as the minimum chip thickness and the size effect. Thus limitations arise because of the size. Experimental work is not always feasible to be carried out in order to overcome

the problems of manufacturing components in the micro and nanoscale. Moreover, increased demand, innovation, reliability and cost reduction need to be satisfied. Modeling and simulation techniques have been developed in aid of engineers and scientists who use them in order to reduce experimental time and testing, give insight on complex phenomena, explore possibilities, reduce complexity and learning cycles of a process, increase accuracy and optimize processes and products.

In engineering the use of modeling methods for the development of a product is common practice. The initial objective of studying and modeling metal machining was to provide a theory which, without any experimental work, would enable researchers to predict cutting performance and thus solve practical problems confronted in industry. The first analytical models set the basis for more advanced methods developed later in the course of time and when the tools for realistic computational cost and analysis time became available with computer advances. Analytical models supported by metal cutting mechanics and with simplifying assumptions began publishing around 1900s. However, it was not until the 1950s that modeling of machining became a key tool used for understanding the mechanisms of material removing process as well as predicting their performance.

In the early 1970s some pioneering works on machining modeling with the Finite Element Method (FEM) began to find their way in scientific journals. Over the years and with the increase of computer power as well as the existence of commercial FEM software, this method has proved to be the favorite modeling tool for researchers of the field. This is established by the vast number of publications on this subject as well as the modeling novelties introduced and used, even by the fact that software dedicated solely for the purpose of modeling machining and lately micromachining operations exist. Finite element models are used today for gaining knowledge on fundamental aspect of material removing mechanisms but more importantly for their ability to predict important parameters such as cutting forces, temperatures, stresses etc. essential for the prediction of the process outcome, the quality of the final product and in a timely and inexpensive way.

It can be stated that the same practice is employed in the manufacturing of micro and nanoproducts, where a lot of work has already been carried out on modeling and simulation. Analytical modeling, often encompassing computer simulations or commercial software programs specializing in small size manufacturing can be found, today. Many engineers and scientists have chosen this approach to conduct their research or aid their production. The techniques used are either the same ones used in traditional machining methods, e.g. the Finite Element Method (FEM), suitably altered so that the particularities of micro and nanoscale phenomena are taken into consideration or methods used only in this area of technology such as the Molecular Dynamics (MD) method. The present paper is mainly concerned with the description of the modeling and simulation techniques used in mechanical micromachining and the benefits deriving from their application. The main principles of the techniques employed in modeling and simulation at micro and nanoscale, the fields of application, limitations, considerations and an up-to-date bibliography are provided within this chapter.

In the following paragraphs an overview of the mechanics of cutting will be provided and then the phenomena appearing in mechanical micromachining will be further discussed. This way the peculiarities of micro-scale machining will be pointed out and explored. Then, an overview of the FEM in machining will be provided and extended to mechanical micromachining. In these paragraphs not all the subjects are directly focused in micromachining, nevertheless, are important for understanding and calling attention to mechanical micromachining modeling and simulation. The last paragraph of this chapter is dedicated to other modeling techniques, such as MD for nanomachining, and the modeling of other micromachining processes, such as laser machining.

## 11.2 Mechanics of Machining and Micromachining

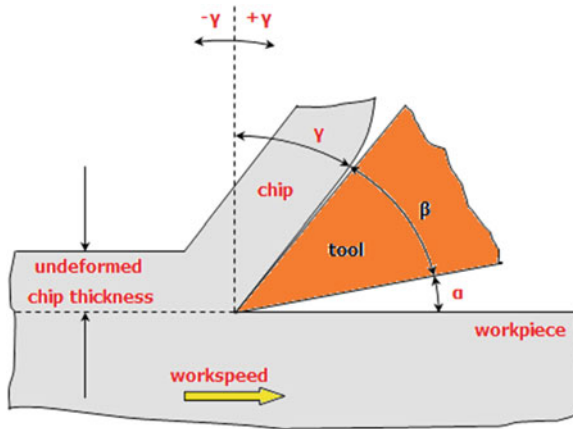
It is useful at this point to provide a few descriptions on the mechanics of cutting and then elaborate on some of the differences as compared to the mechanics of mechanical micromachining. These differences are important and are taken into account in the modeling procedure but are also used to explain various phenomena appearing in the micro-scale.

### 11.2.1 Machining Modeling

The chip flow in all wedged-tool machining processes can be described, in theory, in a common way by two different cutting schemes termed orthogonal cutting and oblique cutting. In orthogonal cutting, see Fig. 11.2, the cutting edge of the tool is perpendicular to the direction of relative workpiece-cutting tool motion and also to the side face of the workpiece. From the relative movement of workpiece and cutting tool, a layer of material in the form of chip is removed. Perpendicular to feed is the depth of cut, which is smaller than or equal to the width of the tool edge. The surface along which the chip flows is the rake face of the tool. The angle between the rake face and a line perpendicular to the machined surface is called rake angle  $\gamma$ , the face of the tool that is near the machined surface of the workpiece is the flank face and the angle between the flank face of the tool and the workpiece is called clearance angle  $\alpha$ . Finally, the angle between the rake face and the flank face is the wedge angle  $\beta$ . The sum of the three angles is always equal to  $90^\circ$ , thus:

$$\alpha + \beta + \gamma = 90^\circ \quad (11.1)$$

In Fig. 11.2 a positive rake angle is shown and the direction for a positive and a negative rake angle can also be seen. For negative rake angles, the tools possess a wider wedge angle. A positive rake angle is used for ductile materials since a “weaker” tool, with smaller wedge angle, will suffice to perform the cutting

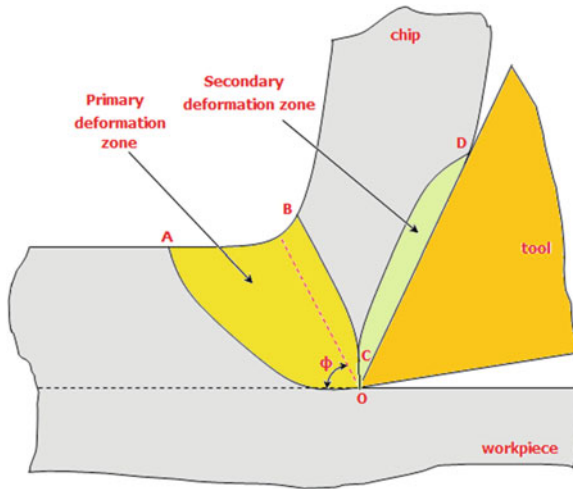


**Fig. 11.2** Orthogonal cutting

operation. For high-strength materials, rake angle is chosen to be negative, thereby increasing the wedge angle and creating a stronger cutting edge. However, stronger cutting edge has the disadvantage of requiring greater power consumption and needing a robust tool-workpiece set-up to compensate for the vibrations. The flank face of the tool does not participate in chip removal; it ensures that the tool does not rub on the newly machined surface and affects its quality. However, the clearance angle affects the cutting tool wear rate. If the tool's clearance is too large it will weaken the wedge angle of the tool, whereas if too small, it will tend to rub on the machined surface.

Orthogonal cutting represents a two-dimensional mechanical problem with no side curling of the chip considered. It represents only a small fragment of machining processes, i.e. planning or end turning of a thin-walled tube. However, it is widely used in theoretical and experimental work due to its simplicity. Because of its 2D nature many independent variables are eliminated, e.g. two cutting forces are only identified to orthogonal cutting problems. On the other hand, oblique cutting, where the cutting tool is inclined corresponds to a three-dimensional problem with more realistic chip flow representation but more complex analysis, i.e. three force components are present and chip curling is accounted for.

There are two deformation areas distinguished in machining, namely the primary and the secondary deformation zones, see Fig. 11.3 the deformation zones thickness, chip thickness and shear angle are not depicted in any scale in this Figure, only the locations are roughly indicated. The primary deformation zone is included in the OAB area. The workpiece material crossing the OA border undergoes large deformation at high strain rates and exits the zone at OB border, work hardened. It is determined by microscopic examination and experiments that chips are produced by shear within this region. Most of the experimental studies conclude that this zone is of average thickness of about one tenth of chip thickness [8]. The secondary deformation zone is included in OCD. Along the OD, the contact length



**Fig. 11.3** Primary and secondary deformation zone and shear plane angle

between the rake face of the tool and the chip, the material is deformed due to intensive interfacial friction. The secondary deformation zone is characterized by two regions, the sticking region, closer to the cutting tool tip and the sliding region, above the previous one [9]. In the sticking region, material adheres to the tool and as a result shear within the chip is observed. Both deformation zones are characterized by temperature rise due to severe plastic deformation in the primary and due to friction in the secondary deformation zone. Furthermore, high cutting speeds do not allow for heat conduction to take place and heat is concentrated at a small area around the cutting tool edge. Strain hardening due to deformation and softening due to temperature alter the chip formation characteristics in every step of its formation. The friction coefficient is very hard to be measured in the secondary deformation zone and several theories have been proposed for the calculation of friction [10].

A simplified approach proposes that shearing in the primary deformation zone takes place along a shear plane, characterized by shear angle  $\phi$ , between the shear plane and the workpiece surface. Although this single shear plane model is criticized, it is usually referred in machining handbooks due to its simplicity and it is the basis for calculating several process parameters. In any case, it is imperative to estimate shear angle and friction parameters in order to calculate cutting forces through analytical or numerical modeling. Analytical models, only briefly described here, are considered the predecessors of numerical models. This is by no way meant to say that numerical models substituted analytical modeling, since a lot of researchers still are working on this subject and the value of these models is paramount. It is meant to say that they have the same origins and form the basis on which e.g. FEM models and simulations are made.



The resultant force  $F$  is resolved in components  $F_N$  and  $F_F$  that are normal to the tool face the former and along the tool face the latter. It is also resolved to  $F_{SN}$  and  $F_S$  that are normal to and along the shear plane respectively. Finally, it can also be resolved into components  $F_c$ , the cutting force, and  $F_t$  the feed or thrust force. Furthermore, the rake angle  $\gamma$ , the shear angle  $\phi$  and the mean angle of friction between chip and tool  $\rho$  are shown. The friction angle  $\rho$  is related to the friction coefficient  $\mu$  through equation:

$$\rho = \arctan(\mu) = \arctan(F_F/F_N) \quad (11.2)$$

According to Ernst and Merchant's theory, an upper bound one, a shear angle needs to be found that the cutting work will reduce to a minimum. In other words, since the work is proportional to the cutting force  $F_c$ , an expression of the cutting force with the shear angle needs to be found and then obtain the  $\phi$  for which  $F_c$  is a minimum. From Fig. 11.4, it can easily be concluded that:

$$F_S = F \cos(\phi + \rho - \gamma) \quad (11.3)$$

Furthermore, the same force component can be calculated in relation to the shear strength of the workpiece material on the shear plane  $\tau_S$ , the cross-sectional area of the shear plane  $A_S$  and the cross-sectional area of the undeformed chip  $A_C$ , via the following equation:

$$F_S = \tau_S A_S = \frac{\tau_S A_C}{\sin \phi} \quad (11.4)$$

Thus from Eqs. 11.3 and 11.4 it is:

$$F = \frac{\tau_S A_C}{\sin \phi} \cdot \frac{1}{\cos(\phi + \rho - \gamma)} \quad (11.5)$$

Geometrically it is deduced that:

$$F_c = F \cos(\rho - \gamma) \quad (11.6)$$

Combining Eqs. 11.5 and 11.6 it may be concluded that:

$$F_c = \frac{\tau_S A_C}{\sin \phi} \cdot \frac{\cos(\rho - \gamma)}{\cos(\phi + \rho - \gamma)} \quad (11.7)$$

If the last equation is differentiated with respect to  $\phi$  and equated to zero, it is possible to calculate a shear angle for which the cutting force is minimum. The equation is:

$$2\phi + \rho - \gamma = \pi/2 \quad (11.8)$$

This equation agreed poorly with experimental results of metal machining. Merchant attempted an alternative solution [12]. When Eq. 11.7 was differentiated it was assumed that  $A_c$ ,  $\gamma$  and  $\tau_S$  were independent of  $\phi$ . In the new theory, deformation and friction are reflected through a change of the force acting in the

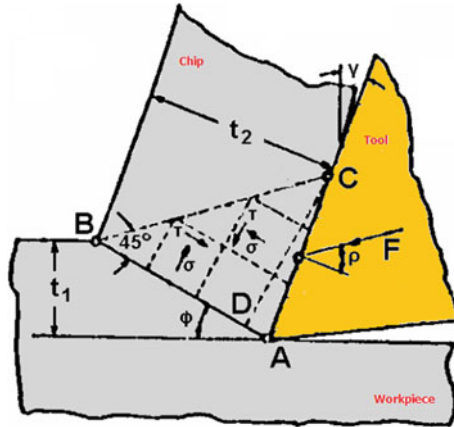


Fig. 11.5 Lee and Shaffer's slip-line field theory for orthogonal cutting

direction perpendicular to the plane of shear, thus the normal stress  $\sigma_S$  of the shear plane affects the shear stress  $\tau_S$ . In the modified analysis a new relation is included:

$$\tau_S = \tau_o + k\sigma_S \tag{11.9}$$

This relation is known as the Bridgman relation and  $k$  is the slope of the  $\tau - \sigma$  relation; the shear stress increases linearly with an increase in normal strength and the lines intersects the shear stress axis at  $\tau_o$ . With this revised theory the new result for shear angle is:

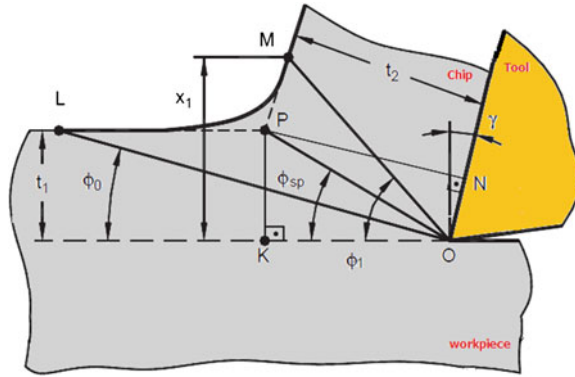
$$2\phi + \rho - \gamma = C \tag{11.10}$$

$C$  is a constant that depends on the workpiece material.

Lee and Shaffer's work was the first contribution of the slip-line field models of chip formation [13]. It was the result of applying simplified plasticity analysis to metal cutting, more specifically to orthogonal cutting with continuous chip. It was assumed that in this plane strain conditions, the workpiece material is rigid perfectly plastic, i.e. the elastic strain is neglected during deformation and once the yielding point is exceeded deformation takes place at constant stress for varying strains, strain rates and temperatures. The constructed slip-line field is shown in Fig. 11.5.

In this lower bound solution all deformations take place in a stress field bounded by rigid bodies; this stress field transmits the cutting forces from the shear plane to the chip resulting in the triangular plastic zone ABC. In this region no deformation occurs but the material is stressed to its yield point, so that the maximum shear stress is the shear stress on the shear plane. The two directions of the maximum shear stress are indicated by the slip-lines. The shear plane AB is the one set of slip lines because the maximum shear stress must occur along the shear plane. Furthermore, BC can be regarded a free surface since no forces act on the chip after BC, stresses cannot be transmitted from there. Thus, according to the second rule mentioned above, ABC is equal to  $\pi/4$ . Assuming that stresses act





**Fig. 11.6** Zorev's simplified model

uniformly at the chip-tool interface, normal stresses will meet the boundary at angles  $\rho$  and  $\rho + \pi/2$ . Maximum shear stresses are  $\pi/4$  to the direction of normal stresses and thus  $ACB$  is  $(\pi/4) - \rho$ . The shear angle can be calculated by equation:

$$\phi + \rho - \gamma = \pi/4 \tag{11.11}$$

It is evident that when the mean angle of friction between chip and tool is  $\pi/4$  and the rake angle is zero, shear plane angle is also zero, which is not possible. Lee and Shaffer proposed a solution for this case of high friction and low rake angle, assuming built-up edge formation. The slip-line theory was also used by other researchers who suggested curved  $AB$  and  $CD$  boundaries [14, 15].

Zorev proposed a qualitative model for which no solution was provided [9]. However, a simplified form was proposed where it is assumed that no shearing occurs along the shear lines adjacent to the tool rake face, see Fig. 11.6. By using geometrical relationships a generalized solution is derived as:

$$2\phi_{sp} + \rho - \gamma \approx (\pi/2) - \psi_{sp} \tag{11.12}$$

In this equation the  $\phi_{sp}$ , the specific shear angle is introduced and  $\psi_{sp}$  is the angle of inclination of the tangent to the outer boundary of the plastic zone. The interesting about this solution is that if various values of  $\psi_{sp}$  are substituted, the shear angle relations by other researchers are derived, i.e. for  $\psi_{sp}$  equal to zero, representing the single shear plane model, the Ernst and Merchant solution is obtained, for  $\psi_{sp} = C_1$  and  $C = (\pi/2) - C_1$  the modified Merchant solution is obtained and for  $\psi_{sp} = \rho - \gamma$  the Lee and Shaffer solution is derived.

Most shear plane models assume that shear stress on the shear plane is uniform, no strain hardening is considered and that friction along the cutting tool-chip interface is characterized by a constant friction coefficient; this last assumption is in contradiction with experimental data. If it is assumed that deformation takes place in a narrow band centered on the shear plane, more general material assumptions can be used. The effects of yield stress varying with strain and

sometimes with strain rate and temperature were considered and simplification of the equilibrium and flow was achieved. Pioneering work in this area is associated with the work of Oxley. Based on experimental data, where the plastic flow patterns are observed, it is assumed that the shear zone thickness is about one tenth of the shear zone length. The shear zone models are an obvious improvement over the preceding models. Many additions to the first model proposed by Oxley have been reported. A full account of these developments would be out of the scope of this work; a detailed description of Oxley's works is given in [16].

There are more than 50 shear angle solutions identified in the relevant literature as it is reported in [17] and it would be impossible to explore all of them within this book chapter. It should be noted that the single-shear plane model has been criticized over the years and experimental data do not correlate with the theory results. Astakhov [18] summarized the major inherent drawbacks of the single-shear plane model as being the infinite strain rate, the unrealistic high shear strain that is in contradiction with material testing results, the rigid perfectly plastic workpiece material assumption, the improper accounting for the resistance of the processed workpiece material, the perfectly sharp cutting edge of the tool and the fact that there is no contact on the tool flank surface that are not realistic for common practice and the inapplicability of the model in brittle material machining. Furthermore, for the Ernst and Merchant theory, drawbacks include the incorrect velocity and force diagrams presented and the assumption of constant friction coefficient. However, this model is still in use by researchers due to its simplicity.

The analyses already presented pertained only to orthogonal cutting with continuous chip. However, the shear plane model has been extended to three dimensions [19] and the slip-line model has been proposed for oblique cutting [20]. A three-dimensional analysis similar to the work of Oxley has been presented by Usui [21–23], which includes secondary cutting edge and nose radius effects; the results apply to turning, milling and groove cutting. However, both Oxley's and Usui's models are quite complex and for their application stress and strain data at the strain rates and temperatures encountered in metal machining are needed. The lack of these data is a significant drawback. These are the reasons that these models, although more complete than all the others since they include temperature effects and can be used in tool wear and segmented chip formation modeling and are in agreement with experimental data, are not widely used outside the research groups that they developed them.

### ***11.2.3 Mechanical Micromachining***

The set-up of mechanical micromachining is quite similar to the one used in the macro-scale traditional cutting processes, i.e. a wedge-like tool is removing material from a surface. All the geometrical features and the kinematic characteristics of the tool and the workpiece can be identified. However, downscaling all the phenomena in order to apply the same theories in both micro and macro regime

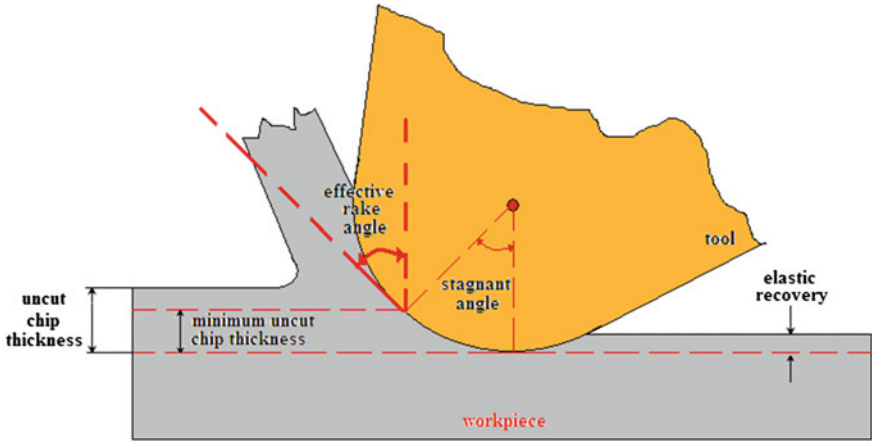


Fig. 11.7 Orthogonal microcutting

has proven to be inadequate. There are features of machining and phenomena that are considerably different in micromachining and do not allow for such a simplification; differences arise when considering the chip formation process, the resulting cutting forces, the surface integrity and tool life.

In Fig. 11.7 the orthogonal cutting model corresponding to microcutting may be observed. At this level, Merchant’s model seems unrealistic. In micro-cutting operations the depth of cut may be below 10  $\mu\text{m}$  and the anticipated surface roughness only a few nm. The cutting edge can no longer be considered sharp; the cutting edge radius is comparable in size to the uncut chip thickness. Note, also that, perfectly sharp cutting tools are practically impossible to be produced due to technological limitations connected to the fabrication of the tools. The rake angle of the tool, which in Fig. 11.7 is positive, is not the actual rake angle participating in the processes. As it can be seen, the effective rake angle is negative. In this case, the elastic-plastic deformation of the workpiece material and the ploughing need to be taken into account, as well as the elastic recovery at the clearance face.

From the above, the existence of a minimum chip thickness that can be removed from the workpiece surface in a mechanical micromachining operation can be deduced. A stagnation point above which a chip is formed and below only elastic-plastic deformation takes place is assumed. The stagnation point is connected to a stagnant angle  $\theta_m$ , which with the tool edge radius determines the value of the minimum uncut chip thickness,  $h_m$  [24]:

$$h_m = r_e(1 - \cos\theta_m) \tag{11.13}$$

The minimum chip thickness determines whether a chip is formed or not because if the depth of cut for a microcutting operation is set below this minimum, then the cutting edge is expected to just plastically deform the workpiece material without producing a chip. This is the ploughing mechanism which except the

obvious effect on the surface integrity and the quality of the finished workpiece alters significantly the cutting forces and thus the process stability in mechanical micromachining and makes the force prediction methods described in the next paragraph ineffective. The minimum chip thickness was identified by Finnie in grinding [25] but several other researchers have performed experimental work to prove its existence in microturning and micromilling. Ikawa et al. used a turning diamond tool with an edge radius of about 10 nm and found that the uncut chip thickness is in the order of 1/10 of the cutting edge radius [26]. Weule et al. studied the minimum chip thickness for micromilling. In this case, during the same pass of a single tooth of the cutter from the workpiece and due to the varying thickness of the chip, the material removal mechanism may change from shearing to ploughing or vice versa, resulting in a saw tooth-like surface profile and deteriorating surface finish [27]. The minimum chip thickness was also reported in micromilling by Kim et al. by comparing the chip volume of the process with the nominal chip volume and the feed marks on the workpiece surface with the feed per tooth. It was concluded that chip was not formed in every pass of a cutting tooth and this was attributed to the minimum chip thickness [28].

Although ploughing may exist in machining, its effect on the overall process may be neglected. However, this effect caused by cutting edge radius is important in micromachining. Many researchers consider this mechanism to be the main reason for the size effect, the non-linear increase in the specific energy and thus in the specific cutting force with decreasing depth of cut, which is observed in micro-cutting. Albrecht argued that there are two areas where ploughing occurs, one on the rake face and another around the tool edge [29, 30]. Masuko introduced a new effect that acts independently from ploughing and called it indenting. According to this theory an indenting force causes the cutting edge to penetrate the workpiece; this force is held responsible for the size effect according to this theory [31].

The size effect was identified in metal cutting operations as early as 1952 [32]. The researchers processed specimens made of SAE 1112 steel with grinding, micromilling, turning and tension test; the processes are listed with increasing size of specimen processed. In Fig. 11.8 the chip thickness versus the resisting shear stress clearly shows the size effect, which was attributed to the significantly reduced amount of imperfections, namely crystallographic defects such as grain boundaries, missing and impurity atoms and inhomogeneities present in all commercial metals, encountered when deformation takes place in a small volume. With smaller uncut chip thickness the material strength is expected to reach its theoretical value of strength.

Many more investigators have acknowledged the size effect experimentally and theoretically. In the following Fig. 11.9a and b the specific cutting force and the specific thrust force can be viewed, respectively. In these Figures the results from many micromachining experiments are gathered and the size effect is obvious, for depth of cut lower than 10  $\mu\text{m}$ . In the same Figures, the results of two models, one for round and one for sharp edge tool, are also depicted; it may be concluded that

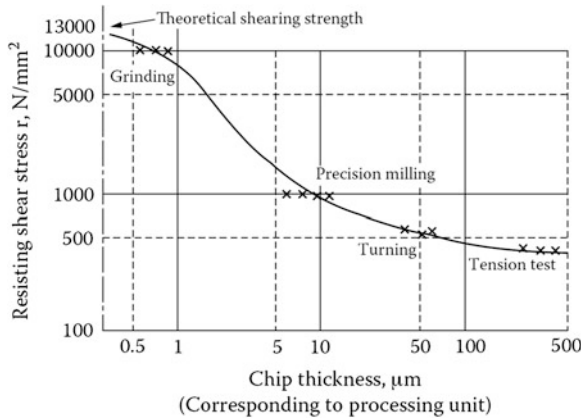


Fig. 11.8 Relation between chip thickness and resisting shear stress (data from [32] modified to include the tension test results) [33]

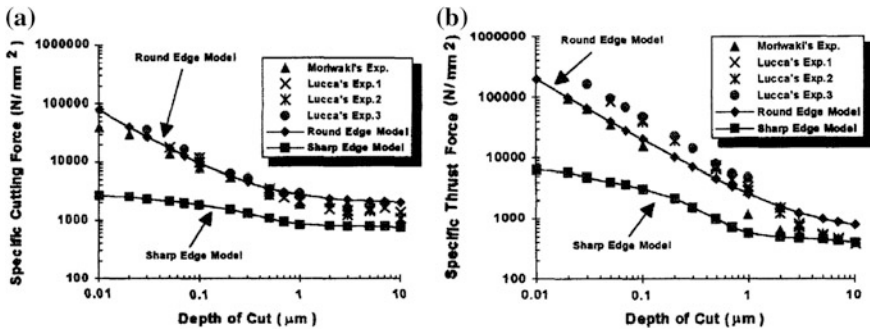


Fig. 11.9 a Specific cutting force and b specific thrust force for various experiments and models indicating the size effect in micromachining [34]

although the size effect is present when a sharp tool is used, it is more intense with the round edge tool.

Although size effect is present in metal cutting, like minimum chip thickness, it is of special importance when it pertains to micromachining. Besides the explanations already mentioned, there are several other discussions on the reasons for the appearance of the size effect. It is also attributed to material strengthening due to an increase in the strain rate in the primary shear zone [35] or the decrease of temperature in the tool-chip interface [36] with decreased chip thickness. Atkins proposed that the size effect is due to the energy required for new surface creation via ductile fracture [37]. Another explanation is based on the size effect appearing in micro-nano-indentation and its extension to machining [38]. The increased hardness of a material with reduced indentation depth is a result of the dependence of material flow stress on the strain gradient in the deformation zone; strain

gradient plasticity can be the reason of size effect in machining because of the intense strain gradients observed. The criticism to this theory pertains to the fact that the size effect in hardness is related to that in cutting when the von Mises criterion is applicable and this assumption is not compatible with the experiments of Merchant [39]. However, some works with analytical and FEM models and experimental validation based on strain gradient plasticity have been published [40, 41].

From the literature review it is evident that many reasons for the size effect in machining and micromachining have been reported. It is not clear which of the above mechanisms is dominant or whether there could be more than one mechanism acting at the same time. Even in the case of multiple mechanisms acting together, there may be factors that alter the contribution of each factor in each case. For further reading on the size effect in machining and especially in micromachining Refs [39, 42, 43] are suggested.

### 11.3 FEM Modeling

In this paragraph the aspects of modeling machining with FEM are explored. The special features required in mechanical micromachining, when needed, are also discussed in this section. The previous section has pointed out the difficulties that are associated with modeling machining processes. First of all, the strain rates observed are very high; this holds true for even low cutting speeds. Furthermore, the plastic deformation takes place in small regions, the primary and secondary deformation zones, around the cutting edge, making difficult the selection of the appropriate boundary conditions. Furthermore, the ploughing-shearing mechanism of micromachining further complicates the matter. There is not a unified and generally accepted theory pertaining to the exact chip formation mechanism, mainly due to the phenomena taking place in the deformed regions. In many analytical models that are proposed, the strain hardening of the workpiece material is not included in the analysis, although it plays a significant role, as is concluded from experimental results. Additionally, the temperature rise in the region due to plastic deformation and friction induce material softening and alter the workpiece material properties in relation to strain rates and temperatures. Data for the workpiece material for varying temperature and strain rate at the levels which occur in metal machining are not easily found in the literature. On top of this non-linear situation, the temperature rise needs to be taken into account to the various calculations performed, which means that besides the mechanical problem, a heat transfer problem must be dealt with simultaneously. Finally, the grain size of the workpiece material, which is comparable to depth of cut, needs to be addressed.

Considering the above, finite elements appear to be the most suitable method for modeling cutting and mechanical micromachining. Due to its inherent characteristics it can solve non-linear problems and with advances in computers and the use of commercial software it can readily perform coupled thermo-mechanical

analysis. This kind of numerical modeling has already been used with success in many scientific and technological areas, modeling of manufacturing processes being one of them [44–46]. Still, chip formation is difficult to be modeled. Except the physical phenomena explained above two more challenges need to be addressed. The first one is to provide accurate data to the model; this is common sense however it can be problematic. The second is to actually choose a finite elements method, meaning that there are different approaches or strategies proposed for metal machining modeling with FEM pertaining to formulation, treatment of friction, material behavior, iteration scheme etc. used for approximating a solution; the combinations that have already been tried by researchers are numerous.

In FEM the basic principle is the replacement of a continuum by finite elements forming a mesh; this procedure is called discretization. Each finite element is simpler in geometry and therefore easier to analyze than the actual structure. Every finite element possesses nodes where the problem initial and boundary conditions are applied and the degrees of freedom are calculated; the finite elements are connected to one another in nodes. Between the nodes, problem variables are derived by interpolation. The problem variables as well as properties applied on the nodes of each element are assembled and global relations are formatted. Usually, the analysis involves a great number of algebraic equations to determine nodal degrees of freedom and that is why a personal computer is employed for processing.

There are two different time integration strategies in order to face non-linear and dynamic models, namely implicit and explicit schemes. The explicit approach determines the solution of the set of finite element equations by using a central difference rule to integrate the equations of motion through time. The equations are reformulated and they can be solved directly to determine the solution at the end of the increment, without iteration. The method is dynamic; it uses a mass matrix and computes the change in displacements from acceleration. On the other hand, the implicit method is realized by solving the set of finite element equations, performing iterations until a convergence criterion is satisfied for each increment. The length of the time step is imposed by accuracy requirements. In the implicit method the state of a finite element model at time  $(t + \Delta t)$  is determined based on data at time  $(t + \Delta t)$ , while the explicit method solves the equations for  $(t + \Delta t)$  based on data at time  $t$ . Both implicit and explicit methods have been used in cutting simulation [47, 48]. There are some papers that elaborate on the use of implicit or explicit techniques that give more information on the matter, with examples including manufacturing processes [49–51].

Another topic pertains to the use of a certain numerical formulation. The ones used in metal cutting FEM models are so far of three types, namely Eulerian, Lagrangian and the newer Arbitrary Lagrangian-Eulerian (ALE) analysis. In the Eulerian approach the finite element mesh is spatially fixed and covers a control volume. The material flows through it in order to simulate the chip formation. This implies that the shape of the chip, shear angle and the contact conditions must be a priori known, derived from experiments, or assumed. An iterative procedure is

used for the convergence of variables and chip geometry is updated. The element sides that are the boundaries of the chip that are adjacent to the rake face and far from the rake face of the tool are repositioned to be tangential to the cutting position. However, strains are derived from the integration of strain-rates along stream lines; this cannot be used for the simulation of discontinuous chips.

In the Lagrangian approach the elements are attached to the material. The material is deformed due to the action of the cutting tool and so is the mesh. This way there is formation of the chip due to deformation from the tool. Unconstrained material flow in Lagrangian formulation allows for simulations from incipient chip formation to steady-state conditions and modeling of segmented chips besides the continuous one. In an explicit approach the displacement of the workpiece and the attached mesh, is a function of the time step and can be related to the material removal rate; in an implicit formulation the size of the time step has no influence on the stability of the solution. Furthermore, several models that depend on strain, strain rate and temperature have been applied for the workpiece material.

A disadvantage of the Lagrange formulation is connected to the large mesh deformation observed during the simulation. Due to the attachment of the mesh on the workpiece material, the mesh is distorted because of the plastic deformation in the cutting zone. Such severe distortions of the mesh may result in the failure of the model as they cannot be handled by the elements applied in the mesh. Pre-distorted meshes [52] and re-meshing techniques are applied in order to overcome these problems [53]. Furthermore, for the formation of the chip, a chip separation criterion in front of the tool edge is applied. This procedure can be quite thorny; it has been the topic of several papers and no generally accepted criterion is adopted. The latest development in the Lagrangian formulation, an updated Lagrangian analysis, has overcome the disadvantage of a chip separation criterion by applying continuous re-meshing and adaptive meshing, dealing at the same time with the mesh distortion; the above are thoroughly discussed in forthcoming paragraphs.

Summarizing a comparison between Eulerian and Lagrangian techniques it can be stated that the Eulerian formulation needs no re-meshing since there is no element distortion involved in the analysis and requires no chip separation criterion because the course of the chip is predetermined. The computational time in such models is reduced due to the few elements required for modeling the workpiece and the chip and simple procedures are used in the relative software. This analysis is suitable for the simulation of steady-state cutting, when the incipient stages of chip formation are not of interest and with continuous chip since no chip breakage criterion can be incorporated in the model for the simulation of discontinuous chip formation. This technique was used in the past, mainly in the first FEM models that appeared for metal cutting and although it is still used today, it is considered that it does not correspond to the real deformation procedure encountered in real metal cutting processes, as the chip thickness, a major outcome of the process cannot be assumed physically [54]. On the other hand, the Lagrangian and the updated Lagrangian formulation can produce non steady-state models with chip breakage considered; a lot of updated Lagrangian models have enriched the relative literature, such as [55–58]. Although the updated Lagrangian



formulation adds considerably to the required calculation time, the advances in computers have made it possible to reduce the time needed for such an analysis to acceptable levels. However, two new aspects of machining modeling are introduced that their application is controversial and needs to be further studied, namely the use of a chip separation criterion and adaptive meshing.

The arbitrary Lagrangian-Eulerian formulation has also been proposed with the aim to combine the advantages of the two aforementioned methods [59–61]. This method uses the operator split procedure. The mesh is neither fixed nor attached to the material. Instead, it is allowed to arbitrarily move relative to the material with the total displacement being the sum of a Lagrangian displacement increment and an Eulerian displacement increment. A Lagrangian step is used in the procedure for the material flow at the free boundaries so that chip formation is the result of material deformation, thus mesh displacement in this step is associated with deformation. Then, in an Eulerian step, the reference system is suitably repositioned to compensate for the distortions during deformation, thus mesh displacement in this step is connected with numerical benefits. The procedure involves small time increments and it does not alter elements and connectivity of the mesh. Additionally, no separation criterion or extensive re-meshing is required. As a result, an ALE mesh is expected to be less distorted and more regular in comparison to a Lagrangian mesh. The drawbacks of the ALE formulation are the re-mapping of state variables, which may be performed inaccurately, and the need for a complete re-meshing [54].

### ***11.3.1 Model Geometry, Meshing and Chip Formation***

For mechanical micromachining, the geometrical characteristics of the tool and the workpiece greatly influence the outcome of the process. More specifically, the tool edge radius is connected to size effect, minimum chip thickness, effective rake angle, stagnation point and ploughing mechanism reviewed in Sect. 11.2. In mechanical micromachining the simulation of the process with a sharp tool would be pointless; the size effect would not be accounted for and so the size of the cutting forces and the chip formation would be unrealistic. Some researchers have investigated the influence of the tool edge radius on the size effect. Weber et al. have performed an analysis using similarity mechanics and various values for the tool edge radius [62]. Woon et al. also investigated the influence of the tool edge radius on the contact length between the tool and the workpiece and the material deformation and validated their numerical results with the aid of a small field-of-view photography technique [63]. Liu and Melkote have concluded from their analysis that the tool edge radius accounts only for part of the size effect in microcutting and that material strengthening is associated with the temperature drop in the secondary deformation zone, for higher cutting speeds [41].

The initial mesh of the workpiece is significant for the results the model will provide. The convergence of the numerical procedure and the accuracy of the

predicted variables depend on it. The obvious is that the mesh must be able to represent accurately the workpiece geometry and be able to handle the analysis to be performed. Structured and unstructured mesh generation procedures have been developed for the arrangement of the elements in the mesh and their individual geometry but there is not only one way to devise a representation of a continuum with finite elements. The size, number and type of the elements used in the mesh play a significant role on the simulation outcome as well. As a rule, a large number of small sized elements increases accuracy but also increases computational time. There is a threshold beyond which further increase in the number of the elements will significantly increase the time of the analysis with marginal gain in accuracy. Usually, coarser meshes are used for testing a model and a finer mesh is applied when the model is checked. Another technique is to identify the regions that are of more interest for the analysis. Finer meshes may be used in these regions that are combined with larger ones in the other regions. In machining the action takes place in the primary and secondary deformation zones; the mesh in these parts of the workpiece is expected to be denser in order to obtain better geometry of the chip and also be able to cope with the strains, strain rates and temperature gradients expected there. These parameters are incorporated in the analysis by a material model suitable for thermal, elastic and plastic effects to be accounted for; material modeling is the topic of the next paragraph. In Fig. 11.10 an example of finer discretization in the regions where primary and secondary deformation zones are anticipated is shown. Note also that the chip has finer mesh than the workpiece, except the deformation zones and tool tip, so that the mesh follows chip shape more accurately.

An element with a compact and regular shape is expected to perform better, i.e. as the aspect ratio of an element increases it loses accuracy. Low order elements, often with formulations to avoid volumetric locking behavior that can halt the analysis due to large incompressible plastic strains in the cutting area, are widely used due to their simplicity. The influence of numerical parameters on the modeling of orthogonal cutting is the topic of [64]. In the relevant literature quadrilateral elements with 8 and 9 nodes, enhanced 4-noded elements and triangular elements with 6 nodes can be found.

The boundary conditions applied in the initial mesh may differ. In the case of workpiece and tool different approaches have been proposed. Some researchers apply boundary conditions that allow the tool to advance towards the non-moving workpiece while others do just the opposite, considering the tool to be constrained from moving in the x-axis. In any case the result is equivalent; the tool and the workpiece have a relative movement to each other equal to the cutting speed. Contact and contact detection between chip and tool is of great importance in machining modeling. The most common algorithms used for solving contact problems are the penalty approach and the Lagrangian multipliers approach. Other procedures such as the augmented Lagrangian technique and the perturbed Lagrangian method are reported [65].

Of interest is the way thermo-mechanical coupling is considered. In cutting processes heat generation originates from the two deformation zones, i.e. the primary and the secondary, due to inelastic and frictional work. The heat is

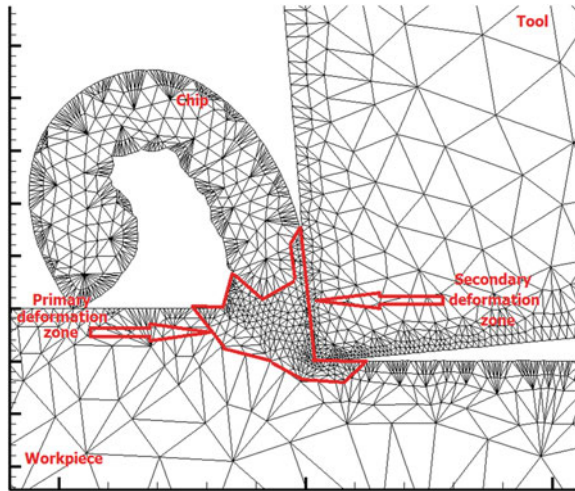


Fig. 11.10 Example of discretization

conducted into the tool and chip and transferred away from the chip to the environment or the cutting fluid by convection. The above are either modeled by heat sources at the heat generation regions or usually with material and tribological models that are functions of mechanical and thermal behavior with strain, strain-rate and temperature. The associated strain hardening and thermal softening is interpreted to non-linear analysis. This feature of the model is important for mechanical micromachining since the size effect is attributed to material strengthening due to decrease of temperature in the tool-chip interface with decreased chip thickness [36]; successful application of the thermo-mechanical coupling results in more accurate models. Moriwaki et al. [66] developed a thermo-mechanical model of micromachining of copper and calculated the stress, strain flow of the cutting heat and temperature of the tool and the workpiece.

In [53] a staggered procedure is adopted for coupling thermal and mechanical equations. There are two different meshes, one for the mechanical and one for the thermal model, which exchange information. A mechanical step is taken first with constant temperature and heat generation is computed. Then it is transferred to the thermal mesh. Temperatures are re-computed and transferred to the mechanical mesh to be inserted to the thermal-softening model used in the analysis. In another approach, all heat generated by deformation and friction is kept inside the mesh, external boundaries of workpiece, chip and tool are insulated, causing temperature rise [64, 67, 68]. If external boundaries are not adiabatic then heat convection takes place through natural convection or forced convection by the cutting fluid, for wet machining. Two difficulties arise in this case: the long standing problem of whether the cutting fluid actually makes its way between chip and tool and what are the actual thermal characteristics of the cutting fluid, e.g. heat transfer coefficient, in this region. Heat loss due to radiation is ignored, as it is considered

negligible. Thermal contact, i.e. the numerical technique to model heat transfer between chip and tool, is realized by several approaches such as the heat conduction continuity, the constant factor, two semi-infinite bodies and the thermal equilibrium approaches. It is worth noticing that for commercial FEM software various options and tools on the above are provided.

Lagrangian formulation based models, simulate chip generation either by plastic deformation considerations or by employing a chip separation criterion. The ideal would be to incorporate to the model the real physical mechanism of chip formation for a material being machined. It is generally thought that chip formation in ductile metal cutting involves only plastic deformation without any fracture. Many researchers prefer this approach as being more realistic, backed-up by the fact that microscopic observations of the cross-sectional areas of the chip revealed no evidence of a crack. It is argued that chip formation cannot be accomplished just by plastic deformation [69]. The fact that no crack is observed in the laboratory tests is attributed to crack stability rather than crack formation; large compressive stresses in the tool edge quench cracks or cracks have the same speed with the tool and cannot be seen [70]. Thus, the implementation of a separation criterion to simulate separation and fracture of the material is not only a modeling technique to overcome the problem of chip formation.

The two main techniques for chip separation are node-splitting and element deletion techniques [71]. In the node-splitting case a chip separation plane is predefined and a separation criterion is applied. There are two types of criteria, namely geometrical and physical. A simple geometrical criterion is to prescribe a critical distance  $d_c$  between the tool tip and the nearest node on the cutting direction [72]. This criterion is easy to control and can be used for cutting tools without sharp edge, which is important for mechanical micromachining, but it cannot account for breakage outside the cutting line. Furthermore, different critical distances result in different plastic strain distribution [73]. Physical criteria use the critical value of a physical quantity to estimate the onset of separation, e.g. in the plastic strain criterion chip separated when the calculated plastic strain at the nearest node to the cutting edge reaches the critical value [74]. The disadvantage of this criterion is that node separation may propagate faster than cutting speed, “unzipping” the mesh in front of the cutting tool. Another physical criterion is connected to stress [75]. Based on the Johnson-Cook yield stress equation a critical strain to fracture criterion is used [64]. An interesting review of the chip separation criteria can be found in [47]. The authors state that the criteria reviewed cannot simulate incipient cutting correctly. The matter remains controversial and more research needs to be carried-out to improve separating criteria for metal machining.

In the Lagrangian formulation, the initial mesh is altered significantly due to plastic deformation and chip separation. The distorted mesh causes numerical errors and the solution is rapidly degraded; the Jacobian determinant becomes negative for severe distortion and the analysis is halted. A strategy to address this problem is to use pre-distorted meshes. The advent of computers has made it possible to apply adaptive meshing techniques. During the simulation certain steps

are taken, e.g. the size of the elements, the location of the nodes or the number of the elements changes so that a new mesh, applicable for the analysis is created; this procedure takes place periodically. Figure 11.11 shows an example of an adaptive meshing technique, through the steps of the analysis.

Adaptive meshing can take place as a re-meshing technique, where the existing distorted mesh is substituted by a new one. The refinement technique increases mesh density by reducing mesh size and smoothing relocates nodes to provide more regular element shapes. Adaptive meshing improves the accuracy of the simulation but at a computational cost. This is attributed to the newer mesh being denser and thus more elements are involved in the analysis, but also to two other very important aspects of re-meshing, namely error and distortion metrics, for the assessment of the quality of the solution and transfer operators, that are responsible for transferring the variables of the old mesh to the new [77].

### 11.3.2 Material Modeling

Material modeling in both machining and micromachining is of great importance. Especially the flow properties of the workpiece material and the corresponding equations that are included into FEM have been extensively studied. These constitutive equations describe the flow stress or instantaneous yield strength at which work material starts to plastically deform or flow; the elastic strains are much lower than plastic strains in metal cutting and so workpiece material flows plastically into the cutting zone. The constitutive models presented in the literature are mainly elasto-plastic, elasto-viscoplastic, rigid-plastic and rigid-viscoplastic.

Machining conditions subject workpiece material to high levels of strain, strain rate and heat which greatly influence flow stress. In the primary zone strain and temperature ranges from 1–2 and 150–250 °C respectively and in the secondary deformation zone from 3 to much higher and 800–1,200 °C, while strain rates reach values of up to  $2 \times 10^4 \text{ s}^{-1}$  and  $10^5 \text{ s}^{-1}$  in the two zones [78]. If  $\sigma$  is stress,  $\varepsilon$  is plastic strain,  $\dot{\varepsilon}$  is plastic strain rate and  $T$  is temperature, a complete constitutive equation is in the form:

$$f(\sigma) = \sigma(\varepsilon, \dot{\varepsilon}, T) \quad (14)$$

The problem is the lack of data for high stresses, strain rates and temperatures as the ones encountered in machining. In many cases the constitutive data are taken from standard tension tests that are not sufficient for machining processes. Dynamic experimental material tests such as Split Hopkinson Pressure Bar (SHPB) impact testing is employed. Samples are deformed under high speed compression with strain rates of up to  $10^5 \text{ s}^{-1}$  and temperatures of up to 700 °C. However, the results are not sufficient for the deformation behavior of metals, especially in high speed machining; values beyond test results are calculated by interpolation. Astakhov and Outeiro criticized the use of SHPB results in

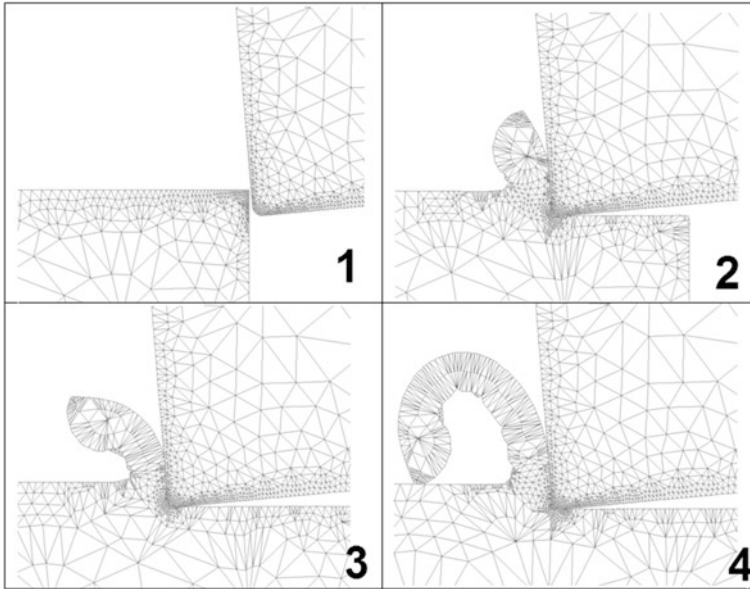


Fig. 11.11 Example of adaptive meshing [76]

machining [54]. They argue that the available data are not from specialized laboratories, generally speaking SHPB requires special equipment; high strain rates in metal cutting is a myth [79], metal cutting is a cold working process, although the chip only is of high temperature; finally, it not clear how to correlate uniaxial impact testing results of SHPB with materials that are triaxially stressed, as in metal cutting. Other tests used are torsion tests, compression ring tests and projectile impact tests [80].

Although many constitutive equations have been employed for the case of metal cutting, some are discussed here. The first is the relation by Usui, Maekawa and Shirakashi [81, 82]:

$$\sigma = B \left[ \frac{\dot{\varepsilon}}{1000} \right]^M e^{-kT} \left[ \frac{\dot{\varepsilon}}{1000} \right]^m \left\{ \int_{Path} e^{kT/N} \left[ \frac{\dot{\varepsilon}}{1000} \right]^{-m/N} d\varepsilon \right\}^N \quad (11.15)$$

In this equation B is the strength factor, M is the strain-rate sensitivity and n the strain hardening index, all functions of temperature T, and k and m are constants. The integral term accounts for the history effects of strain and temperature in relation to strain-rate. In the absence of these effects, Eq. 11.15 is reduced to [83]:

$$\sigma = B \left[ \frac{\dot{\varepsilon}}{1000} \right]^M \varepsilon^N \quad (11.16)$$

Oxley suggested a relation for carbon steel as [84]:

$$\sigma = \sigma_1 \dot{\epsilon}^n \quad (11.17)$$

with  $\sigma_1$  the material flow stress for  $\epsilon = 1$  and  $n$  is the strain hardening exponent. Both are functions of temperature, which is modified as:

$$T_{\text{mod}} = T[1 - 0.09 \log(\dot{\epsilon})] \quad (11.18)$$

for the combined effect of temperature and strain rate.

Among the most used material models is the Johnson-Cook model [85]. The equation consists of three terms the first one being the elasto-plastic term to represent strain hardening, the second is viscosity, which demonstrates that material flow stress increases for high strain rates and the temperature softening term; it is a thermo-elasto-visco-plastic material constitutive model, described as:

$$\sigma = (A + B\epsilon^n) \left[ 1 + C \ln\left(\frac{\dot{\epsilon}}{\dot{\epsilon}_o}\right) \right] \left[ 1 - \left(\frac{T - T_a}{T_m - T_a}\right)^m \right] \quad (11.19)$$

where  $T_a$  is the ambient temperature,  $T_m$  the melting temperature and  $A$ ,  $B$ ,  $C$ ,  $n$  and  $m$  are constants that depend on the material and are determined by material tests [86, 87] or predicted [88]. The influence of the Johnson-Cook constants on the outcome of machining modeling was investigated [89] and was found that FEM results are sensitive to these inputs, which in turn are strongly related to the test method used to derive the constants. On the other hand the results from a test method can be fitted to different constitutive equations and the selection of the material model can influence the predicted results [90, 91]. A review of material models used in manufacturing processes, including machining and micromachining, can be found in [92].

All the above mentioned FEM models refer to isotropic materials; no crystallographic effects are considered in the modeling process. Zerilli and Armstrong developed a constitutive model based on dislocation-mechanics theory and considering crystal structure of materials [93]. They suggested two different models, one for body cubic centered (BCC) and one for face cubic centered (FCC) lattice structure, respectively:

$$\sigma = C_o + C_1 \exp[-C_3 T + C_4 T \ln(\dot{\epsilon})] + C_5 \dot{\epsilon}^n \quad (11.20)$$

$$\sigma = C_o + C_2 \dot{\epsilon}^n \exp[-C_3 T + C_4 T \ln(\dot{\epsilon})] \quad (11.21)$$

where  $C_i$ ,  $i = 0 - 5$ , and  $n$  are material constants determined experimentally, e.g. by the SHPB method [94].

As noted in previous paragraphs, it is suggested by some researchers that the size effect in micromachining is not only induced by the effect of the edge radius but is also connected to strain gradient [38]; size effect is also observed when cutting is performed with a relatively sharp tool. Liu and Melkote [41] proposed a FEM model that uses a Taylor-based non-local theory of plasticity to represent the material under highly localized inhomogeneous deformation. It was shown that the

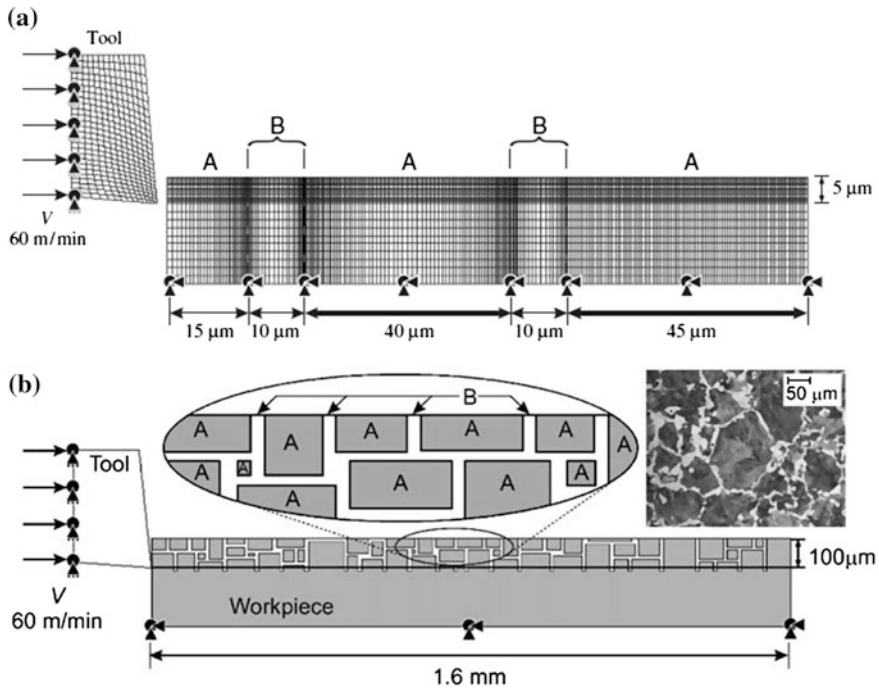
maximum effective stress is higher with strain gradient than without. Their results were validated with experiments.

In mechanical micromachining the cutting tool radius is comparable to the size of the grains of the material being processed. Furthermore, in materials with surface defects or multi phase materials, such as cast iron, the microcutting mechanism is quite different in comparison to non-heterogeneous materials due to the encounter of the cutting tip with these features of the material during the course of the process. Chuzhoy et al. prepared a model that simulates cutting at a microstructure level [95]. Ductile iron and two of its constituents, namely pearlite and ferrite, were model in the same continuum, taking into account the microstructural composition, the grain size and the distribution of each material. The model was in position of predicting stresses, strains and temperatures. Simoneau et al. studied the machining of 1045 steel, considering its microstructure [96–99]. In Fig. 11.12a the initial mesh of the material is presented where material A, with a pearlite-like behavior and material B, with a ferrite-like behavior, are represented as bands of appropriate size. In Fig. 11.12b the sections of the materials representing pearlite and ferrite for the same material, in another approach of the problem, are shown. Material plasticity was formulated with a strain dependent Johnson-Cook model. The results of the analysis showed good correlation between experimental and numerical results regarding the morphology of the chip and the behavior of each phase of the material. The predicted strain resulted larger in a heterogeneous FE model in comparison to a homogeneous one due to strain localization at each phase.

It is worth noticing that two case studies from macro machining are closely related to material modeling in the micro regime. The first pertains to grinding, a precision finishing process. Although this process can be modeled in macro scale as a thermal or thermo-mechanical system [10], it can also be modeled in micro scale. In the latter case the material removal mechanism of a grain of the grinding wheel is considered. In [100] a review of macro and micro scale of models of grinding can be found. The second case study pertains to the machining of metal matrix composite materials. The nature of these materials, a matrix material reinforced with fibers or particles of small size, is similar to the case of multiphase materials discussed above. It is clear that failure mechanisms of the composites materials in turning are different from monolithic materials; cracking of the reinforcement, debonding at the reinforcement-matrix interface, growth and coalescence of voids play an important role in the machining of composite materials. The available FEM models can successfully predict the performance of machining of such materials. For a review on the subject, Ref [101] is suggested.

Finally, most of the available models of mechanical micromachining pertain to continuous chip. However, Wang et al. used a constitutive model based on the Johnson-Cook model, adopted to describe the strain softening effect and simulate a shear localized chip with success [102, 103].





**Fig. 11.12** Representation of the microstructure of 1045 steel with **a** bands [96] and **b** sections [98] (material A: pearlite, material B: ferrite)

### 11.3.3 Friction Modeling

Friction modeling in the secondary deformation zone, at the interface of the chip and the rake face of the tool is of equal importance to the workpiece material modeling presented in the previous paragraph. It is important in order to determine cutting force but also tool wear and surface quality. Once again the detailed and accurate modeling is rather complicated. Many finite element models of machining assume that it is a case of classical friction situation following Coulomb’s law; frictional sliding force is proportional to the applied normal load. The ratio of these two is the coefficient of friction  $\mu$  which is constant in all the contact length between chip and tool. The relation between frictional stresses  $\tau$  and normal stresses may be expressed as:

$$\tau = \mu\sigma \tag{11.22}$$

However, as the normal stresses increase and surpass a critical value, this equation fails to give accurate predictions. From experimental analysis it has been verified that two contact regions may be distinguished in dry machining, namely the sticking and the sliding region. Zorev’s stick-slip temperature independent

friction model is the one commonly used [104]. In this model there is a transitional zone with distance  $\ell_c$  from the tool tip that signifies the transition from sticking to sliding region. Near the tool cutting edge and up to  $\ell_c$ , i.e. the sticking region, the shear stress is equal to the shear strength of the workpiece material,  $k$ , while in the sliding region the frictional stress increases according to Coulomb's law.

$$\tau = \begin{cases} k, & 0 \leq \ell \leq \ell_c \\ \mu\sigma, & \ell > \ell_c \end{cases} \quad (11.23)$$

In machining other approaches, based on Zorev's model, have been reported that include the defining of an average friction coefficient on the rake face or different coefficients for the sliding and the sticking region. In another approach, the constant shear model assumes that the frictional stress on the rake face of the tool is equal to a fixed percentage of the shear flow stress or the workpiece material.

Usui, based on Zorev's model and experimental results [81] proposed a non-linear stress expression:

$$\tau = k \left[ 1 - \exp\left(-\frac{\mu\sigma}{k}\right) \right] \quad (11.24)$$

This equation approaches the sticking region part of Eq. 11.23 for large  $\sigma$  and the sliding part for smaller values. However, the mean friction stress on the tool rake face may differ from the frictional stress in the sticking region. Childs [105] proposed another model:

$$\tau = mk \left[ 1 - \exp\left(-\frac{\mu\sigma}{mk}\right)^n \right]^{1/n} \quad (11.25)$$

In the last equation,  $m$  and  $n$  are correction factors; the former ensures that at high normal stresses the frictional stresses do not exceed  $k$  and the latter controls the transition from sticking to sliding region. These coefficients can be obtained by split-tool tests.

The evaluation of friction models has been the topic of some publications. An updated Lagrangian model to simulate orthogonal cutting of low carbon steel with continuous chip was prepared [56]. In a reverse engineering approach, five different friction models were tested and the results were compared against experimental results to decide which friction model is the most suitable. The results were best when friction models with variable shear stress and coefficient of friction were incorporated with the finite element models. Furthermore, an ALE model was used to measure the influence of friction models on several parameters [106]. It was concluded that friction modeling affects thrust forces more than cutting forces. Furthermore, on the implementation of the stick-slip model it is concluded that a major disadvantage is the uncertainty of the limiting shear stress value. In another work [107], five different friction models were analyzed and the investigators concluded that mechanical result, e.g forces, contact length, are practically insensitive to friction models, as long as the "correct" friction coefficient is

applied, while on the other hand, friction modeling greatly affects thermal results. In [108] an improved friction law formulation is suggested where the constant friction coefficient is replaced by one which increases with plastic strain rate:

$$\mu = \mu_o(1 + \alpha \dot{\epsilon}^p) \quad (11.26)$$

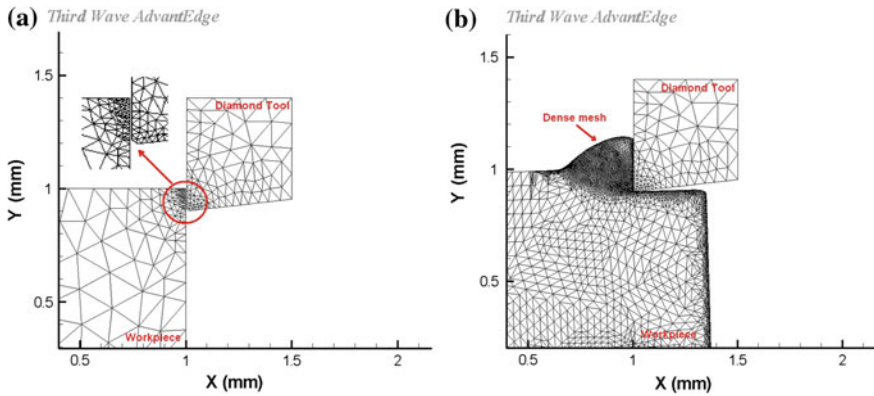
Another parameter, which is closely connected to friction and FEM modeling and is of special interest in micromachining, the contact length, is analyzed in [109]. Several contact length models utilized in the prediction of contact length in machining are analyzed. It should be noted that several papers presume frictionless contact in the chip-tool interface [62].

It is true that modeling with FEM is not at all trivial. However, a solution would be to use a commercial FEM program. For the past twenty years a wide range of commercial FEM packages became available. These programs have been widely accepted by researcher since they can simplify the overall procedure of model building. Commercial FEM add to the quality and accuracy of the produced models. These programs are made by specialists who have tested them and have implemented features and procedures to accelerate the slow process of model building. Most of the software have mesh generation programs, easy to use menus for applying boundary conditions, contact algorithms, automatic re-meshing, material databases etc. Some researchers, however, remain skeptical due to limitations a model can impose, e.g. a model may only be able to solve a problem implicitly or explicitly.

Regarding mechanical micromachining, researchers use either in-house FEM codes, commercial packages, e.g. ABAQUS [41, 110] or commercial packages with specific menus for micromachining, e.g. AvantEdge [111]. Figure 11.13 shows a FEM model of micromachining prepared with AdvantEdge [111]. In this figure, the continuous meshing and the adaptive remeshing procedures can be observed at the simulation progress. In Fig. 11.13a, in the related detail, the mesh of the workpiece is denser near the tool tip, where deformation is about to take place. The mesh coarsens in the areas with certain distance from workpiece surface. In the diamond tool, the mesh is denser near the tip, where more information will be acquired during the analysis. Figure 11.13b clearly indicates that new elements are created in the chip formation zone, where the strain rate is expected to be high; finer mesh can follow the curve of the curling material more closely and, furthermore, provide more accurate results.

## 11.4 Other Modeling Techniques and Micromachining Processes

It is true that the dominating method used in modeling of micromachining and especially microcutting is the finite element method; the number of publications proves it. FEM is used for the simulation of other micromachining methods like

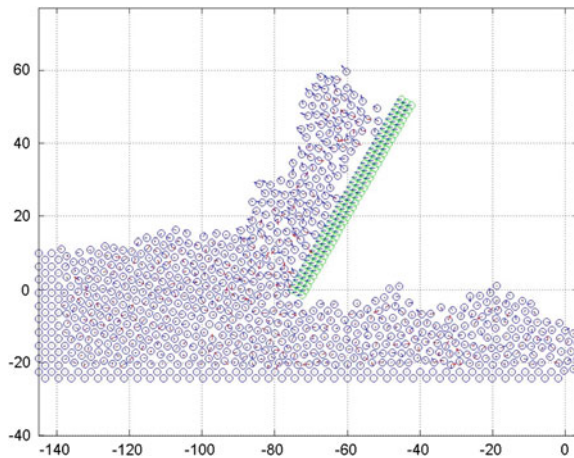


**Fig. 11.13** a Initial mesh, and b Mesh after a few increments

abrasive jet and laser micromachining [112, 113]. In this section some other modeling methods, besides FEM, will be discussed briefly.

Unlike FEM that can be employed for a wide spectrum of modeling, Molecular Dynamics (MD) is used for simulating nanometric cutting, i.e. the uncut chip thickness is in the range of nanometers. FEM is based on principles of continuum mechanics and at nanometric level this is considered a drawback. On the other hand, Molecular Dynamics can simulate the behavior of materials in atomic scale. MD is a modeling method in which atoms are interacting for a period of time, by means of a computer simulation. The interactions between particles are described by potentials, the most known being Lennard-Jones and Morse potential, and the governing equations of motion are the Newton's equations. In order to simulate molecular systems, a very big number of particles is involved and a vast number of equations is produced to describe the properties of these systems; as a multidisciplinary method, laws and theories from mathematics, physics and chemistry consist the backbone of the method. In order to deal with these problems, numerical methods, rather than analytical ones, are used and algorithms from computer science and information theory are employed. Although the method was originally intended to be exploited in theoretical physics, nowadays it is mostly applied in biomolecules, materials science and nanomanufacturing.

MD method was introduced in the simulation of micro and nanomanufacturing in the early 1990s [114, 115]. The results indicated that MD is a possible modeling tool for the microcutting process; atomistic modeling can provide better representation of micro and nanolevel characteristics than other modeling techniques. If FEM and MD simulations could be compared it could be said that FEM mesh is substituted in MD by atoms that play the role of the nodes and the distance between them coincides with interatomic distance; it is not arbitrarily chosen like the distance between the nodes in FEM. This way, phenomena that cannot be investigated with continuous mechanics can be simulated by MD because the atomic interaction and the microstructure are included in the simulation. MD



**Fig. 11.14** MD simulation of nanometric cutting

models are used for the investigation of the chip removal mechanisms, tool geometry optimization, cutting force estimations, subsurface damage identification, burr formation, surface roughness and surface integrity prediction; some relevant works can be found in References [116–127] where it can be seen that the models pertain to various materials, processes and are 2D or 3D. Figure 11.14 shows a typical MD simulation where the chip formation can be observed.

Some disadvantages can be identified in the MD technique as well. MD simulation is based on calculations of interatomic forces among a vast number of atoms that constitute the modeled system. This task requires significant computational power and in order to overcome this problem very small model sizes are simulated; some models are limited to nanometer or Angstrom level. Another feature is that cutting speed is considered to be unrealistically high, of the order of 200–500 m/s when typical speeds range between 2 and  $-10$  m/s in microcutting, in order to bring cutting speed closer to the atomic movement speed and thus save computational time. However, the detailed insight in material behavior in microcutting provided by MD simulation has supported process development and optimization and at the same time has provided information on its theoretical study. It is generally accepted that chip formation in cutting is owed to the shearing effect in the cutting zone of the workpiece. When machining with a depth of cut that is of the same order to the cutting edge radius, regardless of the nominal rake angle of the cutting tool, the effective rake angle in nanometric cutting is always negative, as already discussed in micromachining. Thus, a compressive stress imposes deformation in front of the cutting edge. The subsurface deformation is also described by a MD model in [128] where nanometric cutting is performed by an AFM tip tool, a technique proposed for the fabrication of MEMS and NEMS.

A review on MD simulation of machining at the atomic level can be found in [129]. Applications of MD simulation for grinding are also popular in order to study

the interaction of a single grain with the workpiece [130] and almost in every other micromachining procedure like micro-EDM and laser micromachining [131, 132].

Other modeling methods pertain to analytical [24], mechanistic [34, 40] and even soft computing [133] models for micromachining. It is of interest to examine the case of multiscale modeling that account both for continuum mechanics and atomistic modeling. In these models, methodologies that use hybrid FEM and MD modeling techniques in order to exploit the capabilities of each method and cover a more wide range of material behavior at nanoscale cutting are employed [134, 135].

## 11.5 Conclusions

Modeling and simulation are of capital importance in micromachining. They are used widely by researchers as there can be important reduction in experimentation time and cost. Furthermore, modeling and simulation can provide data and illuminate phenomena, a task that is difficult or even impossible in some cases to be obtained otherwise. This book chapter describes the most important features of modeling micromachining processes with special focus on FEM modeling of mechanical micromachining. In Sect. 11.2 the background of orthogonal modeling and its application in mechanical micromachining are thoroughly discussed; topics like size effect and minimum chip thickness are considered. Section 11.3 deals with the parameters that are taken into account in order to construct a FEM model for mechanical micromachining and the available options for each case are laid down. It is up to the modeler to incorporate some or all of the parameters in his model and also try to figure out which option for which parameter will work better for the at hand problem. It needs to be decided what mesh will be applied on the workpiece geometry, both size and shape, especially for the cutting edge radius, what kinds of elements are to be used, what boundary conditions are to be applied and how surface contacts will be modeled. Then, the formulation must be decided, i.e. Eulerian, Lagrangian or ALE and if the problem will be solved implicitly or explicitly. A real puzzle must be solved in connection with the material and friction model to be used and whether adaptive meshing or a chip separation will be applied on the model and finally make a successful modeling of the microstructure of the material that corresponds to real world parameters.

Providing a reliable and physically sound model is not easy at all. One must have a strong background on the problem dealt with FEM, in this case micromachining, but also on FEM method as well. One simple rule could be to start simple with the model and anticipate the results; then revise the model so that it includes more detail both in a physical respect, e.g. add material properties parameters that are acquired by a proper method for the specific material, and in numerical respect, e.g. use more elements or more focused mesh. However, FEM is considered the best option, in comparison to other methods, to provide reliable results, especially when combined with powerful computers. The accumulated experience on the method is

also an advantage. [Section 11.4](#) describes briefly other modeling methods employed in the micro and nanoworld, e.g. Molecular Dynamics.

The future of modeling and simulation at the micro and nanoscale depends greatly on the available computational power. More powerful systems in the future will be able to provide models where less simplifying assumptions are made and more phenomena are taken into account in one model. Models with the above mentioned characteristics will be able to provide more accurate results since they will be built in closer correlation to the existing systems they represent, allowing for models and simulations with wider fields of application. Furthermore, the computational time will decrease, providing thus the possibility of using simulations on-line with the processes.

In the case of micro and nanomachining several steps can be made in order to obtain more accurate and reliable results. Future trends include larger 3D models with adequate material modeling and longer simulations. Some of the topics to be addressed will be the ductile mode cutting of brittle materials and the chip formation at the nanoscale, surface integrity models and more effective microstructure simulations. Finally, modeling and simulation will be benefitted by future progress in physics and materials science, providing more complex interdisciplinary and thus more accurate models.

## References

1. Byrne G, Dornfeld D, Denkena B (2002) Advancing cutting technology. *Ann CIRP* 52(2):483–507
2. Corbett J, McKeown PA, Peggs GN, Whatmore R (2000) Nanotechnology: international developments and emerging products. *Ann CIRP* 49(2):523–545
3. Masuzawa T (2000) State of the art of micromachining. *Ann CIRP* 49(2):473–488
4. Altıng L, Kimura F, Hansen HN, Bissacco G (2003) Micro engineering. *Ann CIRP* 52(2):635–657
5. Mamalis AG, Markopoulos A, Manolakos DE (2005) Micro and nanoprocessing techniques and applications. *Nanotechnol Percept* 1:31–52
6. Dornfeld D, Min S, Takeuchi Y (2006) Recent advances in mechanical micromachining. *Ann CIRP* 55(2):745–768
7. Madou MJ (2002) *Fundamentals of microfabrication: the science of miniaturization*. CRC Press, Florida
8. Stephenson DA, Agapiou JS (2006) *Metal cutting theory and practice*, 2nd edn. CRC Press, FL, USA
9. Zorev NN (1966) *Metal cutting mechanics*. Pergamon Press, Oxford
10. Markopoulos AP (2012) *Finite element method in machining processes*. Springer, London, UK
11. Ernst H, Merchant ME (1941) Chip formation, friction and high quality machined surfaces. surface treatment of metals. *Am Soc Metals* 29:299–378
12. Merchant ME (1945) Mechanics of the metal cutting process II. Plasticity conditions in orthogonal cutting. *J Appl Phys* 16(6):318–324
13. Lee EH, Shaffer BW (1951) The theory of plasticity applied to a problem of machining. *Trans ASME J Appl Mech* 18:405–413
14. Kudo H (1965) Some new slip-line solutions for two-dimensional Steady-state machining. *Int J Mech Sci* 7:43–55

15. Dewhurst P (1978) On the non-uniqueness of the machining process. *Proc R Soc Lond A* 360:587–610
16. Oxley PLB (1989) *The mechanics of machining: an analytical approach to assessing machinability*. Ellis Horwood, Chichester, UK
17. van Luttervelt CA, Childs THC, Jawahir IS, Klocke F and Venuvinod PK (1998) Present situation and future trends in modeling of machining operations. Progress report of the CIRP working group modeling and machining operations. *Ann CIRP* 47(2):587–626
18. Astakhov VP (2005) On the inadequacy of the single-shear plane model of chip formation. *Int J Mech Sci* 47:1649–1672
19. Shaw MC, Cook NH, Smith PA (1952) The mechanics of three-dimensional cutting operations. *Trans ASME* 74:1055–1064
20. Morcos WA (1980) A slip line field solution of the free continuous cutting problem in conditions of light friction at chip-tool interface. *Trans ASME J Eng Ind* 102:310–314
21. Usui E, Hirota A, Masuko M (1978) Analytical prediction of three dimensional cutting process. Part 1. Basic cutting model and energy approach. *Trans ASME J Eng Ind* 100:222–228
22. Usui E, Hirota A (1978) Analytical prediction of three dimensional cutting process. Part 2. Chip formation and cutting force with conventional single-point tool. *Trans ASME J Eng Ind* 100:229–235
23. Usui E, Shirakashi T, Kitagawa T (1978) Analytical prediction of three dimensional cutting process. Part 3. Cutting temperature and crater wear of carbide tool. *Trans ASME J Eng Ind* 100:236–243
24. Malekian M, Mostofa MG, Park SS, Jun MBG (2012) Modeling of minimum uncut chip thickness in micro machining of aluminum. *J Mater Process Technol* 212:553–559
25. Finnie I (1963) A comparison of stress strain behavior in cutting with that in other materials tests. In: Shaw M (ed) *International research in production engineering: proceedings of the international production engineering research conference*, Carnegie Institute of Technology, Pittsburgh, Pennsylvania, ASME:76–82
26. Ikawa N, Shimada S, Tanaka H, Ohmori G (1991) Atomistic analysis of nanometric chip removal as affected by tool-work interaction in diamond turning. *Ann CIRP* 40(1):551–554
27. Weule H, Huntrup V, Tritschle H (2001) Micro-cutting of steel to meet new requirements in miniaturization. *Ann CIRP* 50(1):61–64
28. Kim CJ, Bono M, Ni J (2002) Experimental analysis of chip formation in micro-milling. *Trans NAMRI/SME* 30:1–8
29. Albrecht P (1960) New developments in the theory of metal cutting processes: Part I-the ploughing process in metal cutting. *ASME J Eng Ind* 82:348–357
30. Albrecht P (1961) New developments in the theory of metal cutting processes: part II-the theory of chip formation. *ASME J Eng Ind* 83:557–568
31. Masuko M (1956) Fundamental research on the metal cutting: second report. *Bull Jpn Soc Mech Eng* 22:371–377
32. Backer WR, Marshall ER, Shaw MC (1952) The size effect in metal cutting. *J Manuf Sci Eng Trans ASME* 74:61–72
33. Taniguchi N (1994) The state of the art of nanotechnology for processing of ultraprecision and ultrafine products. *Precis Eng* 16:5–24
34. Kim JD, Kim DS (1995) Theoretical analysis of micro-cutting characteristics in ultra-precision machining. *J Mater Process Technol* 49:387–398
35. Larsen-Basse J, Oxley PLB (1973) Effect of strain rate sensitivity on scale phenomena in chip formation. In: *Proceedings of the 13th international machine tool design and research conference*, University of Birmingham, pp 209–216
36. Kopalinsky EM, Oxley PLB (1984) Size effects in metal removal processes. *Inst Phys Conf Ser* 70:389–396
37. Atkins AG (2003) Modeling metal cutting using modern ductile fracture mechanics: quantitative explanations for some longstanding problems. *Int J Mech Sci* 45:373–396



38. Dinesh D, Swaminathan S, Chandrasekar S, Farris TN (2001) An intrinsic size-effect in machining due to the strain gradient. In: Proceedings of the 2001 ASME IMECE, NY, 11–16 Nov, pp 197–204
39. Shaw MC, Jackson MJ (2006) The size effect in micromachining. In: Jackson MJ (ed) Microfabrication and nanomanufacturing. CRC Press, Boca Raton, FL, USA
40. Joshi SS, Melkote SN (2004) An explanation for the size-effect in machining based on strain gradient plasticity. *J Manuf Sci Eng Trans ASME* 126:679–684
41. Liu K, Melkote SN (2007) Finite element analysis of the influence of tool edge radius on size effect in orthogonal micro-cutting process. *Int J Mech Sci* 49:650–660
42. Shaw MC (2003) The size effect in metal cutting. *Sadhana Acad Proc Eng Sci* 28(5):875–896
43. Liu X, DeVor RE, Kapoor SG, Ehmann KF (2004) The mechanics of machining at the microscale: assessment of the current state of the science. *J Manuf Sci Eng Trans ASME* 126:666–678
44. Dixit PM, Dixit US (2008) Modeling of metal forming and machining processes by finite element and soft computing methods. Springer-Verlag London Limited, UK
45. Klocke F, Beck T, Hoppe S, Krieg T, Müller N, Nöthe T, Raedt HW, Sweeney K (2002) Examples of FEM application in manufacturing technology. *J Mater Process Technol* 120:450–457
46. Mamalis AG, Manolakos DE, Ioannidis MB, Markopoulos A, Vottea IN (2003) Simulation of advanced manufacturing of solids and porous materials. *Int J Manuf Sci Prod* 5(3):111–130
47. Huang JM, Black JT (1996) An evaluation of chip separation criteria for the FEM simulation of machining. *J Manuf Sci Eng* 118:545–554
48. Owen DRJ, Vaz M Jr (1999) Computational techniques applied to high-speed machining under adiabatic strain localization conditions. *Comput Methods Appl Mech Eng* 171:445–461
49. Lindgren LE, Edberg J (1990) Explicit versus implicit finite element formulation in simulation of rolling. *J Mater Process Technol* 24:85–94
50. Sun JS, Lee KH, Lee HP (2000) Comparison of implicit and explicit finite element methods for dynamic problems. *J Mater Process Technol* 105:110–118
51. Harewood FJ, McHugh PE (2007) Comparison of the implicit and explicit finite element methods using crystal plasticity. *Comput Mater Sci* 39:481–494
52. Shih AJ (1996) Finite element analysis of rake angle effects in orthogonal metal cutting. *Int J Mech Sci* 38:1–17
53. Marusich TD, Ortiz M (1995) Modelling and simulation of high-speed machining. *Int J Numer Meth Eng* 38:3675–3694
54. Astakhov VP, Outeiro JC (2008) Metal cutting mechanics, finite element modelling. In: Davim JP (ed) Machining: fundamentals and recent advances. Springer-Verlag Limited, UK
55. Bäker M, Rösler J, Siemers C (2002) A Finite element model of high speed metal cutting with adiabatic shearing. *Comput Struct* 80:495–513
56. Özel T (2006) The influence of friction models on finite element simulations of machining. *Int J Mach Tools Manuf* 46:518–530
57. Maranhão C, Davim JP (2010) Finite element modelling of machining of AISI 316 steel: numerical simulation and experimental validation. *Simul Model Pract Theory* 18:139–156
58. Özel T, Thepsonthi T, Ulutan D, Kaftanoğlu B (2011) Experiments and finite element simulations on micro-milling of Ti–6Al–4 V alloy with uncoated and cBN coated micro-tools. *CIRP Ann Manuf Technol* 60:85–88
59. Movahhedy MR, Altintas Y, Gadala MS (2002) Numerical analysis of metal cutting with chamfered and blunt tools. *Tran ASMEJ Manuf Sci Eng* 124:178–188
60. Arrazola PJ, Özel T (2008) Numerical modelling of 3-D hard turning using arbitrary eulerian lagrangian finite element method. *Int J Mach Mach Mater* 3:238–249
61. Woon KS, Rahman M, Fang FZ, Neo KS, Liu K (2008) Investigations of tool edge radius effect in micromachining: a FEM simulation approach. *J Mater Process Technol* 195:204–211

62. Weber M, Hochrainer T, Gumbsch P, Autenrieth H, Delonnoy L, Schulze V, Löhle D, Kotschenreuther J, Fleischer J (2007) Investigation of size-effects in machining with geometrically defined cutting edges. *Mach Sci Technol* 11:447–473
63. Woon KS, Rahman M, Neo KS, Liu K (2008) The effect of tool edge radius on the contact phenomenon of tool-based micromachining. *Int J Mach Tools Manuf* 48:1395–1407
64. Barge M, Hamdi H, Rech J, Bergheau J-M (2005) Numerical modelling of orthogonal cutting: influence of numerical parameters. *J Mater Process Technol* 164–165:1148–1153
65. Vaz M Jr, Owen DRJ, Kalhori V, Lundblad M, Lindgren L-E (2007) Modelling and simulation of machining processes. *Arch Comput Methods Eng* 14:173–204
66. Moriwaki T, Sugimura N, Luan S (1993) Combined stress, material flow and heat analysis of orthogonal micromachining of copper. *Ann CIRP* 42(1):75–78
67. Vaz M Jr, Owen DRJ (2001) Aspects of ductile fracture and adaptive mesh refinement in damaged elasto-plastic materials. *Inter J Numer Meth Eng* 50:29–54
68. Wen Q, Guo YB, Todd BA (2006) An adaptive FEA method to predict surface quality in hard machining. *J Mater Process Technol* 173:21–28
69. Atkins AG (2006) Toughness and oblique cutting. *Trans ASME J Manuf Sci Eng* 128(3):775–786
70. Rosa PAR, Martins PAF, Atkins AG (2007) Revisiting the fundamentals of metal cutting by means of finite elements and ductile fracture mechanics. *Int J Mach Tools Manuf* 47:607–617
71. Ng EG, Aspinwall DK (2002) Modeling of hard part machining. *J Mater Process Technol* 127:222–229
72. Usui E, Shirakashi T (1982) Mechanics of Machining—from “Descriptive” to “Predictive” Theory”. In: Kops L, Ramalingam S (eds) *On the art of cutting metals—75 years later: a tribute to F.W. Taylor*, proceedings of the winter annual meeting of the ASME PED, vol 7. pp 13–35
73. Lin ZC, Lin SY (1992) A Couple finite element model of thermo-elastic-plastic large deformation for orthogonal cutting. *ASME J Eng Ind* 114:218–226
74. Liu CR, Guo YB (2000) Finite element analysis of the effect of sequential cuts and tool-chip friction on residual stresses in a machined layer. *Int J Mech Sci* 42:1069–1086
75. Shet C, Deng X (2000) Finite element analysis of the orthogonal metal cutting process. *J Mater Process Technol* 105:95–109
76. Markopoulos AP (2006) *Ultraprecision material removal processes*. Ph.D. Thesis, National Technical University of Athens, Greece
77. Kalhori V (2001) *Modelling and simulation of mechanical cutting*. Ph.D. Thesis, Luleå University of Technology, Luleå, Sweden
78. Jaspers SPFC, Dautzenberg JH (2002) Material behaviour in metal cutting: strains, strain rates and temperatures in chip formation. *J Mater Process Technol* 121:123–135
79. Astakhov VP (2006) *Tribology of metal cutting*. Elsevier, London, UK
80. Athavale SM, Strenkowski JS (1998) Finite element modeling of machining: from proof-of-concept to engineering applications. *Mach Sci Technol* 2(2):317–342
81. Usui E, Maekawa K, Shirakashi T (1981) Simulation analysis of built-up edge formation in machining low carbon steels. *Bull Jpn Soc Precis Eng* 15:237–242
82. Maekawa K, Shirakashi T, Usui E (1983) Flow stress of low carbon steel at high temperature and strain rate (Part 2). *Bull Jpn Soci Precis Eng* 17(3):167–172
83. Childs THC, Otieno AMW, Maekawa K (1994) The influence of material flow properties on the machining of steels. *Proceedings of the Third International Conference on the Behaviour of Materials in Machining*, Warwick, pp 104–119
84. Oxley PLB (1989) *The mechanics of machining: an analytical approach to assessing machinability*. Ellis Horwood, Chichester, UK
85. Johnson GR, Cook WH (1983) A constitutive model and data for metals subjected to large strains, high strain rates and high temperatures. In: *Proceedings of the 7th international symposium on Ballistics*, The Hague, The Netherlands, pp 541–547

86. Jaspers SPFC, Dautzenberg JH (2002) Material behaviour in conditions similar to Metal cutting: flow stress in the primary shear zone. *J Mater Process Technol* 122:322–330
87. Lee WS, Lin CF (1998) High-temperature deformation behavior of Ti6Al4 V alloy evaluated by high strain-rate compression tests. *J Mater Process Technol* 75:127–136
88. Özel T, Karpat Y (2007) Identification of constitutive material model parameters for high-strain rate metal cutting conditions using evolutionary computational algorithms. *Mater Manuf Processes* 22:659–667
89. Umbrello D, M'Saoubi R, Outeiro JC (2007) The influence of Johnson-Cook material constants on finite element simulation of machining of AISI 316L steel. *Int J Mach Tools Manuf* 47:462–470
90. Liang R, Khan AS (1999) A critical review of experimental Results and constitutive models for BCC and FCC metals over a wide range of strain rates and temperatures. *Int J Plast* 15:963–980
91. Shi J, Liu CR (2004) The influence of material models on finite element simulation of machining. *Trans ASME J Manuf Sci Eng* 126:849–857
92. Dixit US, Joshi SN, Davim JP (2011) Incorporating of material behavior modeling of metal forming and machining processes: a review. *Mater Des* 32:3655–3670
93. Zerilli FJ, Armstrong RW (1987) Dislocation-mechanics-based constitutive relations for material dynamics calculations. *J Appl Phys* 61:1816–1825
94. Meyer HW Jr, Kleponis DS (2001) Modeling the high strain rate behavior of titanium undergoing ballistic impact and penetration. *Int J Impact Eng* 26:509–521
95. Chuzhoy L, DeVor RE, Kapoor SG (2003) Machining simulation of ductile Iron and its constituents, part 2: numerical simulation and experimental validation of machining. *Trans ASME J Manuf Sci Eng* 125:192–201
96. Simoneau A, Ng E, Elbestawi MA (2006) Surface defects during microcutting. *Int J Mach Tools Manuf* 46:1378–1387
97. Simoneau A, Ng E, Elbestawi MA (2006) Chip formation during microscale cutting of a medium carbon steel. *Int J Mach Tools Manuf* 46:467–481
98. Simoneau A, Ng E, Elbestawi MA (2007) Modeling the effects of microstructure in metal cutting. *Int J Mach Tools Manuf* 47:368–375
99. Simoneau A, Ng E, Elbestawi MA (2007) Grain size and orientation effects when microcutting AISI 1045 steel. *Ann CIRP* 56(1):57–60
100. Doman DA, Warketin A, Bauer R (2009) Finite element modelling approaches in grinding. *Int J Mach Tools Manuf* 49:109–116
101. Dandekar CR, Shin YC (2012) Modeling of machining of composite materials: a review. *Int J Mach Tools Manuf* 57:102–121
102. Wang H, To S, Chan CY, Cheung CF, Lee WB (2010) Elastic strain induced shearbands in the microcutting process. *Int J Mach Tools Manuf* 50:9–18
103. Wang H, To S, Chan CY, Cheung CF, Lee WB (2010) A study of regularly spaced shear bands and morphology of serrated chip formation in microcutting process. *Scripta Mater* 63:227–230
104. Zorev NN (1963) Interrelationship between shear processes occurring along tool face and on shear plane in metal cutting. *Proceedings of the International Research in Production Engineering Conference, ASME, New York*, pp 42–49
105. Childs THC, Maekawa K (1990) Computer-aided simulation and experimental studies of chip flow and tool wear in the turning of low alloy steels by cemented carbide tools. *Wear* 139:235–250
106. Arrazola PJ, Özel T (2010) Investigations on the effects of friction modeling in finite element simulation of machining. *Int J Mech Sci* 52:31–42
107. Filice L, Micari F, Rizzuti S, Umbrello D (2007) A critical analysis on the friction modelling in orthogonal machining. *Int J Mach Tools Manuf* 47:709–714
108. Childs THC (2006) Friction modelling in metal cutting. *Wear* 260:310–318

109. Iqbal SA, Mativenga PT, Sheikh MA (2008) Contact length prediction: mathematical models and effect of friction schemes on FEM simulation for conventional to HSM of AISI 1045 steel. *Int J Mach Mach Mater* 3(1/2):18–32
110. Woon KS, Rahman M, Liu K (2010) Numerical and experimental study of contact behavior in the tool-based micromachining of steel. *Inter J Precis Eng Manuf* 11(3):453–459
111. Markopoulos AP, Manolakos DE (2010) Finite element analysis of micromachining. *J Manuf Technol Res* 2(1–2):17–30
112. Getu H, Spelt JK, Papini M (2011) Thermal analysis of cryogenically assisted abrasive jet micromachining of PDMS. *Int J Mach Tools Manuf* 51:721–730
113. Gupta MC, Li B, Gadag S, Chou KC (2010) Laser micromachining for fatigue and fracture mechanics applications. *Opt Lasers Eng* 48:441–447
114. Stowers IF, Belak JF, Lucca DA, Komanduri R, Moriwaki T, Okuda K, Ikawa N, Shimada S, Tanaka H, Dow TA, Drescher JD (1991) Molecular- dynamics simulation of the chip forming process in single crystal copper and comparison with experimental data. *Proc ASPE Ann Meet* 1991:100–104
115. Ikawa N, Shimada S, Tanaka H, Ohmori G (1991) Atomistic analysis of nanometric chip removal as affected by tool-work interaction in diamond turning. *Ann CIRP* 40(1):551–554
116. Rentsch R, Inasaki I (1995) Investigation of surface integrity by molecular dynamics simulation. *Ann CIRP* 42(1):295–298
117. Komanduri R, Chandrasekaran N, Raff LM (2001) MD simulation of exit failure in nanometric cutting. *Mater Sci Eng A* 311:1–12
118. McGeough J (ed) (2002) *Micromachining of engineering materials*. Marcel Dekker Inc., New York, USA
119. Luo X, Cheng K, Guo X, Holt R (2003) An investigation on the mechanics of nanometric cutting and the development of its test-bed. *Int J Prod Res* 41(7):1449–1465
120. Rentsch R (2004) Molecular dynamics simulation of micromachining of pre-machined surfaces. In: *Proceedings of 4th euspens international conference*, Glasgow, Scotland, pp 139–140
121. Fang FZ, Wu H, Liu YC (2005) Modelling and experimental investigation on nanometric cutting of monocrystalline silicon. *Int J Mach Tools Manuf* 45:1681–1686
122. Pei QX, Lu C, Fang FZ, Wu H (2006) Nanometric cutting of copper: a molecular dynamics study. *Comput Mater Sci* 37:434–441
123. Cai MB, Li XP, Rahman M (2007) Study of the temperature and stress in nanoscale ductile mode cutting of silicon using molecular dynamics simulation. *J Mater Process Technol* 192–193:607–612
124. Narulkar R, Bukkapatnam S, Raff LM, Komanduri R (2009) Graphitization as a precursor to wear of diamond in machining pure iron: a molecular dynamics investigation. *Comput Mater Sci* 45:358–366
125. Kim CJ, Mayor R, Ni J (2011) Molecular dynamics simulation of plastic material deformation in machining with a round cutting edge. *Inter J Precis Eng Manuf* 13(8):1303–1309
126. Shi J, Shi Y, Liu CR (2011) Evaluation of a three-dimensional single-point turning at atomistic level by a molecular dynamic simulation. *Int J Adv Manuf Technol* 54:161–171
127. Davim JP, Jackson MJ (eds) (2009) *Nano and micromachining*. ISTE Ltd, London, UK
128. Zhang JJ, Sun T, Yan YD, Liang YC, Dong S (2008) Molecular dynamics simulation of subsurface deformed layers in AFM-based nanometric cutting process. *Appl Surf Sci* 254:4774–4779
129. Komanduri R, Raff LM (2001) A Review on the molecular dynamics simulation of machining at the atomic scale. *Proc Inst Mech Eng Part B J Eng Manuf* 215:1639–1672
130. Brinksmeier E, Aurich JC, Govekar E, Heinzel C, Hoffmeister H-W, Klocke F, Peters J, Rentsch R, Stephenson DJ, Uhlmann E, Weinert K, Wittmann M (2006) Advances in modeling and simulation of grinding processes. *Ann CIRP* 55(2):667–696
131. Yang X, Guo J, Chen X, Kunieda M (2011) Molecular dynamics simulation of the material removal mechanism in micro-EDM. *Precis Eng* 35:51–57

132. Cheng J, Liu C-S, Shang S, Liu D, Perrie W, Dearden G, Watkins K (2013) A review of ultrafast laser materials micromachining. *Opt Laser Technol* 46:88–102
133. Ciurana J, Arias G, Ozel T (2009) Neural network modeling and particle swarm optimization (PSO) of process parameters in pulsed laser micromachining of hardened AISI H13 steel. *Mater Manuf Process* 24:358–368
134. Aly MF, Ng E-G, Veldhuis SC, Elbestawi MA (2006) Prediction of cutting forces in the micro-machining of silicon using a “hybrid molecular dynamic-finite element analysis” force model. *Int J Mach Tools Manuf* 46:1727–1739
135. Lin Z-C, Huang J-C, Jeng Y-R (2007) 3D nano-scale cutting model for Nickel material. *J Mater Process Technol* 192–193:27–36

# Chapter 12

## Micro Milling

Emel Kuram and Babur Ozelik

**Abstract** Micro products and components are demanded for industrial applications including electronics, optics, aerospace, medicine, biotechnology etc. in recent years. Laser machining, focused ion beam machining, electrochemical machining, electrodischarge machining and micro mechanical machining have been used to manufacture such components. However, the capability of micro mechanical machining especially micro milling to manufacture a wide range of workpiece materials and complex three-dimensional geometries makes it one of the best candidates to produce the micro parts. Therefore, this chapter provides a review of the current developments in the field of micro milling. Since the mechanism of material removal in micro milling is different from macro milling due to the presence of minimum chip thickness, size effect, elastic recovery and ploughing mechanism, these effects must be taken into consideration in micro milling research. In this chapter, important aspects of the micro milling mechanisms such as minimum chip thickness, size effect, effects of workpiece materials microstructure and cutting edge radius are presented. Furthermore, micro cutting tools are tiny and fragile and can be easily broken due to the excessive deflections, forces and vibrations. Therefore, an appropriate cutting tool geometry and cutting conditions must be selected in micro milling. To this end, the development of micro tool research and applications in micro milling are also presented.

### 12.1 Introduction

Nowadays, many industries such as electronics, optics, aerospace, medicine, biotechnology etc. demand micro products and components. The miniaturization of components has some advantages, such as portability and reduction in space,

---

E. Kuram · B. Ozelik (✉)  
Department of Mechanical Engineering, Gebze Institute of Technology,  
41400 Gebze-Kocaeli, Turkey  
e-mail: ozcelik@gyte.edu.tr

power, energy consumption, materials and costs. To meet this demand, it is mandatory to develop an appropriate manufacturing process. This manufacturing process must be used not only for a variety of workpiece materials but also for manufacturing of three-dimensional (3D) features. Laser machining, focused ion beam (FIB) machining, electrochemical machining and electrodischarge machining (EDM) are used to manufacture micro parts, however these non-traditional fabrication methods have some drawbacks due to the using of limited workpiece material (such as silicon), poor productivity and high cost. Due to these negative aspects of non-traditional fabrication methods, micro mechanical machining (or micro machining) such as micro milling, micro drilling, micro turning and micro grinding has gained increasing interest in recent years. Micro mechanical machining is one of the production methods for micro components, which is the mechanical removal of workpiece materials using miniature cutting tools.

There is no consensus about the dimension definition of micromachining. In the literature, some attempts can be seen to define the micromachining and proposed definitions for micromachining are as follows: Masuzawa and Tönshoff [1] defined the micromachining as the using the undeformed (uncut) chip thickness ranging from 0.1 to 200  $\mu\text{m}$ . Later, Masuzawa [2] stated that the range of micromachining varied according to era, person, machining method, type of product or material. Liu et al. [3] explained that the undeformed chip thickness was comparable to cutting tool edge radius in micromachining. Chae et al. [4] defined the micromachining as the fabrication method for creating miniature devices and components with features that varied from tens of micrometers to a few millimeters in size. Dornfeld et al. [5] defined the micromachining as the mechanical cutting of features with tool engagement less than 1 mm, using geometrically defined cutting edges. Aramcharoen et al. [6] explained that if cutting tool diameter falls in the range of 1–999  $\mu\text{m}$  or if undeformed chip thickness is comparable to the cutting edge radius or material grain size, this could be considered as micromachining. In general, micromachining means the machining of the dimensions below 1 mm (between 1 and 999  $\mu\text{m}$ ) and cutting tools at diameter below 1 mm used in the micro milling literatures have considered in this chapter.

Micro milling is one of the micro mechanical machining processes and it is a downscaled process of the conventional (macro) milling. 3D micro parts with high aspect ratios and complex geometrical shapes in various engineering materials are manufactured with micro milling process. Micro milling is suitable especially for the manufacture of moulds and dies.

Micro milling is similar to conventional milling process however there are some differences between micro and macro milling. In micro milling unlike conventional milling, there are no handbooks available to determine the machining parameters and recommended data by tool manufacturers cannot directly be applied to micro milling. If the recommended cutting speed for conventional milling of aluminium material is applied to micro milling, calculated spindle speed will be about 350,000 rev/min for 100  $\mu\text{m}$  diameter cutting tool, which is difficult to obtain [7].

In micro milling process, the effect of the cutting edge radius should be taken into consideration. The ratio of feed per tooth to radius of the cutting tool in micro milling is greater than conventional milling.

Run-out in micro milling process, even within micrometer levels, affects the cutting process extensively. Although run-out in macro milling is higher than that in micro milling, run-out in macro milling is sometimes negligible.

The control of micro milling process to achieve desired results (longer tool life, better surface quality, burr-free machining) is harder compared to macro milling. This is due to the size effect and when machining hard or difficult-to-cut workpiece materials, this problem is much pronounced. In the case of conventional milling, operators can monitor the machining process visually. However, the micro milling process cannot be monitored visually by the operators.

The primary drawbacks of micro milling are low productivity and high cost. Low productivity in micro milling is owing to low material removal rates. High cost of micro milling is due to the short tool life and high cost of machine and cutting tools.

Unpredictable tool life and premature tool failures are the main drawbacks for micro milling process. This is due to the increasing stress concentrations and decreasing tool stiffness with a decrement of the tool diameter. Micro cutting tools have low flexural stiffness and strength owing to their small diameter. Micro cutting tools can easily bend, which affecting the cutting process negatively and causing sudden tool breakage. Regenerative chatter is also one of the challenges in micro milling due to the resulting poor surface quality and shorter tool life.

During machining various tools have been used and tools are usually changed. The integrity of X and Y directions has not been affected by tool change. However, whenever a tool is changed, the integrity in Z direction is lost and adjusting z-coordinates requires additional efforts and time. Currently available commercial touch probes are not able to handle less than 1  $\mu\text{m}$  resolution. Therefore, it is crucial to find reliable, high sensitivity and easy method for detecting tool-workpiece contact in micro milling [8]. For this purpose, an acoustic emission (AE)-based system with micron-level accuracy for detecting initial tool-workpiece contact in micro machining was proposed by Min et al. [8] and Bourne et al. [9].

## 12.2 Micro Milling Equipments

### 12.2.1 Micro Cutting Tools

The tool geometry of the commercial micro end mill such as rake angle, relief angle etc. is derived from the macro end mills. But, the geometrical values cannot be properly obtained for micro end mills owing to the manufacturing process limitation. The fabrication of micro milling tools is difficult task and micro tools break easily during machining. Some processes such as focused ion beam (FIB),



wire electrical discharge grinding (WEDG) and grinding are used to manufacture the micro cutting tools. All manufacturing processes have some limitations. For example, the size of the cutting edge radius and the quality of micro cutting tools which are manufactured by grinding, depend on the properties of grinding wheel, such as wheel grit, wheel wear and carbide grain size. Therefore attempts have been focused the improvements in tool manufacturing technology, tool geometry, tool material and coatings.

In micro milling, the quality of micro tools should be checked before experimental study and cutting tools produced from the same batch should be used to minimize the geometrical deviations.

The stiffness of the micro cutting tool is low due to the small dimension of the micro tools. Low stiffness can result in the bending of the cutting tool, vibration and chatter during machining. Micro cutting tools can be modelled as a simple cantilever beam (Fig. 12.1) and the deflection and bending stress can be found by the following equations [4]:

$$\delta = \frac{FL^3}{3EI} = \frac{64FL^3}{3Ed^4} \quad (12.1)$$

$$\sigma = \frac{32FL}{\pi d^3} \quad (12.2)$$

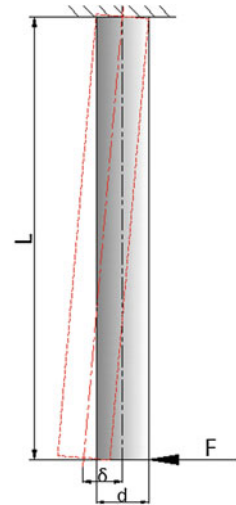
where  $\delta$  is the deflection,  $\sigma$  is the bending stress,  $d$  is the diameter,  $L$  is the length of the micro tool,  $E$  is the Young's modulus for the micro tool material and  $F$  is the force.

### 12.2.1.1 Geometry and Development of Micro Cutting Tool

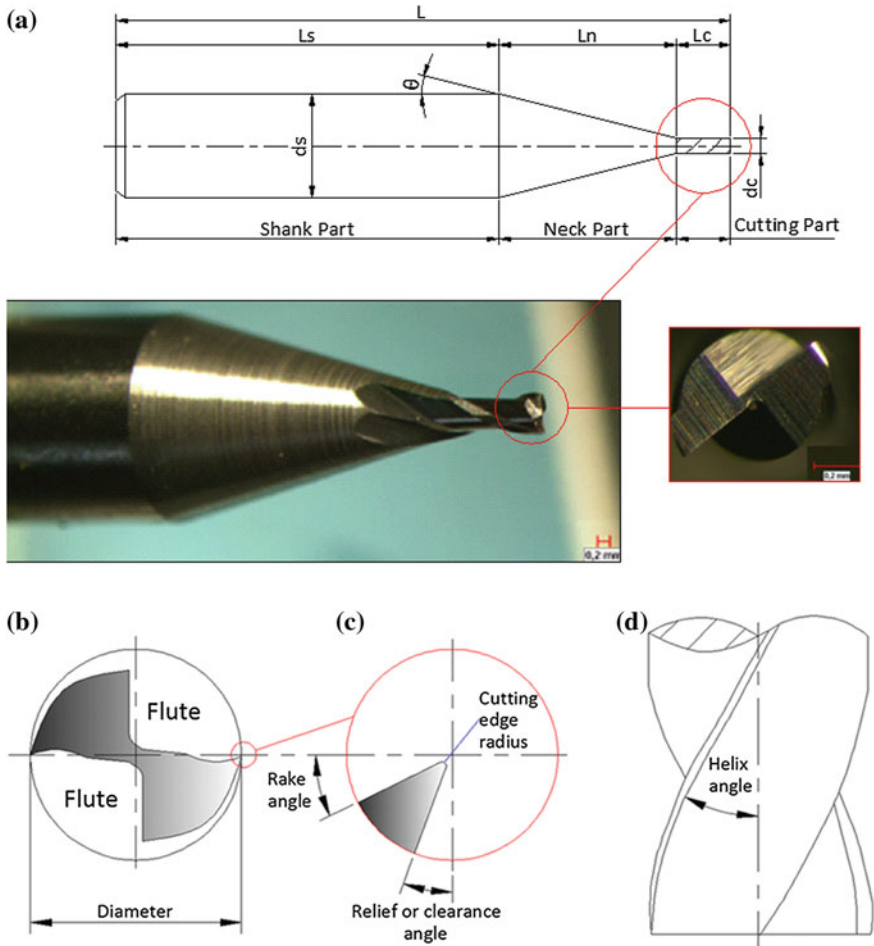
Generally, geometric characteristic of micro milling tools has been adopted from macro milling tools and micro tool is scaled down version of macro cutting tools. Schematic illustration of typical micro end mill [10] is given in Fig. 12.2. The typical geometry of the commercial micro end mill includes three main parts, namely the shank part, the neck part and the cutting part. The shank part connects the cutting tool to the tool holder. The cutting part contains the cutting edges. The role of the neck part is to connect the cutting part with the shank part.

Micro end mills with sharp edges cannot be fabricated with the current manufacturing technology. Currently available micro tools have edge radius ranging from 1 to 5  $\mu\text{m}$ . Rake angle of micro tools is  $0^\circ$  or smaller than  $0^\circ$ . Helix angle affects tool stiffness and chip evacuation therefore a selection of an appropriate helix angle is important at the design of micro tools.  $30^\circ$  helix angle is generally used for commercial micro cutting tools. Although commercial micro cutting tools have generally  $30^\circ$  helix angle, Rahman et al. [11] conducted micro milling experiments using two different helix angles ( $25^\circ$  and  $30^\circ$ ) for pure copper workpiece to determine the effect of helix angle on tool life. Results showed that cutting tool with  $25^\circ$  helix angle had a longer tool life than cutting tool with  $30^\circ$  helix angle. Tooling geometrical effects on the micro milling performance using

**Fig. 12.1** Beam model for elastic deflection of a micro tool



finite element method and experiments were investigated by Wu et al. [10]. Finite element simulations were made under different helix angles ( $0, 15, 30$  and  $45^\circ$ ), rake angles ( $-7, 0$  and  $7^\circ$ ) and cutting edge radius ( $1, 2.5$  and  $4 \mu\text{m}$ ). It was found that the most influential factor on the performance of tool was cutting edge radius, followed by helix angle. Predicted cutting forces, tool stresses, tool temperatures, chip formation and temperatures at different tool geometries were studied. Rising helix angles and cutting edge radius increased cutting forces. However, cutting force changed rather slightly when rake angle varied from negative to positive. The increasing helix angles from  $0$  to  $45^\circ$  induced increasing tool stresses due to weakened cutting edges for cutting tools with high helix angle. Maximum and minimum stress were obtained at the positive rake angle and the negative rake angle, respectively. It is known that in hard milling, negative rake angle is more suitable for maintaining good edge strength. It is also known that a sharp cutting edge can diminish the tool stresses. The  $1 \mu\text{m}$  cutting edge radius gave the least tool stress, followed by  $2.5$  and  $4 \mu\text{m}$  edge radius. Maximum tool temperatures increased with increasing helix angles. A positive rake angle had the most influential effect on tool temperatures. Tool temperatures reduced when a smaller cutting edge radius was used. Cutting edge radius had the most significant effect on the tool temperatures as compared to the helix angle and rake angle. A higher helix angle, positive rake angle and dull cutting edge induced higher tool temperatures.  $30^\circ$  helix angle, positive rake angle and sharp cutting edge were found to be more suitable for the better chip removal ability, machining stability and less power consumption. To maintain edge strength and enhance tool life in micro milling of hard materials, a low helix angle and negative rake angle were suggested. Whereas, in light or finish cutting, a  $30^\circ$  helix angle, positive rake angle and sharp edge corner were more suitable for improving the cutting performance.



**Fig. 12.2** Schematic illustration of (a) a micro cutting tool geometry ( $L$ : tool length,  $L_s$ : shank length,  $L_n$ : neck length,  $L_c$ : cutting length,  $d_s$ : shank diameter,  $d_c$ : cutting diameter,  $\theta$ : neck angle), (b) cross section of a two fluted end mill, (c) enlarged view of a cutting edge, (d) side view of end mill

In the literature, cemented tungsten carbide was used as the tool material due to its high hardness and fracture toughness. Its high electrical conductivity allows to manufacture by EDM. Therefore, Egashira et al. [12] developed cemented tungsten carbide micro tools with a diameter of  $3 \mu\text{m}$  by EDM to conduct micro drilling and micro milling processes. Schaller et al. [13] ground hard metal micro end mills with a diameter of less than  $50 \mu\text{m}$ . Grooves with a width of less than  $50 \mu\text{m}$  for brass and grooves with a width of about  $100 \mu\text{m}$  for stainless steel were machined successfully [13].  $22 \mu\text{m}$  diameter micro milling tools were developed, were fabricated using FIB technique and used for machining of polymethyl methacrylate (PMMA) in the literature [14]. 2, 4 and 6 cutting facets micro end mills having  $\sim 25 \mu\text{m}$  diameters

were made by focused ion beam sputtering by Adams et al. [15]. Experiments were conducted using these micro tools at various workpiece materials, namely, 6061-T4 aluminium, brass, AISI 4340 steel and polymethyl methacrylate. Experimental results showed that machined trench widths were approximately equal to the tool diameters and surface roughnesses at the bottom were  $\sim 200$  nm. Developed micro tools could be used for more than 6 h without fracture [15]. Micro tungsten-carbide tool with a minimum diameter of  $31 \mu\text{m}$  was fabricated by micro-EDM using the WEDG method and this tool was used to produce micro-slot and micro thin-walled structure on Al 6061-T6 materials [16]. Cheng et al. [17] fabricated a polycrystalline diamond micro hexagonal end mill with a diameter of 0.5 mm by wire electrical discharge machining.

In the literature, some micro end mills with new geometry were designed and manufactured. Fang et al. [18] used various types of tool geometry, namely, two-flute end mills (commercial type),  $\Delta$ -type (triangle-based) end mills with a straight body, D-type (semi circle-based) end mills with a straight body,  $\Delta$ -type end mills with a tapered body and D-type end mills with a tapered body. The relative rigidity of these end mills was computed by finite element method (FEM) and it was found that two-flute end mills were 8–12 times weaker than the  $\Delta$ -type and D-type end mills with tapered body. This meant that, under similar experimental conditions there was less deflection for  $\Delta$ -type and D-type end mills with tapered body. Experimental results also showed that machining quality of brass workpiece material with D-type tools was better than that of  $\Delta$ -type end mills.

Ohnishi et al. [19] fabricated various tool shapes micro end mills with a diameter of  $20 \mu\text{m}$  by grinding. Micro grooving experiments were conducted on duralumin workpiece material. Micro tools were named as type a, b, c and d. Type a had one cutting edge, type b had 2 cutting edges with no flank at side cutting edges and relief angle of  $0^\circ$ , type c had the same shape with type b except for relief angle which was  $5^\circ$  and type d had 4 cutting edges. Rake angle of types a, b and c was  $-37^\circ$  and rake angle of type d was  $-45^\circ$ . Type a, having the largest peripheral surface, gave the largest cutting force and burr size and it was followed by types b and c. Type d showed the smallest cutting force and burr size. Tool life was defined as the machined length of the groove by end mill till its breakage and the longest tool life was achieved with type c.

Fleischer et al. [20] developed a single-edged micro milling tool. Stability analyses of developed geometries were carried out by using FEM simulations. Firstly, simulations were done for three different geometries with a diameter of  $300 \mu\text{m}$  to determine the best geometry. Geometry 1 and 2 were trapezium-shaped and Geometry 3 was semi-circular shape. From the results, it was found that the stability of Geometry 3 was 30 % higher than that of Geometry 1 and 2. Therefore, the deflection of Geometry 3 was smaller than the others. Then, the optimized geometry (Geometry 3) was scaled down from 300 to 150, 125, 100, 75, 50 and  $30 \mu\text{m}$ . Micro tools were manufactured using EDM and grooves were machined successfully using these micro tools.

A single edge micro milling tool had been designed and fabricated by using 6-axis wire-cut EDM by Nakamoto et al. [21].

### 12.2.1.2 Effect of Coating

Coatings improve the physical/mechanical properties of cutting tools. In macro milling, appropriate coatings can improve the machining performance and allow the use of higher cutting speeds and dry machining. Diamond cutting tools are often utilized for milling however they cannot use to machine ferrous materials due to the affinity. Therefore, tungsten carbide (WC) is widely used as tool material in micro milling process. In conventional milling, appropriate tool coatings can improve the performance and efficiency of machining process in terms of increased tool life, improved surface quality and prevention the risk of tool breakages.

Diamond coatings for micro end mills can reduce flute clogging by eliminating adhesion of workpiece material to the tool surface, can reduce tool wear rate owing to their high hardness and can reduce cutting forces because of their low friction coefficient against many materials. However, conventional polycrystalline diamond coatings that are currently used in macro cutting tools are too thick (2–100  $\mu\text{m}$ ) for micro cutting tools due to the fact that they increase the cutting edge radius. Increasing the cutting edge radius will negatively affect the performance of micro cutting tool. So, thin coatings are required for micro cutting tools [22].

In the literature, to minimize the cutting edge radius, thin, fine-grained diamond coatings for micro end mills were used [22]. A new approach for coating 300  $\mu\text{m}$  diameter WC micro end mills using a tailored seeding method and hot filament chemical vapour deposition (CVD) was used. It was found that a fine-grained diamond coating reduced the thrust and main cutting forces by about 90 and 75 %, respectively as compared to uncoated cutting tools in micro milling of 6061-T6 aluminium material. Chip adhering was observed when using uncoated cutting tools however, no adhered aluminium was seen for the fine-grained diamond coated tools. The diamond coated tool gave a smoother channel than the uncoated tool. The uncoated tools produced significant burring but diamond coated tools did not give any observable burring. The uncoated tools produced larger continuous chips of various sizes. Whereas, the diamond coated tools produced shorter, more uniform chips.

Torres et al. [23] applied thin (<300 nm) fine-grained diamond (FGD) and nanocrystalline diamond (NCD) coatings using the hot-filament chemical vapor deposition (HF-CVD) process to 300  $\mu\text{m}$  diameter WC micro end mills in order to improve cutting performance. Cutting performance of diamond coated micro tools was compared with uncoated WC micro end mills during slot milling of 6061-T6 aluminium material. NCD coated tools gave the lowest cutting force as compared to FGD coated tools and uncoated tools. Very little difference was observed between the FGD-coated and NCD-coated end mills in terms of produced surface finish and cut chips, therefore comparisons were based on uncoated and diamond-coated cutting tools. In this study, similar results were obtained with reference [22].

Butler-Smith et al. [24] investigated the performance of monocrystalline CVD diamond tools of different crystallographic orientations in micro milling of Ti-6Al-4V. The results showed that the cutting tools with a rake/flank orientation

of  $\{110\}/\{100\}$  extended tool life over the tools having a rake/flank orientation of  $\{100\}/\{110\}$ . However, surface roughness values ( $R_a$  and  $R_z$ ) for two diamond cutting edge orientations were not significantly different.

The performance of different coated tools (TiN, TiCN, TiAlN, CrN and CrTiAlN) was evaluated based on tool wear, surface finish and burr size in micro milling of hardened H13 tool steel (45 HRC). It was found that TiN, TiCN, TiAlN, CrN and CrTiAlN coatings reduced cutting edge chipping and edge radius wear as compared to uncoated micro tools. 41 and 27 % reduction in the average flank wear were observed when using TiN and CrTiAlN coatings, respectively, compared to using uncoated tools. Coating delamination for CrN and TiCN coated micro tools and localised flank wear for TiAlN coated micro tools were observed. All coated cutting tools except for TiAlN, reduced burr size compared to uncoated tools. The results demonstrated that TiN coated tools gave the best performance in terms of tool wear reduction and improvement in machined surface quality [6].

Comparative studies on the performance between CrN and Cr–Si–N coated end mill for micro milling of brass workpiece material were conducted by Shin et al. [25]. Cr–Si–N coatings, in which Si was incorporated into CrN, were deposited on WC–Co substrates using a hybrid coating system of arc ion plating and magnetron sputtering techniques. The hardness of the Cr–Si–N coatings gave a maximum value of 34 GPa at a Si content of 9.3 at. %. Average friction coefficient of Cr–Si–N films decreased with an increase in Si content. Therefore, Cr–Si (9.3 at. %)–N coating with the highest hardness and lower friction coefficient was deposited on WC-based micro end mills by hybrid coating system and cutting performance of this coating was compared with CrN coating. The values of cutting force and flank wear of Cr–Si (9.3 at. %)–N coated tool were much lower than that of CrN coated tool.

Cutting performance of Cr–C–N, Cr–Si–N and Cr–Si–C–N coatings was investigated by Kim et al. in micro milling of brass material [26]. Coatings were deposited on WC–Co substrates by a hybrid coating system combining arc ion plating and sputtering techniques. Cr–Si–C–N coated micro end mill gave the lowest tool wear value.

Micro milling of oxygen-free, high-conductivity copper using tungsten carbide, CVD diamond and single-crystal diamond tools was analyzed. The effect of cutting parameters namely cutting speed, feed rate and depth of cut on surface roughness and burr formation was investigated [27].

Ozel et al. [28] presented experimental study and finite element simulations for micro milling of Ti-6Al-4V alloy using uncoated and CBN coated tungsten carbide micro end mills. Micro cutting tools were coated with CBN using magnetron sputtering system. Surface roughness, burr formation and tool wear were measured experimentally and forces, temperature and wear rate were predicted with finite element modelling. Effects of cutting parameters on measured responses were also investigated. For 635 and 508  $\mu\text{m}$  diameter tools, surface roughness decreased with the increasing cutting speed, however  $\text{Ø}381 \mu\text{m}$  tool showed opposite trend. Increased cutting speed also increased burr. Burr formation was affected

significantly by feed rate but the effect of feed rate was less on surface roughness. Burr width reduced with the increment of feed rate. Lower surface roughness and burr width were achieved with CBN coated cutting tools.

### ***12.2.2 Micro Machine Tools***

Micro cutting tools used in micro milling operations are fragile due to the very small diameter of cutting tools. Therefore, feed rate is very small in micro milling and small feed rate increases machining time. To compensate the machining time, micro milling process requires high speed machines. High spindle speed provides not only necessary cutting speed but also high material removal rate. However, the price of high speed machine is very expensive. To avoid high cost of high speed machine, high speed attachments (or high speed attached spindle) can be used and this spindle is mounted on the taper fit of the main spindle of the machine tool. High speed attachments are driven by brushless motors and a brushless motor is directly connected to the spindle in high speed attachments. The power is supplied to the motor by an external control unit. The basic problem of using high speed attachment is the spindle vibration owing to overhanging. Therefore, tool failure and poor surface quality can occur when this attachment is used.

Various research groups have been interest the building of small scale machine tools to manufacture micro size components. Miniaturization of machine tools has some advantages such as portability and diminishing space, energy, materials and costs. Bang et al. [29] constructed 5-axis micro milling machine for machining micro parts and this machine is composed of micro stages, air spindle and PC-based control board. The overall size of this precision machine was  $294 \times 220 \times 328$  mm (W  $\times$  D  $\times$  H). Developed micro milling machine was capable of manufacturing micro parts such as micro walls, micro columns with rectangular and circular cross-sections, micro impeller and micro blades.

## **12.3 Micro Milling Mechanisms**

Micro milling is not only downscaled versions of conventional milling but also it possesses some distinguished properties such as size effect, chip formation, minimum chip thickness, effect of cutting edge radius and effect of workpiece microstructure. Ignored factors in macro milling such as deflection, workpiece microstructure, tool edge radius, vibration, etc. become significant in micro milling.

In macro milling, the feed per tooth is greater than the cutting tool edge radius and the assumption of a sharp tool that removes the chip without any elastic recovery or ploughing is valid. However in micro milling, because of very small tool edge radius and feed rates, this assumption is no longer valid. A large negative

rake angle in micro milling is due to the comparable of tool edge radius with the chip thickness. This negative rake angle induces elastic recovery of the surface and ploughing.

The effective rake angle can be calculated as following equation:

$$\alpha_e = -\sin^{-1}\left(\frac{2r_e - h}{2r_e}\right) \quad (12.3)$$

where  $r_e$  is the tool edge radius and  $h$  is uncut chip thickness.

In micro milling ploughing easily occurs due to the comparable size of the tool edge radius and feed per tooth. During ploughing the workpiece material is elastically or plastically deformed but no chip is formed. Poor surface quality and short tool life are observed when ploughing occurs. Ploughing can be determined by minimum chip thickness. Ploughing occurs when the uncut chip thickness is less than the minimum chip thickness. When the uncut chip thickness is greater than the minimum chip thickness, chip forms by shearing mechanism as is conventional milling. Thus, the feed per tooth must be selected higher than the minimum chip thickness in order to avoid ploughing. However, the determination of minimum chip thickness is a difficult task. To this end, a ploughing-detection algorithm during micro milling was developed in the literature [30]. This method used the variation in the peak values of the cutting force signal. The proposed method was verified by comparing the surface roughness of the sidewall under ploughing and normal machining conditions. It was found that surface roughness value increased when ploughing occurred [30].

### 12.3.1 Size Effect

A non-linear increase in the specific energy (specific energy is the mechanical energy required to remove a volume of material) with a decrement in the depth of cut is called “size effect” phenomenon in micro milling.

Size effect in micro milling of H13 hardened tool steel was investigated by Aramcharoen and Mativenga [31]. Size effect was observed by studying the effect of the ratio of undeformed chip thickness to the cutting edge radius on specific cutting force, surface finish and burr formation. These researchers also investigated the effect of different micro end mill geometry on product quality (surface finish and burr). The feed and normal cutting force increased with an increment in the ratio of undeformed chip thickness to cutting edge radius. Cutting force increased with an increase in feed per tooth, which is similar to macro milling. Surface roughness decreased with an increase in chip load when the chip thickness was less than the cutting edge radius. At higher feed per tooth, ploughing effect reduced, resulting in less elastic recovery of the workpiece. When chip thickness was greater than the cutting edge radius, surface roughness increased with an increase in feed per tooth, which is same in macro milling. Burr reduced with an



increase in ratio of undeformed chip thickness to cutting edge radius. It was also found that milling modes affected the burr size. Down milling showed larger burr than that of up milling. In order to investigate the effect of cutting edge geometry, the poor quality cutting tools (with chipping at the edge, cracking, chamfer cutting edges and multi-edge) were used in the experiments. Surface finish and burr formation of poor cutting tools were compared to the acceptable rounded edge tools (1.4  $\mu\text{m}$  cutting edge radius). The best surface finish was obtained with the rounded edge tools. Rounded edge tools produced higher burr size than that of chipped and chamfer cutting edge tools.

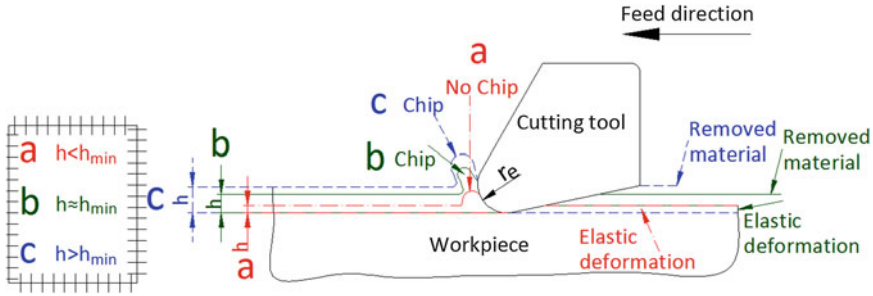
The effect of cutting conditions on the size effect in micro milling of Inconel 718 was experimentally investigated by using an L9 Taguchi orthogonal array. Cutting velocity, the ratio of undeformed chip thickness to cutting edge radius, axial depth of cut and tool coating (TiAlN and uncoated tools) were selected as variables. Analysis of variance (ANOVA) was used to determine the dominant cutting variables for size effect and results showed that not only chip thickness but also cutting velocity was a dominant factor. Cutting velocity and frictional condition (tool coating) were found to be statistically significant parameters affecting specific acoustic emission energy. Cutting velocity and the ratio of undeformed chip thickness to cutting edge radius had significant effect on the surface roughness and burr root thickness [32].

Ding et al. [33] investigated machined surface integrity, part dimension control, size effect and tool wear in micro milling of hardened H13 tool steel. They developed two-dimensional (2D) finite element models with a strain gradient plasticity model.

### ***12.3.2 Chip Formation and Minimum Chip Thickness***

In micro milling, the depth of cut and feed per tooth are very small and comparable to the tool edge radius. No chip can be formed in every pass of the cutting edge below a critical threshold depth of cut value called “minimum chip thickness”. Therefore, the formation of chip depends on the cutting edge radius and its function (minimum chip thickness). Minimum chip thickness is an important parameter, since it determines chip formation and affects cutting process performance in terms of tool wear, cutting forces, burr formation and surface quality.

When the undeformed chip thickness ( $h$ ) is smaller than the minimum chip thickness ( $h_{\text{min}}$ ) (Fig. 12.3a), the workpiece material is forced by the cutting tool and then recovers back after the cutting tool passes over the workpiece. Thus, no material is removed from the surface in the form of a chip due to the elastic deformation. When the undeformed chip thickness is equal to the minimum chip thickness (Fig. 12.3b), chip starts to form however, the removed material is less than the target depth of cut value due to the elastic deformation and recovery. When the undeformed chip thickness is larger than the minimum chip thickness (Fig. 12.3c), the elastic deformation phenomena decreases and material is removed as a chip [31].



**Fig. 12.3** Chip formation in micro scale machining ( $r_e$ : cutting tool edge radius,  $h$ : undeformed chip thickness,  $h_{min}$ : minimum chip thickness)

Son et al. [34] studied the relationship between the friction of a tool-workpiece and the minimum cutting thickness (minimum chip thickness). The minimum cutting thickness is calculated as following equation:

$$h_{min} = r_e \left( 1 - \cos \left( \frac{\pi}{4} - \frac{\beta}{2} \right) \right) \tag{12.4}$$

$$\cos \beta = \frac{1}{\sqrt{1 + \mu^2}} \tag{12.5}$$

where  $\beta$  is the friction angle between a tool and a continuous chip as determined in Eq. 12.5 and  $r_e$  is tool edge radius. Friction coefficient was calculated as the ratio of the tangential force ( $F_T$ ) and the normal force ( $F_N$ ). The friction coefficient of aluminium, brass and oxygen free high conductivity copper (OFHC) were obtained as 0.3, 0.2 and 0.4, respectively. The minimum cutting thickness was predicted using above equation and values for aluminium, brass and OFHC were obtained as 0.1, 0.12 and 0.09  $\mu\text{m}$ , respectively. Cutting forces for all materials increased with the depth of cut. It was also found that a continuous chip generated and the surface quality was the best when cutting at the minimum cutting thickness.

Vogler et al. [35] found that the minimum chip thicknesses were 29–43 % of tool radius for ferrite and 14–25 % of tool radius for pearlite in micro milling. The minimum chip thickness for brass was found to be 22–25 % of tool radius in the literature [36].

The effects of thermal softening, strain hardening, cutting velocity and tool edge radius on minimum chip thickness were taken into account by Liu et al. [37]. The developed model was verified experimentally using 1040 steel and Al6082-T6 workpieces. Normalized minimum chip thickness ( $\lambda_n$ ) was defined as the minimum chip thickness to tool edge radius ratio. In general,  $\lambda_n$  increased with the increasing cutting velocity and tool edge radius when machining 1040 steel, however it was found that neither the cutting velocity nor the tool edge radius had significant effect on  $\lambda_n$  when machining Al6082-T6.

Chip formation of micro milling was investigated by using the finite element simulations in order to find out  $h_{\min}$ . The uncut chip thickness was selected as  $0.1r_c$  ( $r_c$ : tool edge radius,  $0.2 \mu\text{m}$ ),  $0.2r_c$  ( $0.4 \mu\text{m}$ ) and  $0.3r_c$  ( $0.6 \mu\text{m}$ ). It was observed that no chip formed when  $h$  was  $0.1r_c$  and  $0.2r_c$ . From these results, it was proposed that  $h_{\min}$  was  $0.25r_c$  ( $0.5 \mu\text{m}$ ) for OFHC copper when tool edge radius was  $2 \mu\text{m}$  and rake angle was  $10^\circ$  [38, 39]. It was found that the specific energy increased greatly when the uncut chip thickness was smaller than minimum chip thickness. This was attributed to ploughing mechanism and the accumulation of the actual chip thickness [38].

Mian et al. [40] analysed AE signals during micro milling of various workpiece materials (Copper, AISI 1005 steel, Inconel 718, Al6082-T6, AISI 1045 steel and Ti-6Al-4V) to determine the minimum chip thickness value.

Kang et al. [41] experimentally determined the minimum chip thickness by observing the cutting force behaviour. The chip thickness below which the cutting force remained constant was decided as minimum chip thickness. When the undeformed chip thickness exceeded the minimum chip thickness, the cutting force increased. They achieved the minimum chip thickness as  $0.3r_c$  ( $0.28 \mu\text{m}$ ,  $r_c$ : cutting edge radius).

The minimum chip thickness for Al7075 material was found to be approximately  $0.7 \mu\text{m}$  in the literature [42].

The influence of undeformed chip thickness on surface finish, burr formation, cutting forces and tool wear in micro milling of coarse-grained AISI 1045 steel was investigated by Mian et al. [43]. Micro cutting tools deteriorated more rapidly, leading to higher surface roughness and increased burr size at undeformed chip thickness lower than the cutting edge radius. Burr size increased when chip load decreased. Ploughing effect increased specific force, burr size and tool wear.

Malekian et al. [44] modeled the minimum uncut chip thickness taking into account the edge radius and friction angle, based on the minimum energy and infinite shear strain approaches. The proposed model was validated experimentally using Al6061 workpiece. Malekian et al. [44] found that minimum uncut chip thickness was the functions of the edge radius and friction coefficient, which was dependent on the tool geometry and the properties of workpiece material. The minimum uncut chip thickness was determined as about 0.23 of the edge radius.

Summary of researches about the value of the minimum chip thickness is given in Table 12.1.

### 12.3.3 Cutting Edge Radius Effects

Cutting edge radius effect for both macro and micro milling is shown schematically in Fig. 12.4. In macromachining, the cutting edge radius of tools can be negligible and cutting tool is assumed to be sharp because the ratio between the undeformed chip thickness and the cutting edge radius is relatively high. The effective rake angle is positive in macro milling. However, in micro milling, since the ratio between the

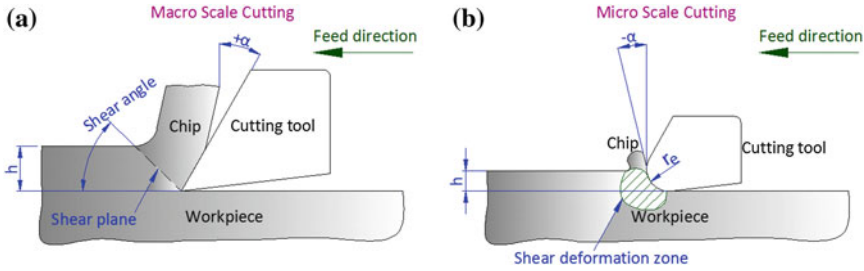
Table 12.1 Researches about the value of the minimum chip thickness

Researchers	Workpiece material	Tool material, geometric properties of tool	Tool diameter ( $\mu\text{m}$ )	Tool edge radius, $r_e$ ( $\mu\text{m}$ )	Estimated minimum chip thickness, $h_{\min}$	Method	Cutting parameters
Vogler et al. [35]	Ferrite Pearlite	No information	508	2–5	29–43 % of $r_e$ 14–25 % of $r_e$	Finite element simulation	Axial depth of cut: 50 and 100 $\mu\text{m}$ Feedrate: 0.25, 0.5, 1, 2 and 3 $\mu\text{m}/\text{flute}$ Spindle speed: 120,000 rpm Axial depth of cut: 12.7 $\mu\text{m}$ Feedrate: 0.375, 0.75 and 1.5 $\mu\text{m}/\text{flute}$
Kim et al. [36]	Brass	Carbide, Helix angle: 30°, Rake angle: 13°	635	3	22–25 % of $r_e$	Indirect method by using cutting force measurement	Spindle speed: 80,000 rpm Axial depth of cut: 100 $\mu\text{m}$ Feedrate: 0.5 and 4 $\mu\text{m}/\text{flute}$ Spindle speed: 84,000 and 150,000 rpm
Liu et al. [37]	Al6082-T6 1040 carbon steel	No information about tool material, Helix angle: 10°, Clearance angle: 10°	508	1–4	38–40 % of $r_e$ 24–35 % of $r_e$	Analytical slip-plasticity model	Spindle speed: 80,000 rpm Axial depth of cut: 100 $\mu\text{m}$ Feedrate: 0.5 and 4 $\mu\text{m}/\text{flute}$ Spindle speed: 84,000 and 150,000 rpm
Lai et al. [38]	Copper (OFCH)	Carbide, Rake angle: 10°	100	2	25 % of $r_e$	Strain gradient plasticity approach and modified Johnson–Cook approach	Axial depth of cut: 6, 10, 15 and 20 $\mu\text{m}$ Feedrate: 0.2, 0.4, 0.6, 0.8, 1 and 1.4 $\mu\text{m}/\text{flute}$ Spindle speed: 20,000 rpm
Li et al. [39]	Copper (OFCH)	Carbide, Helix angle: 30°, Rake angle: 10°	100	2	25 % of $r_e$	Finite element simulation	Axial depth of cut: 6, 10 and 20 $\mu\text{m}$ Feedrate: 24, 48 and 80 mm/min Spindle speed: 20,000, 40,000 and 70,000 rpm

(continued)

Table 12.1 (continued)

Researchers	Workpiece material	Tool material, geometric properties of tool	Tool diameter ( $\mu\text{m}$ )	Tool edge radius, $r_e$ ( $\mu\text{m}$ )	Estimated minimum chip thickness, $h_{\text{min}}$	Method	Cutting parameters
Kang et al. [41]	AISI1045 steel	Carbide, No information about geometric properties of tool	200	0.93	30 % of $r_e$	Using cutting force measurement	Axial depth of cut: 10, 20, 30 and 40 $\mu\text{m}$ Feedrates: 0.5, 1.0, 2.0, 3.0, 4.0 and 5.0 $\mu\text{m}/\text{flute}$ Spindle speed: 120,000 rpm Radial depth of cut: 100 $\mu\text{m}$ Axial depth of cut: 1 mm Feedrate: 0.02, 0.5, 2 and 10 $\mu\text{m}/\text{flute}$ Cutting velocity: 94.2 m/min for Cu, Al6082-T6, AISI 1005 and AISI 1045; 50 m/min for Ti-6Al-4V and 25 m/min for Inconel 718
Mian et al. [40]	Copper (OFCH) Al6082-T6 AISI 1005 steel AISI 1045 steel Ti-6Al-4V Inconel 718	No information about tool material, Helix angle: $0^\circ$	2000	0.40 0.44 0.67 0.84 0.33 0.55	11.2–18 % of $r_e$ 23–38.9 % of $r_e$ 31.3–42 % of $r_e$ 19.7–35.5 % of $r_e$ 19–27.5 % of $r_e$ 18.5–25.5 % of $r_e$	Acoustic emission signals	



**Fig. 12.4** **a** Macro scale and **b** micro scale machining ( $r_e$ : cutting tool edge radius,  $h$ : undeformed chip thickness,  $\alpha$ : rake angle)

undeformed chip thickness and the cutting edge radius is low, the cutting edge radius cannot be omitted and cannot be assumed to be sharp. Cutting edge radius effects play a significant role in micromachining and cannot be negligible. In micro milling, the edge radius of cutting tool is not scaled consistently with the cutting tool diameter and the assumption of the cutting edge is perfectly sharp cannot be valid. The effective rake angle in micro milling becomes negative.

In micromachining, a highly negative rake angle and bluntness of the cutting tool increase the specific cutting forces.

The effect of cutting edge radius in micromachining has been investigated by some researchers by using modelling and experimental methods.

Liang et al. [45] developed three-dimensional microburr formation finite element model for micro milling including the effect of varying tool edge radius (4, 6 and 8  $\mu\text{m}$ ). The simulation results showed that there were three types of burr: entrance, top and exit burr. They explained the formation of entrance burr by the presence of the initial negative shear angle. They also stated that the workpiece was compressed and bulged to the side of entrance until permanent plastic deformation occurred. The morphologies of entrance burr were found to be triangle, ellipse and pin shape. The dimension of top burr was found to be relatively larger than the other burr types and the morphologies of top burr were wavy, serrated, bend and roll-over. The morphologies of exit burrs were roll-over, pin shape and strip shape. The material was pushed out plastically when the tool exited the workpiece edge and the exit burr formed. It was concluded that when the tool edge radius increased from 4 to 8  $\mu\text{m}$ , the dimension of top burr increased greatly [45].

The average effective rake angle can be computed by [46]

$$\alpha_{ave} = -\frac{\pi}{2} + \cos^{-1}\left(1 - \frac{h}{r_e}\right) \quad (12.6)$$

where  $h$  is undeformed chip thickness and  $r_e$  is tool edge radius. From this equation, it can be seen that an increment in tool edge radius induces more negative rake angle therefore, the ploughing mechanism dominates. If feed per tooth is smaller than tool edge radius, chip formation may not occur but the top burrs take place due to the plastic deformation effect. An increment in the tool edge radius

increased the average cutting force in feed direction. Large tool edge radius produced larger top burr due to the ploughing effect. The simulation model was verified by the experiments in micro milling of Al2024-T6 aluminum alloy. It was concluded that the proposed model was suitable for predicting burr morphologies and dimensions in the micro milling [45].

The effect of various tool edge radii (sharp tool, 3.2, 5 and 7  $\mu\text{m}$ ) on the simulated cutting temperature, effective stress and cutting force in micro milling was investigated by numerical simulations. Thermo-mechanical finite element model was developed for micro milling of aluminium alloy Al2024-T6. From simulation results, it was concluded that an increase of the tool edge radius increased the cutting force, while decreased the effective stress and cutting temperature. The maximum effective stress regions were located at the rake face of the micro tool for sharp cutter, at the helix cutting edge corners for micro tool with 3.2  $\mu\text{m}$  edge radii, at the end cutting edge for micro tool with 5  $\mu\text{m}$  edge radii and the flank face of the cutting tool for micro tool with 7  $\mu\text{m}$  edge radii. The experimental verification of the simulation results for temperature was conducted with a high-precision infrared camera. The results of comparison between experiments and predictions were found to be acceptable [47].

#### ***12.3.4 Effect of Workpiece Materials and Microstructure/ Grain Size***

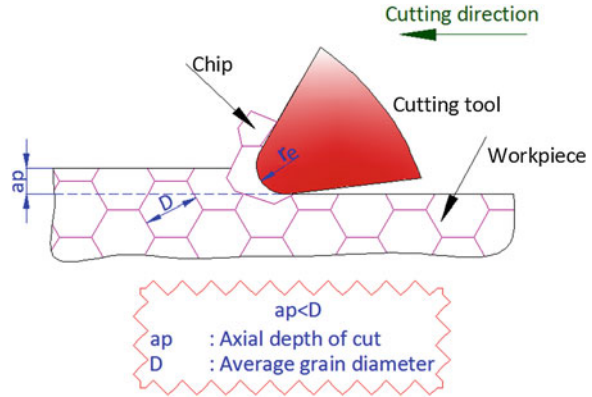
In micromachining, chip formation occurs within a few or even only a grain of the workpiece material (Fig. 12.5), unlike macro milling. The assumption of the microstructure of the workpiece material is homogeneous cannot be valid in micro machining. Cutting performance in micro milling is more dependent to workpiece microstructure than that in macro milling. Therefore, the effects of workpiece material in micro milling were investigated in the literature extensively.

The relationship between grain size and chip thickness size effect in micro-milling was studied by Bissacco et al. [48]. They explained that when shear deformation took place within a single grain, the stress applied to the tool were dependent on individual grain orientations, causing high frequency fluctuations of cutting forces.

Popov et al. [49] investigated the effect of metallurgically and mechanically modified materials on part quality and surface integrity when milling thin features in micro components. It was found that the roughness of micro-features depended on the material grain size and when grain sizes of aluminium alloy were reduced from 100–200  $\mu\text{m}$  to 200 nm, the surface roughness improved more than three times.

Min et al. [50] found a correlation between burr height, critical depth of cut and crystallographic orientation on single crystal materials (copper and magnesium fluoride).

**Fig. 12.5** Grain size in micro milling



Mian et al. [51] studied the micro machinability of a ferrite (AISI 1005) and a near-balanced ferrite/pearlite microstructure (AISI 1045). Higher surface roughness for AISI 1045 was obtained as compared to AISI 1005 steel owing to cutting discontinuities, formation of grain boundary burrs and higher elastic recovery. For both workpiece materials, the best surface was achieved at feed rates closer to the tool edge radius and it was found that surface finish was more sensitive to tool edge radius than material grain size. AISI 1005 steel showed relatively larger burrs in down milling as compared to AISI 1045 steel due to the higher plastic deformation of AISI 1005.

The machining of brittle materials is difficult due to its high brittleness. Brittle materials such as glass, silicon, etc. can be machined in a ductile mode without any surface crack when machining below a critical depth of cut value. In ductile mode machining, material is removed by plastic deformation and surface and subsurface damage can be suppressed.

Morgan et al. [52] demonstrated that polycrystalline diamond (PCD) micro cutting tools could be used in micro milling of glass in ductile mode. PCD tools with conical tips were utilized to machine grooves. It was observed that brittle fractures around the edges of the grooves when the depth of cut was too large.

In the literature, Rusnaldy et al. [53] used the value of  $F_t/F_c$  to determine the ductile cutting mode of silicon. In the micro milling of silicon using diamond coated end mills, the dominant ductile cutting mode was obtained when the force ratio,  $F_t/F_c > 1.0$  [3]. Optimum axial depth of cut, feed rate and spindle speed were found to achieve ductile mode machining. As the axial depth of cut increased, the value of  $F_t/F_c$  decreased. Value of  $F_t/F_c$  rose initially, after that it decreased with increasing feed rate.  $F_t/F_c$  value increased with increasing spindle speed, indicating that cutting regime was in the ductile mode.

The effect of tilt angle along the feed direction on cutting regime transition was investigated in micro milling of glass with micro-ball end mill. Glass was better machined in the ductile regime using  $45^\circ$  tilt angle and feed rates up to 0.32 mm/min [54].



Silicon was machined by cubic boron nitride (CBN) micro ball end mill tool [55]. Effect of workpiece inclination angle in the feed and cross-feed direction on the surface roughness was studied. Surface finish improved when the workpiece was inclined in the feed direction and the best surface finish was obtained at 45° tilt angle. Surface roughness and cutting force increased with an increase in feed rate. Flank wear decreased with an increase in workpiece inclination angle.

## 12.4 Overview of Research in Micro Milling

In recent years, although some difficulties can occur in micro milling, extensive researches about micro milling have been carried out. Approaches to overcome micro milling limitations used in the literature are as follows:

- Coating.
- Cutting fluid.
- Develop hybrid process (laser or vibration assisted).

To reduce friction and associated heat generation at the tool-workpiece interfaces, tool coatings or cutting fluids should be used. The first approach (tool coatings) was explained in Sect. 12.2.1.2.

Cutting fluids not only reduce friction but also improve tool life. Therefore, some studies about using cutting fluids in micro milling were seen in the literature. The micro milling experiments were conducted both with an ester oil for minimum quantity of lubrication (MQL) and without coolant (dry condition). Despite the poor machinability of NiTi due to the high ductility and the strong work hardening, workpiece quality improved and tool life extended when MQL applied. It was found that the optimal cutting ranges for optimal cutting data were very limited for micro milling of NiTi workpiece material. For cutting tool with diameter of 0.4 mm, the depth of cut was restricted to only 10 µm [56].

Performance of MQL was investigated by Li and Chou and tool wear, surface roughness and burr formation were analyzed [57]. The effects of tool materials, oil flow rate and air flow rate on tool performance were also investigated. It was observed that using MQL improved tool life, surface quality and reduced burr formation compared to dry machining. Maximum allowable flank wear for 600 µm diameter tool was found as 80 µm. After that value the surface finish quickly deteriorated. The optimal lubrication condition was determined as oil flow rate of 1.88 ml/h and air flow rate of 40 l/min. It was found that tool life was influenced significantly by air flow rate. The down milling burrs were found to be larger and up milling burrs were smaller. Tool A had a higher content of Co compared to tool B and flank wear of tool A was higher than tool B.

Marcon et al. [58] used a graphite nanoplatelet based cutting fluid in micro-slot milling of H13 tool steel (50 HRC) workpiece material. The effects of graphite size (1 and 15 µm), concentration (0.1 and 0.5 wt %) and flow rate (15 and 100 ml/

min) on the cutting forces, slot depth, tool condition and machined surface were investigated. Graphite nanoplatelet based cutting fluid reduced tangential force but increased axial force as compared to the dry condition. However, no difference in the feed force between cutting fluid and dry condition were observed. 15  $\mu\text{m}$  graphite size showed slightly better performance than 1  $\mu\text{m}$  graphite size. The initial slot depth for the dry and 15  $\mu\text{m}$  graphite size was close to the nominal depth of cut 30  $\mu\text{m}$ , while the depth at the end was  $\sim 1.2\text{--}1.3$   $\mu\text{m}$  lower. In the 1  $\mu\text{m}$  graphite size condition, the initial slot depth was about half the aimed depth while the depth at the end was 2.2  $\mu\text{m}$  lower. Therefore, further experiments were conducted at 15  $\mu\text{m}$  graphite size. Feed and tangential forces were lower at the lower flow rate (15 ml/min). 0.1 wt % graphite concentration gave lowest tangential force.

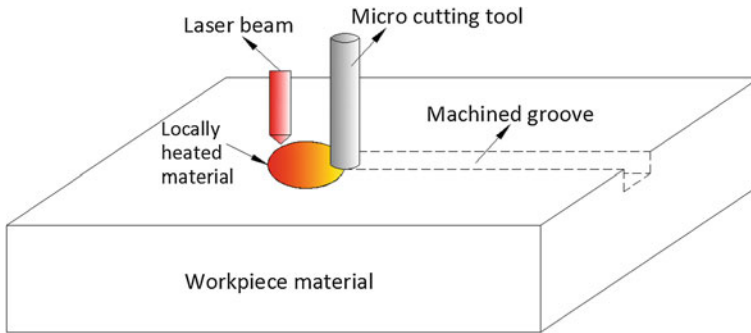
To improve the machinability of hardened tool steels (HRC 55 and HRC 58), two dimensional vibration-assisted micro end milling was applied in the literature [59]. It was found that two dimensional vibration-assisted micro end milling improved the surface roughness and reduced tool wear compared to traditional micro end milling. Larger amplitude and higher frequency were found to be useful in order to improve surface roughness and to reduce tool wear.

A three-dimensional (3D) cutting force model for two dimensional vibration-assisted micro end milling was developed by Ding et al. [60]. Also, a cutting process dynamics of two dimensional vibration-assisted micro end milling was established and this model was combined with the cutting force model. The estimated cutting forces using both models (with and without process dynamics) gave a good agreement with the experimental data. But, model with process dynamics showed close agreement with experimental result.

Ding et al. [61] investigated the influence of the size effect on top burr formation in two dimensional vibration-assisted micro end milling by studying the effects of both the ratio of the undeformed chip thickness to the cutting edge radius and the ratio of the time when the undeformed chip thickness was less than the minimum chip thickness to the total cutting time on top burr formation. It was concluded that the employment of vibration-assisted cutting in micro end milling minimized the size effect and improved the machining performance in terms of reducing the height of the top burr. It was also found that feed per tooth had a significant influence of the height of the top burr.

Laser-assisted micro milling is a machining process which uses a laser to preheat locally the workpiece prior to chip removal (Fig. 12.6) and can improve productivity by reducing the yield strength of workpiece material at the cutting region. To improve the performance of micro milling of difficult-to-cut materials such as hardened tool steels, stainless steels, titanium alloys, laser-assisted micro milling has been used recently.

The effect of laser preheating on cutting forces, specific cutting energy, burr formation, surface finish and temperature in micro milling of 6061-T6 aluminium and 1018 steel was investigated [62]. For Al6061-T6, average cutting and thrust forces reduced by 39 and 36 %, respectively when using laser-assisted micro milling. For 1018 steel, average cutting and thrust forces decreased by 32 and



**Fig. 6** Schematic view of laser-assisted micro milling

56 %, respectively. Specific energy is the mechanical energy required to remove a volume of material and is calculated by

$$u_c = \frac{F_{c,ave} V_c}{d_T t_m f_r} \quad (12.7)$$

where  $F_{c,ave}$  is average cutting force in N,  $V_c$  is cutting velocity in m/s,  $d_T$  is tool diameter in m,  $t_m$  is depth of cut in m and  $f_r$  is feed rate in m/s. Specific cutting energy for Al6061-T6 and 1018 steel decreased by 39 and 32 %, respectively. Results showed that chipload (defined as the maximum uncut chip thickness) and productivity in micro milling could be increased by localized preheating of the workpiece material. Laser-assisted micro milling reduced cutting forces and specific cutting energy, but increased burr formation and surface roughness.

Melkote et al. [63] examined the effect of laser heating on the dimensional accuracy of the feature (groove depth and width) and surface finish in micro milling of hardened A2 tool steel (62 HRC). Groove depth in the presence of laser heating was closer to the set depth of cut than in the absence of laser heating. Groove width was also closer to the nominal value with laser heating. In general, the average surface roughness with laser heating was lower than without laser. Surface roughness increased with an increment in cutting speed in the presence of laser heating. In contrast, the effect of cutting speed on surface roughness was found to be unclear in the absence of laser heating.

Shelton and Shin [64] evaluated the effects of laser-assisted micro milling on surface finish, edge burrs, tool wear and workpiece microstructure for AISI 316, AISI 422, Ti-6Al-4V and Inconel 718. The effect of laser-assisted micro milling on side-cutting surface finish was insignificant. Laser-assisted micro milling reduced edge burrs of Ti-6Al-4V and Inconel 718. Laser-assisted micro milling increased the tool life by approximately four times for AISI 422.

Kumar and Melkote [65] investigated cutting forces, tool wear, material removal rate, burr formation and surface roughness in micro milling of hardened A2 tool steel (62 HRC) with laser. Laser-assisted micro milling had significant advantage on cutting forces, tool wear and material removal rate. The average

reduction in the resultant cutting force was up to 69 % with the presence of laser. When the laser spot size was larger than the tool diameter, the burr height and surface roughness increased in the presence of laser due to the thermal softening of the work material.

Ding et al. [66] focused on numerical modeling of laser-assisted micro milling of Ti6-Al-4V, Inconel 718 and AISI 422 stainless steel and experiments were conducted on these materials in side cutting of bulk and fin workpiece configurations. Results showed that laser-assisted micro milling could eliminate or reduce BUE formation in micro milling of difficult-to-cut materials.

## 12.4.1 Performance Measures

### 12.4.1.1 Tool Wear and Tool Life

Tool wear means that the change of tool shape from its original shape during machining. No unified approach for the evaluation of the tool wear in micro milling is available. Tansel et al. [67] redefined the tool wear in micro milling. Any sense of changes between a new and used tool is defined as tool wear. Three types of breakage in micro milling are as follows according to Tansel et al. [67]:

- Chip clogging due to the removing of chips slowly.
- Breakages owing to the tool wear.
- Excessive stress.

The average reduction in the tool radius,  $w$ , was defined as tool wear by Li et al. [37] and  $w$  was calculated by

$$w = \frac{d_0 - d_{wear}}{2} \quad (12.8)$$

where  $d_0$  is the diameter of the new tool and  $d_{wear}$  is the diameter of the worn tool. Because the tool diameter was very difficult to measure, the widths of the machined channels were used to measure the tool wear. Thus,  $d_0$  was obtained as the channel width at the beginning of experiment, while  $d_{wear}$  was obtained as the channel width when the tool wear occurred.

The dwindling height of the tool edge tip was defined as tool wear by Nakamoto et al. [21]. The quality of machined surfaces with new tools and progressively worn tools was investigated experimentally. Experiments were conducted at WC workpiece material by a poly-crystalline diamond (PCD) tool. It was found that PCD tool wear did not affect the quality of the machined surface.

The change in tool diameter [68], edge radius roundness [69] and flank wear measurement from the bottom view of the micro end mills [6, 51] were also used as the tool wear evaluation criteria in the literature.

Rahman et al. [11] conducted micro milling experiments for pure copper workpiece. They found that the tool failure occurred before the tool was significantly worn out. The reason for this was explained with the highly ductile and malleable properties of pure copper. It was also observed that non-uniform wear happened in both the cutting edges. These authors also investigated the effect of cutting parameters on tool life. It was concluded that the higher the depth of cut, the longer the tool life and tool life decreased with the increasing of cutting speed. Both spiral and broken chips were observed in the micro milling of pure copper and chips could not be seen with naked eyes. The chip size was different from the macro milling, but chip shape was similar to macro milling.

A micro milling tool has very low strength and stiffness. Owing to the low stiffness and very small tool diameter, micro milling tools tend to deflect and can be broken easily. Due to the small diameter of cutting tools, the tool breakage in micro milling cannot be detected visually by the operators. However, to improve the tool life, the monitoring of micro milling process is mandatory.

Some difficulties are encountered in the monitoring of tool condition in micro milling as compared to conventional milling. Firstly, owing to minimum chip thickness effect, observing of chip flow characteristics and vibration in micro machining is difficult. Secondly, problems with miniature size of the components and tools used are seen. Thirdly, the noise in the signal for monitoring micro machining is usually very high and difficult to separate [70]. Since the detection of tool breakage is very difficult task, several attempts that depend on various methods, such as cutting force, vibration and acoustic emission (AE) signals have been made in order to monitor tool condition in micro milling. Kang et al. [71] used AE signals for the indirect monitoring of tool wear in the micro-lens machining process with a micro ball end mill. Jemielniak and Arrazola [72] utilized AE and cutting force signals in micro milling of cold-work tool steel for tool condition monitoring. The total wear in the flank wear  $VB_{B_{max}} = 0.11$  mm was used as the tool life criterion for ball end mills with 400  $\mu\text{m}$  radii [71].

To model tool conditions in micro milling, Zhu et al. [73] presented a discriminant feature selection approach for hidden Markov model. In another study Zhu et al. [70] used a multi-category classification approach for tool flank wear state identification in micro milling and continuous hidden Markov models are used to model of tool wear. Malekian et al. [74] investigated factors affecting tool wear and used various sensors, such as accelerometers, force and acoustic emission sensors in order to monitor tool wear. To determine whether the tool is in good condition or is worn, the neuro-fuzzy method was used.

#### 12.4.1.2 Cutting Force

Since measurement of cutting force during machining provides valuable information about cutting tool condition and represents the state of machining, cutting force is measured by employing table dynamometers or load cells in the literature

to monitor machining process. Knowledge of cutting forces can prevent potential tool breakages and can be useful for improving production rate. However, in micro milling, measurement of cutting force is a challenging task and requires high precision equipment. In macro machining the noises in cutting force signals are very low and the effect of noise can be ignored. Unlike macro machining, the cutting force is low and the noise is high in micro machining.

In the micro milling process, measurement methods and numerical analysis have been employed by many investigators in order to study the cutting force. Tansel et al. [67] investigated the cutting force variation of micro milling for aluminium and steel workpiece materials. When chip clogging occurred cutting forces increased. It was also found that in slot milling operation, the static part of the feed direction force indicated the tool condition.

Uhlmann et al. [75] observed that both an increment of the feed per tooth and cutting speed increased the cutting force during micro milling of sintered tungsten-copper composite materials.

It is known that the vibration amplitude is dependent on the cutting force and this vibration amplitude was used to estimate the cutting force of micro milling in the literature [76].

Kang et al. [41] observed that the increasing of feed per tooth increased the cutting forces during micro milling of AISI 1045 steel. The effect of the axial depth of cut on the cutting force was found to be insignificant.

Banerjee and Bordatchev [77] investigated the effect of circular tool path on the cutting force in micro milling of brass material. A circular tool path was compared to linear tool path. Cutting force measurements were conducted at different tool path radii, tool path rotation angles and feed rates. The difference in the cutting force profiles in micro milling along a circular tool path and a linear tool path was observed. In circular tool path, cutting force profiles changed with the tool path rotation angle. It was observed that at around  $90^\circ$  tool path rotation angle, the correlation between the profiles along a circular tool path and a linear tool path was the highest. The lower tool path radius gave lower cutting force amplitudes due to the lower engagement domain and material removal. Higher cutting force was observed for higher tool path radius greater than the tool radius owing to the complete engagement domain. The increase in the feed rate increased the cutting force.

There are two types of mechanistic force models. In the first model, for each cutting force component, i.e., tangential, radial and axial cutting force, only one specific cutting force coefficient is used. In the second model, two different cutting force coefficients are utilized. These cutting force coefficients ( $k_c$  and  $k_e$ ) include both the shearing and cutting edge forces. The procedure of calculating cutting force with mechanistic force model is given in below:

Firstly, the cutting tool is discretized into thin slices. Then, cutting force applied to each slice is calculated. Lastly, cutting forces for all slices and flutes are summed.

$$\begin{aligned} dF_{t,j}(\phi) &= k_{te}dz + k_{tc}h_j(\phi)dz \\ dF_{r,j}(\phi) &= k_{re}dz + k_{rc}h_j(\phi)dz \\ dF_{a,j}(\phi) &= k_{ae}dz + k_{ac}h_j(\phi)dz \end{aligned} \quad (12.9)$$

where  $k_{tc}$ ,  $k_{rc}$  and  $k_{ac}$  are the cutting force coefficients in the tangential ( $F_t$ ), radial ( $F_r$ ) and axial ( $F_a$ ) directions, respectively and  $k_{te}$ ,  $k_{re}$  and  $k_{ae}$  are the edge force coefficients in the tangential, radial and axial directions, respectively. In macro milling, the uncut chip thickness  $h$  is calculated as:

$$h_j(\phi) = f_t \cos \phi_j \quad (12.10)$$

where  $f_t$  is the feed per tooth and  $\phi$  is the instantaneous cutting angle of the flute  $j$ . However, the uncut chip thickness in micro milling may be different from that of macro milling due to the large ratio of feed per tooth to cutting tool radius. Therefore, some attempts were carried out to calculate the uncut chip thickness in micro milling using the analytical model in the literature. Bao and Tansel [78] calculated the uncut chip thickness using the below analytical model.

$$h = f_t \sin \theta - \frac{N}{2\pi R} f_t^2 \sin \theta \cos \theta + \frac{1}{2f} f_t^2 \cos^2 \theta \quad (12.11)$$

where  $N$  is the number of tool teeth,  $R$  is the cutter radius,  $\theta$  is the tool cutting angle,  $f_t$  is the feed per tooth and  $f$  is the feed rate.

Li et al. [79] proposed the chip thickness equation model for micro milling and the model is given as:

$$h = R \left[ 1 - \sqrt{1 - \frac{2f_t \sin \varphi_i}{R + \frac{Nf_t}{2\pi} \cos \varphi_i} - \frac{f_t^2 \cos 2\varphi_i}{\left(R + \frac{Nf_t}{2\pi} \cos \varphi_i\right)^2} + \frac{f_t^3 \sin \varphi_i \cos^2 \varphi_i}{\left(R + \frac{Nf_t}{2\pi} \cos \varphi_i\right)^3}} \right] \quad (12.12)$$

where  $R$  is the cutter radius,  $N$  is the number of flutes and  $\varphi_i$  is the angular position of the  $i$ th tooth.

A Fourier force model with shearing and plowing cutting mechanism was developed for micro milling by Kang and Zheng [80]. They proposed a new chip thickness expression in Fourier series due to the periodic nature of the milling process and this expression is given as:

$$h = a_0 + \sum_{n=1}^2 (a_n \cos n\theta + b_n \sin n\theta) \quad (12.13)$$

where  $a_0$ ,  $a_n$  and  $b_n$  are the Fourier coefficients.

The tangential, radial and axial forces are transformed into the feed (X), normal (Y) and axial (Z) directions by the following equation:

$$\begin{bmatrix} dF_{x,j}(\phi) \\ dF_{y,j}(\phi) \\ dF_{z,j}(\phi) \end{bmatrix} = \begin{bmatrix} \sin\phi & \cos\phi & 0 \\ -\cos\phi & \sin\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} dF_{t,j}(\phi) \\ dF_{r,j}(\phi) \\ dF_{a,j}(\phi) \end{bmatrix} \quad (12.14)$$

The total cutting force produced by the flute:

$$\begin{aligned} F_x(\phi) &= \int_{z_1}^{z_2} dF_x(\phi) dz \\ F_y(\phi) &= \int_{z_1}^{z_2} dF_y(\phi) dz \\ F_z(\phi) &= \int_{z_1}^{z_2} dF_z(\phi) dz \end{aligned} \quad (12.15)$$

where  $z_1$  and  $z_2$  are the lower and upper axial engagement limits of the flute  $j$  within the cut. To obtain the total  $F_x$ ,  $F_y$  and  $F_z$ , forces acting on all cutting edges are summed:

$$\begin{aligned} F_x(\phi) &= \sum_{j=1}^N F_{xj}(\phi) \\ F_y(\phi) &= \sum_{j=1}^N F_{yj}(\phi) \\ F_z(\phi) &= \sum_{j=1}^N F_{zj}(\phi) \end{aligned} \quad (12.16)$$

where  $N$  is the number of flutes on the cutting tool.

A number of works have presented on the development of analytical and mechanistic models for predicting the forces in micro milling. Vogler et al. [81, 82] developed a mechanistic force model for heterogeneous materials (for ferrite and pearlite, the major components of ductile iron). Zaman et al. [83] developed a three-dimensional analytical cutting force model incorporating the theoretical chip area with the variation in tool rotation angle for micro end milling. The mathematical model was validated experimentally and it was found that the proposed model could be used to simulate the cutting forces at 90 % average accuracy.

Different ways of fitting the cutting force coefficients were investigated during micro milling. The mechanistic force model was validated through experiments and the predicted cutting forces gave good agreement with measured cutting forces [84].

Li et al. [85] developed a three-dimensional cutting force model for micro milling, taking into account the trochoidal trajectory of the tool tip, tool run-out and minimum chip thickness effect. The proposed model was validated by experiments using copper workpiece material. The difference between the predicted and experimental maximum cutting forces was less than 7 %.



A mechanistic cutting force model based on the concept of the partial rake angle was used and cutting force coefficients of the proposed model were independent of the milling cutting conditions. The results showed that the predicted cutting forces were very close to measured cutting forces [86].

Bissacco et al. [87] presented a theoretical cutting force model for micro milling by considering the cutting edge radius size effect, the tool run-out and the deviation of the chip flow angle from the inclination angle. The model was verified experimentally in micro milling of Al6082-T6 workpiece material and it was concluded that predicted and measured forces showed good agreement.

Lai et al. [38] modelled the material strengthening behaviours by using a modified Johnson–Cook constitutive equation. A finite element model for micro scale orthogonal machining process was developed including the material strengthening behaviours, micro tool edge radius and fracture behaviour of the workpiece material. Then, an analytical micro milling force model based on the finite element simulations using the cutting principles and the slip-line theory was developed. Experiments were conducted at OFHC copper workpiece material and good agreements were obtained between the predicted and the experimental results. It was also found that the cutting forces increased with feed per tooth.

Afazov et al. [88] developed a new cutting force model in micro milling of AISI 4340 steel using the finite element model considering the trajectory of the tool, run-out, spindle angular velocity, uncut chip thickness, tool edge radius, rake angle, tool-workpiece contact, chip formation and the thermo-mechanical behaviour of the workpiece material. It was found that the predicted and the measured forces were in very good agreement. The temperature increased with increasing the velocity and the uncut chip thickness. The workpiece material softened at high spindle angular velocities and this resulted the lower cutting force. Lower cutting forces at high spindle angular velocities was also due to the inertia effect. It was found that the cutting forces in the cutting direction decreased by increasing velocity whereas tangential forces were independent from velocity. The forces in tangential and cutting directions slightly increased with increasing the edge radius. This was due to the fact that the contacting length at larger radii was longer and this created more friction.

Li et al. [89] investigated the effect of feed per tooth, depth of cut and width of cut on the force theoretically by force models. In addition, the effect of different milling strategies (up and down milling) and tool paths on the quality of thin features were studied. In micro milling, the cutting forces were modelled using the mechanistic force model used in macro milling. The errors between the predicted forces and measured forces in X- and Y- directions were found to be less than 11 %. The up milling was a preferred strategy for the micro milling of thin features due to the lower average force amplitude as compared to down milling. Force increased with the width of cut and the feed per tooth for up milling. However, the effect of width of cut was found to be more significant than that of feed per tooth. The increase of the depth of cut did not increase the force significantly thus a larger depth of cut value could be utilized to improve productivity without affecting the workpiece quality. For down milling, the amplitude of average force

increased with an increase in width of cut, feed per tooth and depth of cut. Width of cut had the most significant effect on the average force. Thin ribs about 15  $\mu\text{m}$  wide and with an aspect ratio of more than 50 were machined successfully (with good form and surface quality).

A slip-line field model including the temperature, stress, strain-rate and strain hardening effects [90] and finite element model [91] were developed to predict the cutting forces in micro milling of brass 260 workpiece in the literature. When the cutting speed increased from 12 to 25 m/min, the force variation was less than 5 % and thus the effect of cutting speed on the force could be neglected during micro milling of brass 260 material [91].

An analytical force model of micro milling, considering tool run-out, tool deflection, size effect and the entry or exit angles of the tool in the workpiece was developed for up and down milling in the literature [92].

A mechanistic model, considering both the shearing and ploughing dominant cutting mode, was developed to predict micro milling forces of Al7075 material [42]. This model assumes that there is a critical chip thickness that determines whether the cutting regime is shearing or ploughing dominant.

In shearing case, the cutting mechanism is similar to the conventional macro cutting mechanism. When the uncut chip thickness is greater than the minimum chip thickness ( $h > h_{\min}$ ), shearing cutting forces can be modeled as follows:

$$\begin{aligned} dF_{rs} &= [K_{rc}h(\theta_i(z)) + K_{re}]dz \\ dF_{ts} &= [K_{tc}h(\theta_i(z)) + K_{te}]dz \end{aligned} \quad (12.17)$$

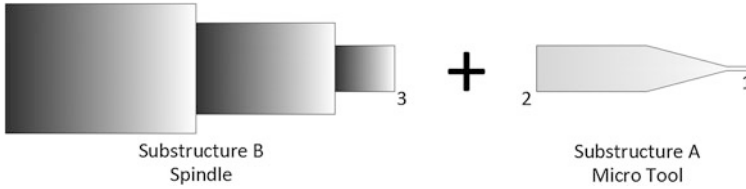
where  $K_{rc}$ ,  $K_{tc}$ ,  $K_{re}$  and  $K_{te}$  are the radial and tangential cutting and edge coefficients, respectively. The cutting coefficients ( $K_{rc}$  and  $K_{tc}$ ) denote shearing of the workpiece and the edge coefficients ( $K_{re}$  and  $K_{te}$ ) denote friction between the cutting tool and the workpiece.

When the undeformed chip thickness is less than the minimum chip thickness, chip formation does not occur due to the ploughing and partial elastic recovery of the material. Shearing mechanism occurs when chip thickness increases due to the rotation. Cutting forces for ploughing dominant cutting regime can be modeled as follows:

$$\begin{aligned} dF_t &= \begin{cases} (K_{tc}h + K_{te})dz & \text{when } h \geq h_{\min}(\text{shearing}) \\ (K_{tp}A_p + K_{te})dz & \text{when } h < h_{\min}(\text{ploughing}) \end{cases} \\ dF_r &= \begin{cases} (K_{rc}h + K_{re})dz & \text{when } h \geq h_{\min}(\text{shearing}) \\ (K_{rp}A_p + K_{re})dz & \text{when } h < h_{\min}(\text{ploughing}) \end{cases} \end{aligned} \quad (12.18)$$

where  $K_{rp}$ ,  $K_{tp}$  are ploughing constants and  $A_p$  is the interference volume.

At very high speeds, the dynamics of micro machining process change due to the centrifugal and gyroscopic effects. Although dynamic modeling of micro end mills is a challenging task as compared to macro milling due to the high speed spindle, the dynamics of micro milling have been investigated by some researchers [93, 94]. Jun et al. [93] developed a dynamic micro end milling model to predict



**Fig. 12.7** Receptance coupling of a spindle and a micro tool

cutting forces and vibrations in the presence of alignment errors at the spindle and manufacturing errors at the cutting edges. Filiz et al. [94] used an analytical model of the transverse vibration of rotating micro end mills in the presence of three-dimensional tilt and rotary axis misalignment.

Chatter is an unstable, self-excited vibration which occurs as a result of an interaction between the dynamics of the machine tool and the workpiece [4]. Regenerative chatter reduces machined surface quality and tool life. Chatter stability can be determined by using stability lobe diagrams. Stability lobe diagrams plot the boundary that separates stable and unstable machining in the form of the axial depth of cut limit versus spindle speed for a given specific step over and workpiece/cutting tool combinations [95].

Unlike macro scale mills, conducting the experimental impact hammer test at the tool tip of micro mills is not possible, owing to the fragility of the small diameter of micro tools [95]. Therefore, to achieve the tool tip dynamics, an indirect method, known as the receptance coupling (RC) method is used [95–98], as shown in Fig. 12.7. RC method mathematically combines the results of experimental modal analysis (EMA) of the spindle and machine tool (i.e. Substructure B) with the result of finite element method of the cutting tool (i.e. Substructure A). The dynamics of tool tip can be determined by using the following equation:

$$G_{11} = \frac{X_1}{F_1} = H_{11} - H_{12}(H_{22} + H_{33})^{-1}H_{21} \quad (12.19)$$

where  $G$  and  $H$  mean the assembled and substructure dynamics, respectively [95].

### 12.4.1.3 Surface Quality

To determine surface quality, surface roughness is an extensively utilized parameter in machining area. Process kinematics, process dynamics, cutting edge geometry, elastic recovery of the workpiece material, minimum chip thickness effect, ploughing and microburr formation affect the machined surface roughness in micro milling.

Vogler et al. [35] investigated the surface generation in the micro milling of both single-phase and multiphase workpiece materials. The surface roughness  $R_a$  at the bottom of slots obtained in single-phase ferrite and pearlite did not increase

with increasing feed rate. It was also found that the axial depth of cut did not have a significant influence on the Ra for single-phase materials. The Ra values for multiphase ductile iron workpieces were found to be larger than the Ra values for single-phase materials. A combination of geometry, minimum chip thickness and burr formation at the grain boundaries affected the surface roughness for multiphase materials.

Uhlmann et al. [75] found that the decreasing of cutting speed induced lower surface roughness values during micro milling of sintered tungsten-copper composite materials.

Surface generation models of micro end milling for both sidewall and floor surfaces were developed [99] and developed models were experimentally calibrated and validated [100]. Experiments were conducted at Al6061-T6 workpiece material. Partial immersion peripheral down milling tests were employed to study of the sidewall surface generation and full immersion tests were conducted to study floor surface generation. Surface roughness models for the sidewall surface and floor surface predicted the sidewall and floor surface roughness with the average error of 10 and <15 %, respectively. Larger edge radius induced higher surface roughness for both the sidewall and floor surfaces. It was also concluded that 3D surface roughness (Sa) increased with the increasing of feed rate.

The effects of tool wear, minimum chip thickness and micro tool geometry on the surface roughness were investigated by Li et al. [37] in the micro end milling of oxygen-free high conductivity copper type 101 (OFHC) workpiece material. A trajectory-based surface roughness model was developed. This model predicted the surface roughness accurately. Tool wear was found to have a great effect on the surface roughness. The effects of the cutting velocity and material removal volume on the tool wear were found to be significant. The higher cutting speed induced an increase in tool wear and surface roughness. It was also found that the depth of cut and feed per tooth had small effect on the tool wear.

Min et al. [8] studied the effects of cutting parameters on the form error of vertical side walls and the surface roughness in step-milling. From the results it was found that down milling gave better surface quality and dimensional accuracy than up milling. For down milling, feed rate significantly affected surface roughness. An increase in form error size was observed with increasing feed rate and depth of cut. But, the effect of feed rate was found to be lower.

3D arithmetical mean deviation of the surface roughness (Sa) was measured in the literature to determine surface quality [101]. Elgiloy<sup>TM</sup> (cobalt-based alloy) was used as a workpiece material and the effect of machining parameters on surface roughness was investigated. It was found that Sa values increased with an increment of feed speed.

#### 12.4.1.4 Burr

Burr formation in micro milling depends on a number of factors such as cutting parameters, workpiece material properties, tool geometry and coating. Burrs can be

classified according to their position: entrance, exit, top and bottom burrs. Burrs induce challenges to assembly and handling. Burrs affect machining process negatively in terms of deburring cost, time and quality of parts. Since removing of burrs from micro parts is very difficult task, burrs are undesirable. In order to avoid the burr formation, machining parameters and tool path strategies should be optimized. Appropriate selected cutting tool geometry can also minimize burr formation.

3-dimensional micro ball end milling operation FEM models on Ti-6Al-4V were used to understand burr formation process. According to simulation, burrs were classified into four types: entrance burr, exit burr, top burr and slot base burr. Also, in micro ball end milling, the correlation between cutting parameters and top burr sizes was analysed. Chips stick to the workpiece in the beginning of tool-workpiece engagement, due to the insufficient tool-chip interaction. Remains of chips evolve into entrance burrs. At the end of the slot, workpiece material to be cut ahead of the tool becomes very thin and it deforms easily. When the tool starts to move out of the slot, enough tool-chip engagement is not guaranteed. Thus, the cutting edge of the tool cannot cut the material down, leading to failure in chip separation, and then exit burr is generated. During each cutting tool pass cycle, the chip generated flows upwards along the rake face of the tool. In the cutting zone, the material tears under large tensile stress. Part of this deformed material is not taken away with the chip and remains along the top of slot walls. Small top burrs start to form and during the following tool pass cycle, these burrs are pushed outwards by tool edge, then top burrs are formed. Burr located slot base occurs due to the insufficient engagement with the chip. According to the simulation, massive slot burr occurs on the slot base, thus leading to an undesired surface quality and increased surface roughness. Slot burr is divided into three types: entrance slot base burr, exit slot base burr and slot side base burr [102].

The size of top burr increased as the axial depth of cut increased. It was also found that top burr on the down milling was larger than that on the up milling. Feed per flute and spindle speed on top burr size had less significant effects than axial depth of cut. At spindle speed higher than 30,000 r/min, burr size increased with an increment in spindle speed [102].

Chern et al. [16] investigated the burr formation in micromachining experimentally and classified into burr four types: primary burr, needle-like burr, feathery burr and minor burr. It was found that both the axial engagement and the feed must be kept at minimum to diminish burr formation. The effect of spindle speed on burr formation was found to be negligible.

#### 12.4.1.5 Temperature

It is known that cutting temperatures in micro milling are lower than that in macro milling owing to the very small chip loads. Machined surface quality and tool wear rate are influenced by the temperature at the tool-chip and tool-workpiece interfaces. Properly controlling temperature during machining can improve the tool life, thus the tool cost can diminish.

The measurement of temperature is a difficult task in micro milling and limited temperature measurement results can be seen in the literature. Wissmiller and Pfefferkorn [103] measured temperatures using an infrared camera during slot milling of Al6061-T6 and 1018 steel with 300  $\mu\text{m}$  diameter two-flute tungsten carbide end mills. Measured temperatures were compared with predictions of a two-dimensional, transient, heat transfer model. It was concluded that as the feed rate increased, the tool temperatures increased. The heat transfer model was in good agreement with the experimental data.

### *12.4.2 Design of Experiments and Optimization Studies*

If cutting conditions in micro milling are not selected properly, cutting tools will break in very short times, even in seconds. Therefore, cutting conditions must be optimized to enhance micro milling performance. However, limited research has been presented for optimization in micro milling process thus far. Also, the effect of cutting parameters on process performance should be investigated.

The effect of spindle speed, feed rate, depth of cut and tool diameter on surface roughness in micro milling of brass was examined by Wang et al. using response surface methodology [104]. ANOVA method was used to determine the main and interaction effects of factors on surface roughness. From ANOVA it was concluded that tool diameter was the most influential factor on surface roughness. In order to analyze the effects of cutting parameters on surface roughness, response surface plots were presented in this study. From these response surface plots some results were obtained. Surface roughness increased with decreasing tool diameter. Because the tool deformation increases when the tool diameter decreases. Surface roughness increased with increasing spindle speed. High spindle speed increased vibration so, the surface quality decreased as the spindle speed increased. Surface roughness increased with increasing feed rate and did not change a lot with depth of cut. These results are the same with the macro milling [104]. Wang et al. [105] used Taguchi method in order to design of experiment. Neural network was developed, trained and used to predict the surface roughness based on the experimental results. The effects of spindle speed, feed rate and depth of cut on surface roughness were also analyzed. It was found that surface roughness decreased with the increase of the spindle speed. Minimum surface roughness value was achieved at the spindle speed of 40,000 rpm. After that value, surface roughness increased with the increase of the spindle speed. The reason of why surface roughness increased when spindle speed  $> 40,000$  rpm was explained with the increased effect of vibration induced by the spindle rotor imbalance. It was also found that surface roughness increased with the increase of depth of cut. Developed neural network model was verified with the validation tests and the maximum prediction error was about 10 %.

The design of experiments was organized as full factorial with five variable factors (spindle speed, depth of cut per pass, channel depth, feed per tooth and coolant) and two levels per variable factor. Ra at the bottom of the micro-channel, micro-channel width dimension and micro-channel shape were selected as the response variables. When the feed per tooth increased, the micro-channel lost the rectangular shape in micro milling of aluminium. The combination of lower feed rate and coolant gave closer micro-channel width to the target of 200  $\mu\text{m}$ . The use of coolant decreased the surface roughness. From ANOVA it was found that micro-channel average width was highly affected by feed per tooth. Ra was mainly influenced by feed per tooth and coolant. Micro-channel manufacturing in copper lacked the intended rectangular shape. It was found that standard machine tool was capable of applying micro milling to manufacture micro-channels [106].

The effects of vibration assisted cutting on micro milling of Al6061-T6 aluminium alloy were investigated. The vibration was provided from the workpiece side by a two-dimensional vibrating worktable. They found that slot oversize, displacement of slot center and slot surface roughness could be improved by using vibration assisted cutting. Vibration assisted cutting increased the tool life when high amplitude and proper frequency were used. Higher frequencies affected tool life negatively. They used also Taguchi method and ANOVA to determine the effects of parameters on slot-width accuracy. Amplitude in X direction, frequency in X direction, spindle speed, feed, frequency in Y direction and amplitude in Y direction were selected as control parameters and slot width was the output. The employment of second directional vibration cutting minimized slot-width oversize. From ANOVA it was found that vibrating amplitude in Y direction had a dominant influence in biaxial vibration cutting [107].

The effects of cutting speed, feed rate, depth of cut, tool diameter and number of flutes on the burr formation in micro milling of stainless steel and aluminium were studied by Lekkala et al. [108]. Experiments were carried out using Taguchi method. Lateral deformation of material, bending and tearing of the chip were found to be dominate burr formation mechanisms. Also, in down milling poisson and rollover burr were observed while in up milling tear burr was observed. It was found that burr height and thickness were affected significantly by the depth of cut and the tool diameter. The burr height decreased with increasing feed rate, tool diameter and number of flutes. The burr height and width of stainless steel were observed to be larger than that of aluminium. In addition to the experimental analysis, an analytical model to predict the height for exit burr was presented. The prediction errors of model were between 0.65 and 25 %.

Cardoso and Davim [109] analyzed the effect of machining strategies (constant overlap spiral, parallel spiral and parallel zigzag) and cutting parameters on surface roughness and optimized the surface roughness and machining time. In this study, a simple method that based on analytical relationship between two important machining parameters (average roughness,  $R_a$  and machining time,  $t$ ) was used. This analytical expression is as follows:

$$f(R_a, t) = \frac{1}{\alpha \frac{R_a}{R_{amax}} + \beta \frac{t}{t_{max}}} \quad (12.20)$$

where  $\alpha$  and  $\beta$  are the weights in percentage of  $R_a$  and  $t$ , respectively and  $\alpha + \beta = 1$ . Different  $\alpha$  (0.8, 0.5, 0.3) and  $\beta$  (0.2, 0.5, 0.7) values were used. The aim was to maximize  $f(R_a, t)$  value to determine the most adequate strategy. In conclusions, best results were obtained with constant overlap spiral strategy.

The optimal result of one response may result in an unacceptable result of another response. Therefore, multi-objective optimization is required and such optimization involves trade-offs.

Thepsonthi and Ozel [110] used multi-objective particle swarm optimization to minimize average surface roughness and burr formation concurrently and to determine optimum machining parameters in micro milling of Ti-6Al-4V titanium alloy. The experiments were organized using Taguchi L9 orthogonal array with three factor three level design. Spindle speed, feed per tooth and axial depth of cut were selected as control parameters. Measured responses were average surface roughness ( $R_a$ ) and total top burr width. Top burr width was defined as a horizontal length of burr from the channel wall in order to measure the burr quantitatively. One side wall was machined with down milling, while the other was machined with up milling. Down and up milling showed different degree of top burr. Therefore, total top burr width was defined as a summation of top burr width measured from down and up milling. The most influential factor on surface roughness was found to be feed per tooth. An increment in feed per tooth resulted in a decrement of surface roughness. The effect of each tested parameters on the total top burr width was not significantly different but axial depth of cut had a slightly higher effect than the other factors. Top burr width of up milling was found to be larger than that of down milling. The minimum surface roughness was obtained at the high levels of spindle speed and feed per tooth and at the medium level of axial depth of cut. The high levels of spindle speed and axial depth of cut and low level of feed per tooth gave minimum total top burr width. Optimum process parameters that satisfy both objective functions in minimizing surface roughness and top burr width were determined as spindle speed of 60,000 rpm, feed per tooth of 0.5  $\mu\text{m}/\text{tooth}$  and axial depth of cut of 60  $\mu\text{m}$ .

Kuram and Ozcelik [111] optimized tool wear, cutting forces and surface roughness simultaneously using grey relational analysis during micro milling of aluminium material with ball nose end mill. The experiments were conducted according to Taguchi L9 orthogonal array with three factor three level design. From multi-objective optimization results it was concluded that the best combination for minimizing the tool wear, cutting forces ( $F_x$  and  $F_y$ ) and surface roughness were spindle speed of 10,000 rpm, feed per tooth of 0.5  $\mu\text{m}/\text{tooth}$  and depth of cut of 50  $\mu\text{m}$ .



## 12.5 Conclusions

Increasing demands for micro parts and components require manufacturing method that is capable of producing three dimensional structures. Laser machining, focused ion beam machining, electrochemical machining and electrodischarge machining are used for manufacturing micro parts. Apart from these manufacturing methods, micro mechanical machining especially micro milling can be used for manufacturing three dimensional structures and variety of workpiece materials. Micro milling is the mechanical cutting process that employ cutting tools with diameters below 1 mm. Generally, geometric characteristic of micro milling tools has been adopted from macro milling tools and tool coating is used to improve machining performance. Some characteristics such as rotation of tool and feed motion of the cutting tool through the workpiece are similar to macro milling, however micro milling process has some distinguished properties. Micro milling process is conducted at high speeds and this requires high speed machine or high speed attachments, which increase the cost. Although micro milling is downscaled version of macro milling, some ignored issues in macro milling such as run-out, tool edge radius, workpiece material grain size and deflection become significant in micro milling. Since the tool diameter is small as compared to macro milling, the phenomenon of size effect arises in micro milling and this phenomenon affects process performance.

Although, micro milling is the scaling down of the macro milling, micro milling is different from macro milling in some aspects and knowledge in macro milling cannot be directly applied to micro milling. Therefore, literatures associated with micro milling have been presented in this chapter. From literature, it was seen that micro products made from a great variety of engineering materials such as polymer, steel, aluminium, copper, brass, silicon, glass and composites were machined successfully by using micro milling process. Tool wear, cutting force, surface roughness, burr formation and temperature were investigated as output responses in these studies. Some attempts about optimization in micro milling can be also seen in the literature.

**Acknowledgments** The authors would like to thank Gebze Institute of Technology for supporting the project (Project Number: GYTE BAP 2012-A19).

## References

1. Masuzawa T, Tönshoff HK (1997) Three-dimensional micromachining by machine tools. *CIRP Ann Manuf Technol* 46:621–628
2. Masuzawa T (2000) State of the art of micromachining. *CIRP Ann Manuf Technol* 49:473–488
3. Liu X, DeVor RE, Kapoor SG, Ehmann KF (2004) The mechanics of machining at the microscale: assessment of the current state of the science. *J Manuf Sci Eng* 126:666–678

4. Chae J, Park SS, Freiheit T (2006) Investigation of micro-cutting operations. *Int J Mach Tools Manuf* 46:313–332
5. Dornfeld D, Min S, Takeuchi Y (2006) Recent advances in mechanical micromachining. *CIRP Ann Manuf Technol* 55:745–768
6. Aramcharoen A, Mativenga PT, Yang S, Cooke KE, Teer DG (2008) Evaluation and selection of hard coatings for micro milling of hardened tool steel. *Int J Mach Tools Manuf* 48:1578–1584
7. Popov K, Dimov S, Pham DT, Ivanov A (2006) Micromilling strategies for machining thin features. *Proc Inst Mech Eng Part C: J Mech Eng Sci* 220:1677–1684
8. Min S, Sangermann H, Mertens C, Dornfeld D (2008) A study on initial contact detection for precision micro-mold and surface generation of vertical side walls in micromachining. *CIRP Ann Manuf Technol* 57:109–112
9. Bourne KA, Jun MBG, Kapoor SG, DeVor RE (2008) An acoustic emission-based method for determining contact between a tool and workpiece at the microscale. *J Manuf Sci Eng* 130:031101-1–8
10. Wu T, Cheng K, Rakowski R (2012) Investigation on tooling geometrical effects of micro tools and the associated micro milling performance. *Proc Inst Mech Eng Part B: Eng Manuf* 226:1442–1453
11. Rahman M, Kumar AS, Prakash JRS (2001) Micro milling of pure copper. *J Mater Process Technol* 116:39–43
12. Egashira K, Hosono S, Takemoto S, Masao Y (2011) Fabrication and cutting performance of cemented tungsten carbide micro-cutting tools. *Precis Eng* 35:547–553
13. Schaller Th, Bohn L, Mayer J, Schubert K (1999) Microstructure grooves with a width of less than 50  $\mu\text{m}$  cut with ground hard metal micro end mills. *Precis Eng* 23:229–235
14. Friedrich CR (2000) Near-cryogenic machining of polymethyl methacrylate for micromilling tool development. *Mater Manuf Processes* 15:667–678
15. Adams DP, Vasile MJ, Benavides G, Campbell AN (2001) Micromilling of metal alloys with focused ion beam-fabricated tools. *Precis Eng* 25:107–113
16. Chern G-L, Wu Y-JE, Cheng J-C, Yao J-C (2007) Study on burr formation in micro-machining using micro-tools fabricated by micro-EDM. *Precis Eng* 31:122–129
17. Cheng X, Wang Z, Nakamoto K, Yamazaki K (2011) A study on the micro tooling for micro/nano milling. *Int J Adv Manuf Technol* 53:523–533
18. Fang FZ, Wu H, Liu XD, Liu YC, Ng ST (2003) Tool geometry study in micromachining. *J Micromech Microeng* 13:726–731
19. Ohnishi O, Onikura H, Min S-K, Aziz M, Tsuruoka S (2007) Characteristics of grooving by micro end mills with various tool shapes and approach to their optimal shape. *Mem Fac Eng Kyushu Univ* 67:143–151
20. Fleischer J, Deuchert M, Ruhs C, Kühlewein C, Halvadjiysky G, Schmidt C (2008) Design and manufacturing of micro milling tools. *Microsyst Technol* 14:1771–1775
21. Nakamoto K, Katahira K, Ohmori H, Yamazaki K, Aoyama T (2012) A study on the quality of micro-machined surfaces on tungsten carbide generated by PCD micro end-milling. *CIRP Ann Manuf Technol* 61:567–570
22. Heaney PJ, Sumant AV, Torres CD, Carpick RW, Pfefferkorn FE (2008) Diamond coatings for micro end mills: Enabling the dry machining of aluminum at the micro-scale. *Diam Relat Mater* 17:223–233
23. Torres CD, Heaney PJ, Sumant AV, Hamilton MA, Carpick RW, Pfefferkorn FE (2009) Analyzing the performance of diamond-coated micro end mills. *Int J Mach Tools Manuf* 49:599–612
24. Butler-Smith PW, Axinte DA, Limvachirakom V (2010) Preliminary study of the effects of crystal orientation of a CVD monocrystalline diamond in micromilling of Ti-6Al-4 V. *Proc Inst Mech Eng Part B: Eng Manuf* 224:1305–1312
25. Shin SH, Kim MW, Kang MC, Kim KH, Kwon DH, Kim JS (2008) Cutting performance of CrN and Cr-Si-N coated end-mill deposited by hybrid coating system for ultra-high speed micro machining. *Surf Coat Technol* 202:5613–5616

26. Kim MW, Tak HS, Kang MC, Kim KH, Park ID, Je TJ (2009) Cutting performance of nanocomposite Cr-C-N, Cr-Si-N and Cr-Si-C-N coated tools for micro end-milling operation. *Curr Appl Phys* 9:201–204
27. Huo D, Cheng K (2010) Experimental investigation on micromilling of oxygen-free, high-conductivity copper using tungsten carbide, chemistry vapour deposition, and single-crystal diamond micro tools. *Proc Inst Mech Eng Part B: Eng Manuf* 224:995–1003
28. Ozel T, Thepsonthi T, Ulutan D, Kaftanoglu B (2011) Experiments and finite element simulations on micro-milling of Ti-6Al-4 V alloy with uncoated and cBN coated micro-tools. *CIRP Ann Manuf Technol* 60:85–88
29. Bang Y-B, Lee K-M, Oh S (2005) 5-axis micro milling machine for machining micro parts. *Int J Adv Manuf Technol* 25:888–894
30. Yun HT, Heo S, Lee MK, Min B-K, Lee SJ (2011) Ploughing detection in micromilling processes using the cutting force signal. *Int J Mach Tools Manuf* 51:377–382
31. Aramcharoen A, Mativenga PT (2009) Size effect and tool geometry in micromilling of tool steel. *Precis Eng* 33:402–407
32. Mian AJ, Driver N, Mativenga PT (2011) Identification of factors that dominate size effect in micro-machining. *Int J Mach Tools Manuf* 51:383–394
33. Ding H, Shen N, Shin YC (2011) Experimental evaluation and modeling analysis of micromilling of hardened H13 tool steels. *J Manuf Sci Eng* 133:041007-1–11
34. Son SM, Lim HS, Ahn JH (2005) Effect of the friction coefficient on the minimum cutting thickness in micro cutting. *Int J Mach Tools Manuf* 45:529–535
35. Vogler MP, DeVor RE, Kapoor SG (2004) On the modeling and analysis of machining performance in micro-endmilling, Part I: Surface generation. *J Manuf Sci Eng* 126:685–694
36. Kim CJ, Mayor JR, Ni J (2004) A static model of chip formation in microscale milling. *J Manuf Sci Eng* 126:710–718
37. Liu X, DeVor RE, Kapoor SG (2006) An analytical model for the prediction of minimum chip thickness in micromachining. *J Manuf Sci Eng* 128:474–481
38. Lai X, Li H, Li C, Lin Z, Ni J (2008) Modelling and analysis of micro scale milling considering size effect, micro cutter edge radius and minimum chip thickness. *Int J Mach Tools Manuf* 48:1–14
39. Li H, Lai X, Li C, Feng J, Ni J (2008) Modelling and experimental analysis of the effects of tool wear, minimum chip thickness and micro tool geometry on the surface roughness in micro-end-milling. *J Micromech Microeng* 18:025006 (12 pp)
40. Mian AJ, Driver N, Mativenga PT (2011) Estimation of minimum chip thickness in micro-milling using acoustic emission. *Proc Inst Mech Eng Part B: Eng Manuf* 225:1535–1551
41. Kang I-S, Kim J-S, Seo Y-W (2011) Investigation of cutting force behaviour considering the effect of cutting edge radius in the micro-scale milling of AISI 1045 steel. *Proc Inst Mech Eng Part B: Eng Manuf* 225:163–171
42. Park SS, Malekian M (2009) Mechanistic modeling and accurate measurement of micro end milling forces. *CIRP Ann Manuf Technol* 58:49–52
43. Mian AJ, Driver N, Mativenga PT (2009) Micromachining of coarse-grained multi-phase material. *Proc Inst Mech Eng Part B: Eng Manuf* 223:377–385
44. Malekian M, Mostofa MG, Park SS, Jun MBG (2012) Modeling of minimum uncut chip thickness in micro machining of aluminum. *J Mater Process Technol* 212:553–559
45. Liang YC, Yang K, Bai QS, Chen JX, Wang B (2009) Modeling and experimental analysis of microburr formation considering tool edge radius and tool-tip breakage in microend milling. *J Vac Sci Technol B* 27:1531–1535
46. Fang FZ, Liu YC (2004) On minimum exit-burr in micro cutting. *J Micromech Microeng* 14:984–988
47. Yang K, Liang Y-C, Zheng K-N, Bai Q-S, Chen W-Q (2011) Tool edge radius effect on cutting temperature in micro-end-milling process. *Int J Adv Manuf Technol* 52:905–912
48. Bissacco G, Hansen HN, Chiffre De (2005) Micromilling of hardened tool steel for mould making applications. *J Mater Process Technol* 167:201–207

49. Popov K, Dimov S, Pham DT, Minev R, Rosochowski A, Olejnik L, Richert M (2006) The effects of material microstructure in micro-milling. In: Second international conference on multi-material micro manufacture, pp 127–13
50. Min S, Dornfeld D, Inasaki I, Ohmori H, Lee D, Deichmueller M, Yasuda T, Niwa K (2006) Variation in machinability of single crystal materials in micromachining. *CIRP Ann Manuf Technol* 55:103–106
51. Mian AJ, Driver N, Mativenga PT (2010) A comparative study of material phase effects on micro-machinability of multiphase materials. *Int J Adv Manuf Technol* 50:163–174
52. Morgan CJ, Vallance RR, Marsh ER (2004) Micro machining glass with polycrystalline diamond tools shaped by micro electro discharge machining. *J Micromech Microeng* 14:1687–1692
53. Rusnaldy Ko TJ, Kim HS (2008) An experimental study on microcutting of silicon using a micromilling machine. *Int J Adv Manuf Technol* 39:85–91
54. Foy K, Wei Z, Matsumura T, Huang Y (2009) Effect of tilt angle on cutting regime transition in glass micromilling. *Int J Mach Tools Manuf* 49:315–324
55. Arif M, Rahman M, San WY (2012) An experimental investigation into micro ball end-milling of silicon. *J Manuf Processes* 14:52–61
56. Weinert K, Petzoldt V (2008) Machining NiTi micro-parts by micro-milling. *Mater Sci Eng, A* 481–482:672–675
57. Li K-M, Chou S-Y (2010) Experimental evaluation of minimum quantity lubrication in near micro-milling. *J Mater Process Technol* 210:2163–2170
58. Marcon A, Melkote S, Kalaitzidou K, DeBra D (2010) An experimental evaluation of graphite nanoplatelet based lubricant in micro-milling. *CIRP Ann Manuf Technol* 59:141–144
59. Ding H, Ibrahim R, Cheng K, Chen S-J (2010) Experimental study on machinability improvement of hardened tool steel using two dimensional vibration-assisted micro-end-milling. *Int J Mach Tools Manuf* 50:1115–1118
60. Ding H, Chen S-J, Cheng K (2010) Two-dimensional vibration-assisted micro end milling: cutting force modelling and machining process dynamics. *Proc Inst Mech Eng Part B: Eng Manuf* 224:1775–1783
61. Ding H, Chen S-J, Ibrahim R, Cheng K (2011) Investigation of the size effect on burr formation in two-dimensional vibration-assisted micro end milling. *Proc Inst Mech Eng Part B: Eng Manuf* 225:2032–2039
62. Jeon Y, Pfefferkorn F (2008) Effect of laser preheating the workpiece on micro end milling of metals. *J Manuf Sci Eng* 130:011004-1–9
63. Melkote S, Kumar M, Hashimoto F, Lahoti G (2009) Laser assisted micro-milling of hard-to-machine materials. *CIRP Ann Manuf Technol* 58:45–48
64. Shelton JA, Shin YC (2010) Comparative evaluation of laser-assisted micro-milling for AISI 316, AISI 422, Ti-6Al-4 V and Inconel 718 in a side-cutting configuration. *J Micromech Microeng* 20:075012 (12 pp)
65. Kumar M, Melkote SN (2012) Process capability study of laser assisted micro milling of a hard-to-machine material. *J Manuf Processes* 14:41–51
66. Ding H, Shen N, Shin YC (2012) Thermal and mechanical modeling analysis of laser-assisted micro-milling of difficult-to-machine alloys. *J Mater Process Technol* 212:601–613
67. Tansel I, Rodriguez O, Trujillo M, Paz E, Li W (1998) Micro-end-milling—I. Wear and breakage. *Int J Mach Tools Manuf* 38:1419–1436
68. Filiz S, Conley CM, Wasserman MB, Ozdoganlar OB (2007) An experimental investigation of micro-machinability of copper 101 using tungsten carbide micro-endmills. *Int J Mach Tools Manuf* 47:1088–1100
69. Lee K, Dornfeld DA (2005) Micro-burr formation and minimization through process control. *Precis Eng* 29:246–252
70. Zhu KP, Wong YS, Hong GS (2009) Multi-category micro-milling tool wear monitoring with continuous hidden Markov models. *Mech Syst Signal Process* 23:547–560

71. Kang IS, Kim JS, Kang MC, Lee KY (2008) Tool condition and machined surface monitoring for micro-lens array fabrication in mechanical machining. *J Mater Process Technol* 201:585–589
72. Jemielniak K, Arrazola PJ (2008) Application of AE and cutting force signals in tool condition monitoring in micro-milling. *CIRP J Manuf Sci Technol* 1:97–102
73. Zhu KP, Hong GS, Wong YS (2008) A comparative study of feature selection for hidden Markov model-based micro-milling tool wear monitoring. *Mach Sci Technol* 12:348–369
74. Malekian M, Park SS, Jun MBG (2009) Tool wear monitoring of micro-milling operations. *J Mater Process Technol* 209:4903–4914
75. Uhlmann E, Piltz S, Schauer K (2005) Micro milling of sintered tungsten-copper composite materials. *J Mater Process Technol* 167:402–407
76. Huang BW, Cai JZ, Hsiao WL (2010) Cutting force estimation in a micromilling process. *Proc Inst Mech Eng Part B: Eng Manuf* 224:1615–1619
77. Banerjee A, Bordatchev EV (2012) Effect of circular tool path on cutting force profile in micro-end-milling. *Proc Inst Mech Eng Part C: J Mech Eng Sci* 226:1589–1600
78. Bao WY, Tansel IN (2000) Modeling micro-end-milling operations. Part I: analytical cutting force model. *Int J Mach Tools Manuf* 40:2155–2173
79. Li HZ, Liu K, Li XP (2001) A new method for determining the undeformed chip thickness in milling. *J Mater Process Technol* 113:378–384
80. Kang Y-H, Zheng CM (2012) Fourier analysis for micro-end-milling mechanics. *Int J Mech Sci* 65:105–114
81. Vogler MP, DeVor RE, Kapoor SG (2003) Microstructure-level force prediction model for micro-milling of multi-phase materials. *J Manuf Sci Eng* 125:202–209
82. Vogler MP, Kapoor SG, DeVor RE (2004) On the modeling and analysis of machining performance in micro-endmilling, Part II: cutting force prediction. *J Manuf Sci Eng* 126:695–705
83. Zaman MT, Senthil Kumar A, Rahman M, Sreeram S (2006) A three-dimensional analytical cutting force model for micro end milling operation. *Int J Mach Tools Manuf* 46:353–366
84. Pérez H, Vizán A, Hernandez JC, Guzmán M (2007) Estimation of cutting forces in micromilling through the determination of specific cutting pressures. *J Mater Process Technol* 190:18–22
85. Li C, Lai X, Li H, Ni J (2007) Modeling of three-dimensional cutting forces in micro-end-milling. *J Micromech Microeng* 17:671–678
86. Lee HUI, Cho D-W, Ehmann KF (2008) A mechanistic model of cutting forces in micro-end-milling with cutting-condition-independent cutting force coefficients. *J Manuf Sci Eng* 130:031102-1–9
87. Bissacco G, Hansen HN, Slunsky J (2008) Modelling the cutting edge radius size effect for force prediction in micro milling. *CIRP Ann Manuf Technol* 57:113–116
88. Afazov SM, Ratchev SM, Segal J (2010) Modelling and simulation of micro-milling cutting forces. *J Mater Process Technol* 210:2154–2162
89. Li P, Zdebski D, Langen HH, Hoogstrate AM, Oosterling JAJ, Munnig Schmidt RH, Allen DM (2010) Micromilling of thin ribs with high aspect ratios. *J Micromech Microeng* 20:115013 (10 pp)
90. Altintas Y, Jin X (2011) Mechanics of micro-milling with round edge tools. *CIRP Ann Manuf Technol* 60:77–80
91. Jin X, Altintas Y (2012) Prediction of micro-milling forces with finite element method. *J Mater Process Technol* 212:542–552
92. Rodríguez P, Labarga JE (2013) A new model for the prediction of cutting forces in micro-end-milling operations. *J Mater Process Technol* 213:261–268
93. Jun MBG, Liu X, DeVor RE, Kapoor SG (2006) Investigation of the dynamics of microend milling-Part I: model development. *J Manuf Sci Eng* 128:893–900
94. Filiz S, Ozdoganlar OB, Romero LA (2008) An analytical model for micro-endmill dynamics. *J Vib Control* 14:1125–1150

95. Park SS, Rahnama R (2010) Robust chatter stability in micro-milling operations. *CIRP Ann Manuf Technol* 59:391–394
96. Mascardelli BA, Park SS, Freiheit T (2008) Substructure coupling of microend mills to aid in the suppression of chatter. *J Manuf Sci Eng* 130:011010-1–12
97. Rahnama R, Sajjadi M, Park SS (2009) Chatter suppression in micro end milling with process damping. *J Mater Process Technol* 209:5766–5776
98. Malekian M, Park SS, Jun MBG (2009) Modeling of dynamic micro-milling cutting forces. *Int J Mach Tools Manuf* 49:586–598
99. Liu X, DeVor RE, Kapoor SG (2007) Model-based analysis of the surface generation in microendmilling—Part I: model development. *J Manuf Sci Eng* 129:453–460
100. Liu X, DeVor RE, Kapoor SG (2007) Model-based analysis of the surface generation in microendmilling—Part II: experimental validation and analysis. *J Manuf Sci Eng* 129:461–469
101. Zhang P, Wang B, Liang Y, Jackson MJ (2011) Experimental study on the surface roughness of micromilled Elgiloy<sup>TM</sup>. *Proc Inst Mech Eng Part B: Eng Manuf* 225:2138–2143
102. Chen MJ, Ni HB, Wang ZJ, Jiang Y (2012) Research on the modeling of burr formation process in micro-ball end milling operation on Ti-6Al-4 V. *Int J Adv Manuf Technol* 62:901–912
103. Wissmiller DL, Pfeifferkorn FE (2009) Micro end mill tool temperature measurement and prediction. *J Manuf Processes* 11:45–53
104. Wang W, Kweon SH, Yang SH (2005) A study on roughness of the micro-end-milled surface produced by a miniaturized machine tool. *J Mater Process Technol* 162–163:702–708
105. Wang J, Gong Y, Shi J, Abba G (2009) Surface roughness prediction in micromilling using neural networks and Taguchi's design of experiments. *Industrial Technology, ICIT 2009, IEEE International Conference*, pp 1–6
106. Vázquez E, Rodríguez CA, Elías-Zúñiga A, Ciurana J (2010) An experimental analysis of process parameters to manufacture metallic micro-channels by micro-milling. *Int J Adv Manuf Technol* 51:945–955
107. Chern G-L, Chang Y-C (2006) Using two-dimensional vibration cutting for micro-milling. *Int J Mach Tools Manuf* 46:659–666
108. Lekkala R, Bajpai V, Singh RK, Joshi SS (2011) Characterization and modeling of burr formation in micro-end milling. *Precis Eng* 35:625–637
109. Cardoso P, Davim JP (2010) Optimization of surface roughness in micromilling. *Mater Manuf Processes* 25:1115–1119
110. Thepsonthi T, Ozel T (2012) Multi-objective process optimization for micro-end milling of Ti-6Al-4 V titanium alloy. *Int J Adv Manuf Technol* 63:903–914
111. Kuram E, Ozcelik B (2013) Multi-objective optimization using Taguchi based grey relational analysis for micro milling of Al7075 material with ball nose end mill. *Measurement*. doi:[10.1016/j.measurement.2013.02.002](https://doi.org/10.1016/j.measurement.2013.02.002)

# Chapter 13

## Digital Image Processing in Machining

Samik Dutta, Surjya K. Pal and Ranjan Sen

**Abstract** This chapter speaks about the application of digital image processing in conventional machining. Advantages and disadvantages of digital image processing techniques over the other sensors used in machining for product quality improvement is also discussed here. A short introduction to image processing techniques used in machining is presented here. A detailed review of image processing applications in machining for over the past decade is discussed in this chapter. Also, an example of an image texture analysis method utilized for cutting tool condition detection through machined surface images is presented. An overall conclusion leading to future work required in this field has been mentioned.

### 13.1 Introduction

Digital image processing is a tool of machine vision technique which is used to extract useful information from real physical objects. It helps to take a decision from the sensed images in a non-invasive way with minimal human intervention. Digital image is a two dimensional representation of a scene where luminance of objects in that scene are assigned in each picture element or pixel. Thus digital image is a matrix and digital image processing is a collection of algorithms applied on that image or image matrix to extract useful information of the captured scene. Machine vision system consists of image sensor, image processing algorithm and pattern recognition tool. Digital image processing or machine vision sensors have

---

S. Dutta · R. Sen  
Precision Engineering and Metrology Lab, CSIR-Central Mechanical Engineering  
Research Institute, Durgapur 713209, India

S. K. Pal (✉)  
Mechanical Engineering Department, Indian Institute of Technology,  
Kharagpur 721302, India  
e-mail: skpal@mech.iitkgp.ernet.in

wide industrial application from micro-level surface inspection to macro-level robot navigation. A wide application of machine vision system is addressed in the field of manufacturing industry for non-invasive inspection of manufactured products. In manufacturing, machine vision systems are applied for inspection of metal injection moulded parts [1], casting defects of continuous cast surface [2, 3], inspection of friction stir welded and seam welded zone [4, 5], sheet metal formed parts [6], monitoring and control of rolling process [7] and most widely used in tool condition monitoring of conventional machining [8].

The principal metal removal process in manufacturing is the conventional machining. Turning, milling and drilling are three primary important conventional machining processes used in industry. High product quality resulting from less costly machining is the primary concern to industries. Aiming these requirements, unmanned machining in flexible manufacturing system is now the ultimate target for manufacturing industries which reduces the machine tool downtime. Tool condition monitoring (TCM) technique is the only way to fulfil these requirements. A complete TCM system can accomplish the monitoring of tool and workpiece alignment, cutting tool wear and breakage monitoring, machining parameters adjustment, monitoring chip congestion, corrective actions towards adverse machining and unmanned machining. The cost benefits of using TCM in machining are reduction of cutting tool cost by monitoring of progressive tool wear, reduction of product cost by reducing downtime, maintenance cost by monitoring of machine tool and cutting tool [9]. In a TCM system, machining process parameters viz. cutting force signal, sound energy signal, power, current, surface finish, vibration, temperature, which are influenced by cutting tool geometry and machining process conditions, are sensed through high level intelligent sensors viz. dynamometer, acoustic emission sensor, power and current sensor, surface profiler or vision based system, accelerometer, pyrometer [10] etc. in the form of signals. Acquired raw signals are filtered and analyzed for extracting relevant features which carry the machining process information. An artificial intelligence (AI) approach has been applied on the extracted signal features for decision making about the condition of cutting tool. Prediction about the future tool condition can also be possible by using AI based learning techniques. This decision of prediction is then fed to a controller for communicating with the machine tool to take proper action.

### ***13.1.1 Different Types of Cutting Tool Wear***

The product quality is mainly controlled by the process of machining. Machined surface finish, drilling and milling burrs are some key attributes to determine the product quality which are dependent mainly on the condition of cutting tool wear. Tool wear is dependent on machining conditions, machine tool condition, combination of cutting tool and work piece material, work piece geometry, tool



geometry, alignment of work piece and cutting tool, cutting chip condition etc. Different types of tool wear resulting from machining are stated below [9].

**Flank wear**—The abrasive action due to hard particle cutting results wear at the flank face of cutting tool. The profile of flank wear is shown in Fig. 13.1. Three zones are present in flank face of cutting tool insert. The worn region is divided into Zone A, Zone B and Zone C which represent notch wear zone, regular flank wear zone and wear zone near the tool nose, respectively. For an worn out tool, the maximum flank wear width ( $VB_{Bmax}$ ) is greater than 0.6 mm and the average flank wear width should be greater than 0.3 mm [11]. Poor machined surface texture is produced due to high flank wear. Consequently, the standard tool life criteria is considered from measurement of flank wear width.

**Crater wear**—Crater wear results from the abrasion and diffusion wear at the rake face due to chip-tool interaction. The strength of cutting edge reduces due to crater wear. It is mainly occurred at high temperature. Crater wear depth and area are two main attributes to the quantification of crater wear. A three dimensional view of crater wear is shown in Fig. 13.2, where the scale at right hand side is the representation of crater depth.

**Notch wear**—Notch wear is being seen at the trailing edge where cutting edge and work piece material are separated. Notch wear results from more abrading action due to work hardening. Notch wear is shown in Fig. 13.1.

**Plastic deformation**—At high cutting speed and feed rate, a material compression due to plastic deformation takes place on cutting edge at high temperature. As a result, there is a bulging of edge at tool nose region.

**Built-up edge (BUE)**—Welding of chip with principal cutting edge happens at high temperature and pressure. This welded material at the cutting edge is named as built-up edge. Chatter may result due to formation of BUE. Also chipping of BUE changes the tool geometry severely.

**Chipping and fracture**—Mechanical fatigue shock in interrupted machining and intermittent cutting is the cause of tool chipping and tool fracture. This fracture is not only damage the cutting tool, but also affects the machine tool catastrophically.

Thus, it is required to monitor the condition of cutting tool wear to achieve better performance of machining, avoiding machine tool damage and accomplishing the required product quality.

### 13.1.2 TCM Techniques

Tool condition monitoring can be achieved by using direct and indirect methods. In direct methods, different wear profiles viz. flank wear, crater wear, nose wear, chipping, breakage and fracture are measured directly using tool maker's microscope, optical microscope, radioactive sensors, laser based system and machine vision etc. Though the direct monitoring system, are mainly offline, the systems are more accurate. In indirect tool condition monitoring, machining process outputs

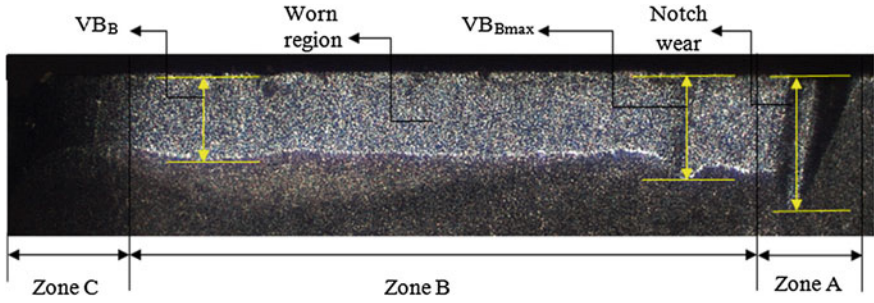


Fig. 13.1 Flank wear profile of an worn out cutting tool insert

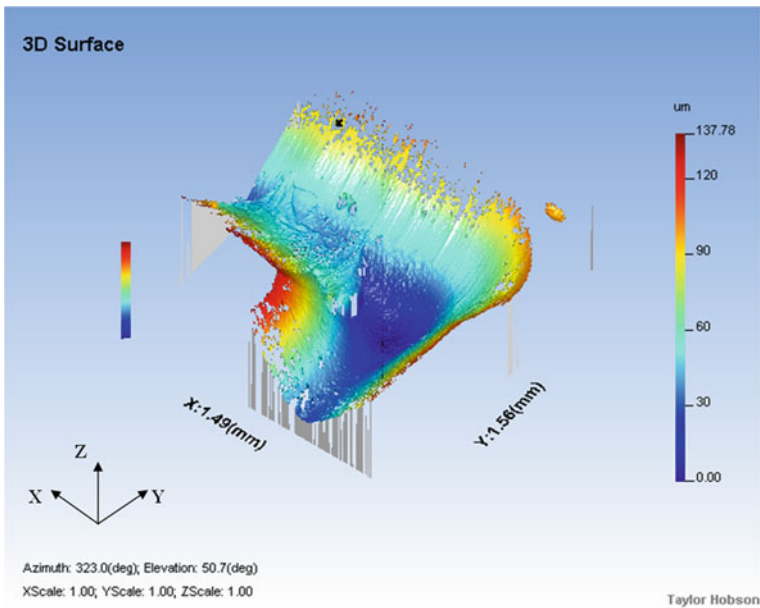


Fig. 13.2 Crater wear region of a cutting tool insert (grooved)

viz. cutting force, cutting power, sound of machining, vibrations and surface finish, which are influenced by tool condition, are monitored via different types of sensors, signal processing techniques and automatic decision making algorithms. Some commonly used TCM sensors for machining are described briefly.

**Force sensor**—The cutting force signatures for characterizing tool conditions are captured using force sensor viz. strain gauge or piezoelectric dynamometers. Tangential, axial and radial force components are resulting from cutting speed, feed force and angle between tool and workpiece, respectively. Feed force is mainly sensitive to flank wear and a substantial drop in tangential force signal can be noticed at the time of tool breakage. The elastic deformation of transducer or

sensing material due to the change of cutting forces is the principle of force sensors. However, very high frequency chatter detection is difficult for dynamometers. Also high cost and less flexibility are the two main disadvantages of force sensors [12].

**AE sensor**—A spontaneous energy release from a localized source, called acoustic emission (AE), is realized at the time of material fracture, phase transformation and deformation. This AE can be recognised by microphones, thin film piezoelectric sensors etc. [13]. AE sensors are used to delineate tool wear, tool breakage, fracture and chipping. The chattering phenomenon is well recognised by the AE sensor. The AE sensor should be placed very near to the machining site such as that can be considered as a point source. The dominance of background noise over the required signal can then be suppressed for improving signal to noise ratio. The non-homogeneity of the structures through which the AE signals can propagate is another difficulty for AE sensor [14]. However, the problem of chip and coolant fouling are the most challenging phenomena to use AE techniques in an industrial environment. This technique is mostly used for breakage and chatter detection. Generally, low frequency signals below kHz range cannot be detected by AE techniques. Thus, accurate recognition of all tool wear states are very difficult to achieve by using AE techniques.

**Vibration sensor**—Vibration due to machining operation can affect the dimensionality of the machined parts, surface finish etc. Tool wear and breakage monitoring is done by vibration signals acquired from accelerometers during machining. Accelerometer should be kept very near to the cutting site for getting proper signal. However, accelerometers are not capable to detect machining chatter as chatter is a high frequency signal.

**Current and power sensors**—Spindle motor torque is proportional to the motor armature current. This spindle torque increases with the increase in tool wear. Consequently, spindle current also increase. Thus current measurement by current sensors is an indirect way to measure cutting force. However, current sensors have a limitation to detect high frequency force components. Spindle drive power are sensed by power sensors. A rough estimation of tool condition, mainly, tool breakage is accomplished by power sensors.

**Temperature sensors**—Temperature and temperature gradient near the cutting edge are sensed by temperature sensors viz. tool work thermocouple, pyrometer, inserted thermocouple, thermal imager etc. Tool-work thermocouple only gives the average temperature information. Inserted thermocouple is able to give the temperature gradient information, but a modification of tool shape is needed for fixing inserted thermocouple. Thermal imager is capable to produce the thermal image from heat radiation of tool-workpiece interface.

**Surface profiler**—Surface profiler senses the topography of machined surface by tracing a line over the surface in contact. This method is useful for product quality inspection offline. However, the tracing may cause small scratches on the surface of soft materials.

### ***13.1.3 Importance of Machine Vision Techniques in TCM***

There are some advantages and disadvantages of machine vision or digital image processing techniques compared to the other methods applied in machining, mainly in tool condition monitoring (TCM).

#### **13.1.3.1 Advantages**

1. Machine vision systems for quality check of machined parts or for measurement of tool wear are non invasive in nature, whereas other TCM sensors viz. dynamometer, accelerometer, power sensor, current sensor, surface profiler are contact type sensors.
2. Machine vision systems are comparatively low cost systems.
3. Flexibility is another important advantage of machine vision system where the system can be moved, assembled and disassembled as per requirements.
4. Surface defects are easily detected by machine vision, apart from surface roughness profile.
5. There is a chance of damage on machined surfaces by the contact type surface profiler, whereas the surface finish measurement by machine vision does overcome this limitation.
6. Only one dimensional trace can be available by contact type surface profiler whereas a more detailed two dimensional quality check is possible via area scan camera.
7. Machine vision can be applied from a remote location, which is applicable for unmanned machining.
8. Acoustic emission (AE) techniques, used in machining, are dependent on frequency of chatter [15–17]. Also these sensors should be placed very near to the site of machining for getting proper attenuation of AE signals which may affect the condition of these sensors [18]. These limitations can be overcome by using machine vision system.
9. Tool breakage, out of tolerance of the job can only be monitored by using accelerometer. However, machine vision can monitor the progressive tool wear, tool breakage, out of tolerance, chatter and vibrations.
10. Dynamometers, in general, cannot recognize high frequency forces, also machine tool vibration can affect dynamometers [19]. Machine vision system are able to overcome these limitations by evaluating the machined surface image or tool wear image.
11. Machined surfaces are affected by tool geometry [20]. Thus, the information about the condition of cutting tool can be available by analyzing the machined surface images.
12. The information about machining parameters can be obtained from the machined surface images [21].

13. Three dimensional quality check of surface finish can also be possible, accurately, by using scanning type non-contact 3D surface profiler [22–25]. However, these 3D measurements are not useful at all for in-process quality check because of its cost and less flexible set up.

### 13.1.3.2 Disadvantages

However, there are some limitations for using machine vision system in TCM techniques.

1. An appropriate illumination system, robust image processing algorithm, protection from machining noises (chips, dirt etc.) are very much essential for the successful implementation of machine vision technique in industrial environment [20].
2. Monitoring of drill parts using digital image processing are very difficult due to its inaccessibility [26]. However, a method to monitor deep hole parts has been developed, recently [27].

## 13.2 Image Acquisition System for Machining

Camera and illumination system are two key elements for capturing an image. A brief introduction about these systems are given as follows.

### 13.2.1 Camera

Image acquisition should be done carefully for getting an informative digital image by using a digital camera and an illumination system. In case of tool condition monitoring in machining, images of cutting tool (rake face or flank surface) or work piece surface are captured with a Charged Coupled Device (CCD) camera or Complementary Metal-Oxide Semiconductor (CMOS) digital camera. For obtaining an highly magnified image of machined surface or tool wear, optical microscope is also used. However, capturing of images using optical microscope is completely an offline technique which is not suitable for capturing the machined surfaces of a large workpiece. CCD camera is comprised of CCD sensor which is an array of photosensitive elements to collect electrical charges generated by absorbed photons. Those electrical charges are then converted to an electrical signal which is converted to a digital image via frame grabber. Finally, the image is transferred to a computer for processing purpose [28]. CMOS is different from CCD sensor by its faster capturing rate. But the sensitivity of CMOS sensor is

much less than that of the CCD sensor. To create a digital image, a conversion is needed from the continuous sensed data into digital form. Also to get a magnified image, magnification lens (preferably long working distance lens) with focus and aperture adjustment capability is needed.

### ***13.2.2 Illumination Systems***

Illumination or lighting system is very important aspect in machine vision system to capture an image with high dynamics (number of intensity steps) and contrast. Due to inhomogeneous illumination for improper illumination set up, many spurious points or false edges formed as wrong informations in captured images. Different types of lighting systems are explained below [29].

Fluorescent lamp is a source with large homogeneous illumination field. However, the light frequency and image readout frequency differs and it results an interference effect on image. Thus, a frequency rectifier is used to avoid this noise.

Halogen lamps are used to get constant light intensity. However it is mainly coupled with fibre optic guided light. Fibre optic guided light source are used to illuminate small objects and difficult to access areas by its highly directional illumination property. The illumination strength can be regulated in this lighting system. Any light source (halogen light, light emitting diodes or Laser) is coupled with fibre optic cable in this system.

Light emitting diodes (LED) reacts instantly over a wide range of illumination. The most advantageous lighting system in industrial image processing is the LED light due to monochromatic nature, low cost, light weight, flexible, inexpensive operation, long life and less temperature. They are used in ring light, array of lights, stroboscopic light and in fibre optic guided lighting.

Laser source is a monochromatic and coherent light which is available as small laser diode modules. Laser diode modules are capable to project points, lines, circles etc. Laser scattering is also useful to get any surface information.

Flash lights or strobe lights are used to capture the images of moving objects by flashing the light on the object for a short period. The shutter speed of the camera should be synchronized with the strobe lighting system to get an illuminated image.

Infrared band filters allow the infrared part of the light only for getting a more clear picture of tool wear profile having very less spurious effect.

Also illumination systems required are different for the requirement of applications viz. for capturing tool wear image and machined surface image.

#### **13.2.2.1 Illumination Systems for Tool Wear Evaluation**

Highly illuminated and directional lighting is required to capture the tool wear region as to get a very accurately illuminated image.

Weis [30] tried to capture the tool wear image using a diode flash light incorporated with an infrared band filter, which helped to enhance the tool wear region with respect to the background. Also the flash lights or strobe lights have a capability to capture the images of a moving object. However, the service life of strobe light is very less and it is sensitive to vibrations [29]. Kurada and Bradley [31] used two fibre-optic guided lights to capture the tool wear regions. They used it to obtain adequate contrast between the worn and unworn tool regions due to the specular effect of this lighting. Kim et al. [32] used a fibre optic guided light surrounding the lens to illuminate the flank face portion of a 4-fluted end mill. They also examined that the best measurement of flank wear can be possible with a high power lighting (60 W). Wang et al. [33, 34] used a fibre optic guided light to illuminate the flank portion of each insert attached to a 4-fluted milling tool holder and capture the successive images in a slow rotating condition by using a laser trigger with very less blurring. The laser triggering is very useful for capturing less blurred images of moving object. Pfeifer and Weigers [35] used ring of light emitting diodes (LED) attached with camera to capture the proper illuminated images of tool inserts. The viewing angle has been varied by them for capturing same portion of the insert. Jurkovic et al. [36] utilized a halogen light to illuminate the rake and flank face of the cutting tool and a laser diode and accessories to obtain a structured light pattern on the face of the tool to detect the tool wear by the deformation of structured light on the rake face. This set up helps to obtain 3D information of tool wear. However, this lighting system is quite complex and costly. A white light from a fluorescent ring as well as light from a fibre bundle was used to minimize specular reflections to capture the tool images by Kerr et al. [37]. The fluorescent ring light is used in lowering the effect of highly specular reflections of fibre optic guided light. Backlighting are also used for capturing the drill bit to get a silhouette image of drill bit for determination of tool run out and profile deviation [38]. However, backlighting only help to capture the overall outer profile of any object.

### 13.2.2.2 Illumination Systems for Surface Texture Evaluation

Diffused frontlighting should be used to get a shadowless image of machined surfaces as most of the machined parts are specular surface having glinting highlights. A DC regulated lighting system with infrared interference filter can be used for getting the effect of diffused illumination [39]. Dome light with light shields can also be used for this purpose. Tsai et al. [40] tried to obtain a homogeneously illuminated machined surface image by a regular fluorescent light source which was situated at an angle of approximately  $10^\circ$  incidence with respect to the normal of the specimen surface. The camera was also set up at an angle of approximately  $10^\circ$  with respect to the normal of the specimen surface to obtain image at the direction of light. But this set up may only be useful for flat specimens, not for curved surfaces. Bradley and Wong [41] used a fibre optic guided illumination source and a lighting fixture. A uniform illumination of the machined

surface was achieved by changing the position of lighting fixture. During surface assessment, the specimen was positioned on the platform so that the lay marks were perpendicular to the longer dimension of the CCD sensor. In this technique, the images of flat specimens (end milled) were captured but the images of turned surface (i.e. curved surfaces) were not obtained.

### 13.3 Image Processing Methods Used in Machining

The digital images of tool wear regions, machined surface or machined parts are captured by CCD camera or CMOS camera or optical microscope. The conversion of the optical and electrical signals into digital images are achieved by sampling and quantization techniques. In sampling technique, the image frame is divided into a number of elements called pixels. The intensity values of an image are assigned to each pixel by quantization technique. In gray level images,  $2^8$  or 256 numbers of gray levels are assigned. The techniques of image processing used in machining are explained below.

**1. Image Interpolation**—Interpolation is a reconstruction technique to find the in-between values of discrete samples. Image interpolation is a method where some new pixels are re-sampled systematically and the intensity values are assigned to those new pixels on the basis of interpolation techniques viz. neighbourhood, bilinear, bicubic, B-spline etc. applied on the corresponding pixel values of the original image. Image interpolation techniques are mainly used to magnify an image. Kumar et al. [42] applied cubic interpolation technique for magnifying the machined surface images. Image registration can also be used to get a high resolution image of a full object from the low resolution part images of that same object. This technique is mainly useful to get a high resolution image from a low resolution camera [43].

**2. Image Smoothing**—Image smoothing operation is performed on the images to remove the image noise by blocking the high frequency components and passing the low frequency parts (low pass filtering). It has been accomplished by convolving a filter mask with the original image. In the convolution process, the filter mask is rotated  $180^\circ$  and then passes through each pixel of the original image [44]. The sum of product depicted in Eq. (13.1) is then calculated and replaced at the target pixel.

$$g(x, y) = \sum_{p=-a}^a \sum_{q=-b}^b h(p, q) f(x + p, y + q) \quad (13.1)$$

where,  $f(x, y)$ ,  $h(p, q)$  and  $g(x, y)$  are original image, filter mask and filtered image, respectively.

Averaging, weighted averaging are some linear filters. However, many information along with noises may be lost by applying these linear filtering techniques. So, to overcome these situations, median filtering technique has been used by



several researchers to remove the high frequency impulse noises present in the tool wear or machined surface images [31, 33, 34, 45–51]. Median filter is a non-linear filter. In this method, pixels encompassing the filter region are ranked and the median of those values is replaced at the target pixel. Less blurring is achieved using this method.

**3. Image Sharpening**—After removal of noises, the image becomes quite blurred. Thus, it is required to sharpen the fine details of an image by image sharpening operation. Image sharpening is an image enhancement technique where a differentiation operation is performed on original image for enhancing the intensity discontinuity. First order intensity differentiation or gradient detection has been accomplished by Eqs. (13.2) and (13.3) for  $x$  and  $y$  directions, respectively [44].

$$g_x = \frac{\partial f}{\partial x} = f(x + 1, y) - f(x, y) \quad (13.2)$$

and

$$g_y = \frac{\partial f}{\partial y} = f(x, y + 1) - f(x, y) \quad (13.3)$$

where,  $f$  is an image.

A gradient operator is used to enhance the prominent edges along proper edge direction as per Eqs. (13.4) and (13.5), respectively.

Gradient magnitude,

$$M(x, y) = \sqrt{(g_x^2 + g_y^2)} \quad (13.4)$$

and, Gradient direction,

$$\alpha(x, y) = \frac{g_y}{g_x} \quad (13.5)$$

Gradient operator,

$$\nabla f = [g_x \quad g_y]^T \quad (13.6)$$

The sum of all the coefficients of filter mask for gradient operator should be zero. Roberts operator, Sobel operator and Frei-Chen operator are used for this purpose [31]. However, the first order differential operators are not been able to enhance the fine details, which can be enhanced by second order differential operator or Laplacian operator due its zero-crossing capability [44]. The Laplacian filters are rotation invariant also. Laplacian of an image,  $f$  is given in Eq. (13.7) and the enhanced image,  $r$  is given in Eq. (13.8).

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = f(x + 1, y) + f(x - 1, y) + f(x, y + 1) + f(x, y - 1) - 4f(x, y) \quad (13.7)$$

and,

$$r = f + c[\nabla^2 f] \quad (13.8)$$

where,  $c$  is a scaling factor.

However, gradient operator is better than Laplacian operator for noise removal purpose. For this reason, the image can be smoothed by gradient and then the resulting image can be multiplied by the Laplacian image for getting the effect of both of the methods [44]. Unsharp masking is also another image sharpening technique which is used for processing tool wear and machined surface images [37, 42]. The histogram equalization and histogram stretching techniques are also widely used for image enhancement. In these techniques, the frequency distribution of intensity levels are equalized (contrast enhancement) over an image. The darker and brighter areas can be well visualized after performing this operation [33, 46, 51, 52].

**4. Frequency-domain Operations**—Discrete Fourier transform (DFT) is applied to convert an image into its frequency domain as per Eq. (13.9) [44].

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)} \quad (13.9)$$

where,  $F(u, v)$  is the Fourier transform of image,  $f(x, y)$  of size  $M \times N$  pixel and  $u = 0, 1, 2, \dots, M - 1$  and  $v = 0, 1, 2, \dots, N - 1$

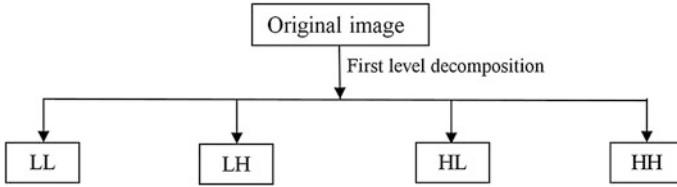
Each element of the transformed image can represent all modified values of original image. Low frequency and high frequency components of the Fourier transform corresponds to slowly varying and faster varying intensities of original image, respectively. Also, translation of original image has no effect on Fourier transform. Some features, namely, frequency spectrum, power spectrum and phase angle of a Fourier transform carry the information of intensity variation and object discrimination.

Butterworth, Gaussian and Weiner filters are widely used to remove noises from tool wear and machined surface images [41, 42, 53–55]. These filters are applied on the Fourier transformed image and then the inverse Fourier transform has been performed to get reconstructed image. The formula of inverse DFT is given in Eq. (13.10) [44].

$$f(x, y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) e^{j2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)} \quad (13.10)$$

For image sharpening, the high pass filter in frequency domain is obtained by subtracting the low pass filter from one.

Wavelet transform is an improved tool of frequency transform for multi-resolution processing. In wavelet transform, a bank of filters are applied on images to get the image information at different resolution levels. Wavelet transform can decompose an image into four lower resolution images containing low frequency



**Fig. 13.3** Wavelet decomposition

components in both the directions (LL), high frequency component in  $x$ -direction and low frequency component in  $y$ -direction (HL), low frequency component in  $x$ -direction and high frequency component in  $y$ -direction (LH) and high frequency components in both directions (HH) as shown in Fig. 13.3. A mother wavelet viz. Morlet, Coiflet, Daubechies etc. are selected for image analyses. Mainly, machined surface texture analyses are performed using wavelet transform [56–60].

**5. Image Deblurring**—Images can be blurred for motion or defocusing at the time of image acquisition or for image smoothening. Thus deblurring of image is required to enhance the image. The blurred images can be deconvolved or deblurred using a point spread function. In this regard, the point spread function (PSF), which is an impulse response of a focused image, should be known. This deblurring technique has been used to deblur the machined surface images captured at moving condition by using Lucy-Richardson algorithm [61]. Thus it may be very useful technique to preprocess the machined surface images captured at the time of machining. The defocused images of drilling burr have also been deblurred using a Gaussian PSF [62].

**6. Morphological Processing**—In morphological processing, the tool wear profile is processed for closing at the break points. In this processing, mainly the mathematical set theory is used. Dilation, erosion, closing, opening and thinning are mainly used for joining the break points situated at the profile. A structural element,  $S$ , with a defined shape is slid over a noise-free image for proper bridging of the unwanted breaks. The dilation operation is stated in Eq. (13.11) [44].

$$f \oplus S = \{z | [(S)_z \cap f] \subseteq f\} \quad (13.11)$$

The structuring element,  $S$  is flipped and then slid over an image,  $f$ , in dilation to bridge the gap.

Erosion is just opposite to dilation. The unwanted points are eroded by this operation. In opening operations, the unwanted narrow or thin points are removed from the profile. Closing operation is used to eliminate small holes and then to fuse narrow breaks. Thinning operation is used to eliminate the multiple unwanted lines detected on a profile. Morphological processing are mainly used to completing the tool wear profile for accurate wear evaluation [30, 31, 33, 35, 46–51, 63].

**7. Image Segmentation**—Image segmentation is mainly used for detecting proper edges and for delineating a profile from its background. This technique is mainly useful to detect or delineate tool wear contour from its background. Edge based

segmentation, region based segmentation, contour detection and thresholding are used in this regard. In edge based segmentation, edge detection is mainly performed on the basis of local variations of intensity values. Edge detection is done to detect the edges of tool wear profile as well as to detect the feed marks created on machined surfaces [38, 39, 45, 52, 55, 64–70]. Edges are basically a localized sets of connected pixels which have intensity variation with respect to the background. Image smoothening for noise elimination, detection of edge points and the connected component analysis for edge localizations are three major steps for any edge detection. First order differentiation of an image is helpful to detect the presence of an edge and second order differentiation of an image is helpful to delineate the edge from its background. Magnitude and direction of an edge can be detected by using Eqs. (13.4) and (13.5), respectively, using Roberts, Sobel and Prewitt operators. However, these operators are limited to detect the horizontal, vertical and diagonal edges, which are not feasible for detection of tool wear profile. To overcome this situation, Canny edge detection technique is a suitable option [38, 52, 55, 65]. This technique is required to detect the edges of machined surface as an image pre-processing technique [68, 69]. In this method, firstly, the image is smoothened by using 2-D Gaussian function. Secondly, the presence of edges is enhanced by taking the gradient of the image in X and Y-directions. Thirdly, the points are suppressed where the gradient value is not maximum (non-maximal suppression); and finally, a two-level thresholding and edge thinning have been done to extract the optimal edge image. The error of Canny edge detection method is very low and the detection of localized edge points are two key advantages of this technique.

Edge linking is a technique for betterment of profiles by linking the breaks resulted due to inhomogeneous illumination problem to detect flank wear profile [49, 50].

Region based segmentations are used to delineate the worn profile of tool wear from its unworn background and for region growing technique to improve the accurate profile detection [31, 71, 72]. Region based segmentation is a similarity based approach where similar pixels are clustered in sub-regions [41]. In region growing technique, the sub-regions are grouped based on pre-defined criteria for growth. It starts from a set of “seed” points and the predicate of connectivity and similarity is used for this purpose. Texture based segmentation is also used for clustering of different sub-regions for tool wear and machined surface images [46, 73].

Contour detection algorithms are applied after region based segmentation to detect the tool wear contour. Snake algorithm is used for active contour detection where contour points are attracted by the edges [52]. SUSAN algorithm is also used in this regard for corner detection. Freeman chain code is also an useful algorithm for contour detection [74, 75]. There is another algorithm called chain coding to detect the contour [35, 49, 50]. However, chain coding cannot be used for accurate application.

Image thresholding is mainly used to differentiate the worn region of tool wear profile from the unworn part [33, 46–50, 71, 73, 76, 77]. However, there are some applications of thresholding for drilling and milling burr profile detection [54, 78]. Global thresholding is used by determining a threshold ( $T$ ) from image histogram. Then, the pixels having intensity values less than  $T$  are assigned as 0 and all other pixels are assigned as 1, in this technique. The determination of optimal threshold value is very important for improving the accuracy of this technique. Otsu's method can determine the optimal threshold value by maximizing the between-class variance [33, 46]. In many cases, variable thresholding are also utilized by using multiple levels of threshold [54, 71].

**8. Image Representation and Description**—Boundary of a region in an image is represented by image representation method. Shape property of the region is represented by its length, orientation, perimeter etc. using image description techniques. The tool wear profile is accurately detected after all the pre-processing of tool wear profile. Then the shape operators, namely, area, perimeter, distance, compactness (perimeter divided by area) are applied on tool wear profile for determination of wear area, wear perimeter, wear land width and wear compactness, respectively [31, 33, 49, 50, 71]. Statistical moments are also a good descriptors to describe tool wear for their scale and rotation invariant properties [79–81].

**9. Texture Analysis**—Texture is a slowly varying set of local statistics or attributes of a repeated pattern. Texture primitive is a connected set of pixels, characterized by a set of attributes (coarseness and directionality). For example, in case of turned surface, a repetitive feed marks can be obtained as texture primitives. Texture analysis are applied mostly on pre-processed images of machined surface to extract some features of surface finish. However, some researchers analysed the texture of tool wear regions to extract wear features [37].

- Texture analysis can be done using statistical, geometrical, model-based and frequency-domain based methods [82]. In statistical method a texture is modelled as a random field and a statistical probability density function model is fitted to the spatial distribution of intensities in the texture. As the tool gets worn out, the resulting machined surface finish gets coarser which in turn results a more diffused reflection and a brighter machined surface image [83]. Thus, the histogram of the machined surface is changing with the change of surface finish. This property has been utilized to get some features of surface finish, namely, gray level average ( $G_a$ ), standard deviation ( $SD$ ), root mean square value ( $RMS$ ), skewness, kurtosis, maximum value of gray level distribution, maximum height of gray level histogram, width of the histogram base, two optical roughness parameters ( $R1$ ,  $R2$ ) etc. in first order statistical texture analysis [20, 37, 39, 41–43, 84–100]. However, the first order statistical texture analysis is very sensitive to lighting conditions.  $R1$  and  $R2$  are two optical parameters, which are less sensitive to light variations, given in Eqs. (13.12) and (13.13), respectively [20].

$$R1 = \frac{SD}{RMS} \quad (13.12)$$

and

$$R2 = \frac{SD}{G_a} \quad (13.13)$$

However, co-occurrence or relations between pixels cannot be obtained through first order statistical texture analysis technique.

- Gray level co-occurrence technique is a second order texture analysis technique proposed by Haralick et al. [101]. A gray level co-occurrence matrix (GLCM), whose elements are the co-occurrence values of a particular pixel pair occurring with a particular pixel pair spacing (PPS) in a defined direction (horizontal, vertical or diagonal), is constituted in this technique. The construction process of GLCM from an original image fragment (Fig. 13.4a) is shown in Fig. 13.4b. The co-occurrence of gray level intensity value 50 with PPS one in horizontal direction is four, marked within a circle, shown in Fig. 13.4b. This way the co-occurrence of other pixel values are obtained and the GLCM of Fig. 13.4a is formed in Fig. 13.4b. This technique is particularly suitable to classify micro-textures [102, 103]. Thus the small variations of machined surface texture with progressive tool wear can be depicted by texture features extracted from GLCM of machined surface images. The optimization of pixel pair spacing and direction is most important to get accurate result [104]. Contrast, correlation, homogeneity, energy, entropy, diagonal moment, secondary angular moment, dissimilarity, similarity, mean, variance, maximum probability, cluster shade, cluster prominence etc. are some extracted features from GLCM of machined surfaces to define surface finish [105–107].
- Run-length statistical (RLS) texture analysis technique is a higher order statistical technique to describe surface texture images proposed by Galloway [108]. Consecutive pixels of same gray level intensity value in a particular direction (horizontal, vertical or diagonal) construct a run. This method is mainly useful for textural classification. A systematic arrangement of number of runs for different gray value constitute the run length matrix. An example of the construction of run length matrix is shown in Fig. 13.4c. A gray level intensity value of 50 occurs four times consecutively in vertical direction shown by a rectangular box in Fig. 13.4a (original image fragment). Thus the run of 50 at that position is four and number of runs with run length 4 of gray level value 50 is one which is marked with a rectangle in Fig. 13.4c. This way, the run length matrix (Fig. 13.4c) is constructed from original image fragment (Fig. 13.4a) where  $i$  denotes the gray level intensity value and  $k$  denotes length of runs. Some textural features are extracted from run length matrix for texture classification. Short run emphasis, long run emphasis, run length non uniformity, run percentage, gray level non uniformity etc. are some features to describe the machined surface texture [66, 67, 109, 110]. However, the detailed analysis of

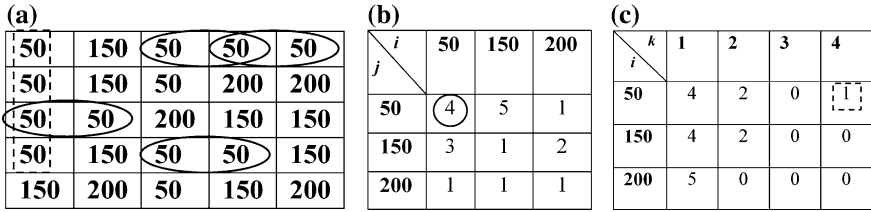
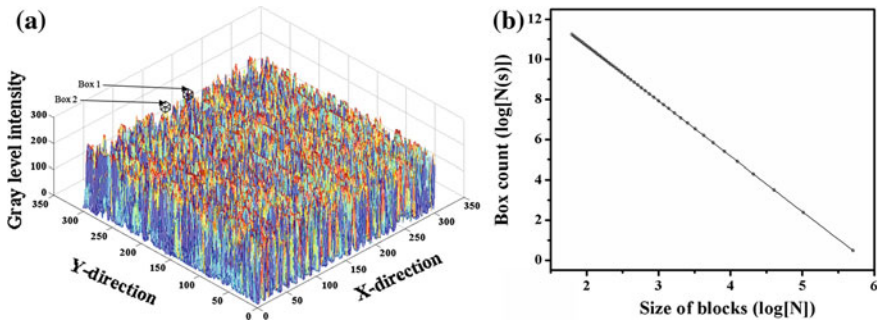


Fig. 13.4 a Original image fragment and corresponding, b GLCM and c run length matrix

micro-texture cannot be possible through this RLS method. It is mainly suitable for macro-texture classification.

- Hough transform is a structural or geometrical texture analysis method used to define surface finish. In this technique, the orientation of different lines are tabularized in an accumulator space and then line length, angle metrics, fitting error etc. are extracted as features to define the surface finish [68, 69]. It is mainly used on edge detected binary images.
- Fractal analysis technique is a model based texture analysis technique, which is used for extracting fractal dimension to define surface finish [111, 112]. A shape, which has same statistical properties at all scale, is called fractal. Thus, all fractals have a self-affine nature which is quantified by fractal dimension (FD). Fractal analysis is a model-based texture analysis technique which can define complex and intricate shapes. In this chapter, box-counting technique is described to compute FD. A three dimensional mesh plot of a machined surface image is shown in Fig. 13.5a. The total three dimensional region is covered by boxes with side lengths of  $s, a.s$  and  $b.s$  where  $s$  is a variable within a range. The boxes, through which any portion of the surface passes, are taken into count as number of boxes,  $N(s)$ . In this example, Box 1 is counted where as Box 2 cannot be counted as no portion of the surface passes through Box 2. As per fractal theory,  $N(s)$  is proportional to  $s^{-FD}$ . Thus,  $\log[N(s)]$  is plotted and fitted linearly against  $\log[s]$  as shown in Fig. 13.5b, where slope of that plot is the value of FD. This is a scale invariant technique for its property of self-affinity. Shape property of an image is mainly described by fractal dimension. This analysis has also been used for extracting the tool wear metric from tool wear images [37]. However, the roughness of a surface cannot always be described by fractal dimension [101].

Machined surface images and tool wear images are also analysed by using frequency domain based texture analyses [37, 40, 41, 43, 61, 93]. In this technique, surface finish descriptors, namely, major peak frequency, principal component magnitude squared, average power spectrum, central power spectrum percentage etc. are extracted from the frequency spectrum of an image. Autocorrelation of speckle patterns of machined surfaces is also another good option to define surface finish [53].



**Fig. 13.5** a Box-counting method, b plot to calculate fractal dimension

Also a texture analysis based on Law's descriptor is used to enhance the micro-texture features by Law's filter bank [39].

GLCM technique can be used for more accurate analysis of machined surfaces as it can find the pixel to pixel correlation. However, GLCM is not a brightness invariant technique. Frequency domain based texture analyses are scale and rotation invariant techniques. Also these techniques are invariant from inhomogeneous illumination of the image.

**10. Stereoscopic Imaging**—Two dimensional images are projection of three dimensional real world scenes. Thus, depth information cannot be obtained in two dimensional images. In stereoscopic vision, two images captured by two cameras simultaneously, are combined and mapped into an image to get depth information where the relative position of both the cameras should be known. The stereo imaging techniques are useful for getting the depth information of crater wear [113, 114]. In this technique, the images of crater area or drilling burr are captured with various object to camera distances and then all the images are stitched or mapped to get a full three dimensional representation.

## 13.4 Literature Review

Direct measurement of tool wears (flank wear, crater wear and nose wear) have been accomplished by processing the tool images in direct monitoring. On the other hand, surface finish features are extracted from the machined surface images in indirect tool condition monitoring. A good comparison of different image processing techniques used for tool condition monitoring has been reviewed in [8]. Different measurement techniques of cutting tool wear and surface finish using image processing are described in the following sections.



### **13.4.1 Tool Wear Evaluation**

Evaluation of different wear profile images of cutting tool used in turning, milling, drilling and grinding are explained in this section.

#### **13.4.1.1 Turning**

Single point cutting tool is used in turning operation. Flank wear and crater wear are two main types of wear produced due to turning. Processing two dimensional images of flank face and rake face of cutting tools is performed to obtain the tool wear measurement. Pre-processing is the main step for noise reduction to enhance tool wear image. Accurate image segmentation is then applied to extract the wear profile from the unworn background. Then an automatic determination of wear width, area and perimeter is performed to get the tool wear attributes. In flank wear measurement of cutting tool, the flank face image is captured where the worn region should be much brighter than the background. Two-dimensional image processing technique is sufficient to determine the flank wear attributes. Flank face image of tool insert is captured to evaluate flank wear and notch wear. Image segmentation techniques are used to segment the tool flank wear profile from the unworn region. Then thresholding and morphological techniques are used to obtain a boundary of wear profile. Flank wear width, wear area, perimeter and compactness are then determined automatically in micron scale using image calibration [33, 45, 55, 72, 115]. Schmitt et al. [116] has used a dynamic contour detection algorithm to get more accurate detection of flank wear profile. Texture analysis methods viz. histogram based processing, GLCM technique, frequency domain based approach and fractal based box counting techniques can also be applied to get tool flank wear metrics [37]. However, average flank wear or maximum flank wear measurement cannot be possible from these texture analysis techniques. Five statistical moment invariant techniques, viz. Zernike, Legendre, Hu, Taubin and Flusser on flank wear images for extracting moments as flank wear descriptors can also be used [79–81]. Pattern classification technique to classify sharp, semi-dull, dull and broken inserts are also used [52, 79–81].

Three dimensional information is required to determine the crater depth of crater wear, which cannot be possible to obtain from two dimensional images. Thus, an autofocusing technique using a microscope to get an image series of crater zone captured at different depths is required to get three dimensional information of rake face of a tool insert. Median filtering, image averaging, deblurring techniques, optimal thresholding, Laplacian contour detection method, edge linking and chain coding techniques can subsequently be applied to get a noise free and segmented worn region. An infinite impulse response (IIR) filter with Laplacian criterion function can be used for automatic focusing [49, 50]. Ramamoorthy and co-workers [113, 114] captured images of rake face at different known positions by shifting the position of the tool insert, systematically. Also

they have performed the depth calibration. Then a stereoscopic imaging technique has been performed to model the three dimensional image of crater zone and for getting a crater depth. Trends of tool wear pattern were then analyzed with a multi-layer perceptron neural network (MLPNN) algorithm, where inputs were speed, feed, depth of cut and cutting time and output parameters were flank wear width and crater wear depth. However, the small crater depth estimation could not be obtained, accurately, by their technique. Another technique, utilizing the principle of fringe pattern deformation due to crater wear can be utilized for obtaining the three dimensional shape of crater region. Four fringe patterns are generated at four phase shifting angles to get a 3D profile of crater wear for measuring crater depth, crater width, crater centre and crater front distance [117]. White light interferometry technique are also used to measure crater depth [118, 119]. In white light interferometry technique, a vertical scanning has been performed to get the best focus positions for each and every point presented in the object to be measured. White light is used to get the high resolution (sub-nanometer) and high precision measurements over an wider area.

Sahabi and Ratnam [47] measured nose radius of the turning insert online from the silhouette image of the cutting tool tip. Median filtering, Wiener filtering, morphological operations, image transformation by rotation and scaling, thresholding and image subtraction methods were subsequently applied by them for noise reduction, image reconstruction, boundary completion, image alignment, worn area segmentation, respectively. The information obtained about the tool wear profile has been utilized to get the surface finish profile, according to their method. However, these methods were utilized only the images of the vicinity of tool nose portion.

### 13.4.1.2 Milling

Milling is an machining operation by multi-point cutting tool. Intermittent cutting is performed by milling. Shape of the milling tool is more complex than the single point cutting tool used for turning. Thus, getting an accurate flank wear profile is a challenging task in milling. Online capturing of flank wear images of a four-fluted milling tool in very slow moving condition by using laser trigger, flash light and CCD camera are accomplished by Wang et al. [33, 34]. Indexing of the tool inserts to get proper numbering of flute has been performed by cross-correlation technique. Median filtering, histogram stretching, edge enhancement, edge detection, Otsu's optimal thresholding, morphological technique and parallel scanning methods can be applied subsequently on flank face images before automated wear measurement. A parallel scanning technique can be applied to omit false edges due to unwanted reflections. Hough transform can be applied to get a reference line for automated measurement. Then a statistical moment based method, which is invariant to rotation, can be applied to extract wear metrics. Inhomogeneous illumination on milling tool surface is a major problem for accurate wear measurement. To overcome this situation, Pfeifer and Wieggers [35] captured images of

milling tool inserts with a ring light in different illumination angles. High pass filtering, edge thinning and contouring techniques by chain coding are applied subsequently on the captured images. Then a pair-wise comparison of a series of images captured at the same location of tool insert by varying illumination condition. Translation, rotation and scaling techniques are also applied to set the same position of all the images, which is very much useful for online capturing of milling cutter [63]. A three dimensional shape of the wear zone of a micro milling cutter has been obtained by Ng and Moon [120], by varying the tool and camera plane distance with high resolution.

### 13.4.1.3 Drilling

Drilling operation is performed for hole making with a complex shaped twist drill. Capturing the wear profile of twist drill is much more complex due to its complex geometry. Online capturing of flank wear profile is very much difficult task. Different image processing techniques are used to obtain wear information of twist drill. Linearity of lip portion of drill bit is deviated due to wear. This can be measured from the edge detected image of drill bit [65]. The tool run-out effect of twist drill can be evaluated by capturing drill bit image at rotating condition using backlighting [38]. Spatial moment based edge detection technique with sub-pixel accuracy, edge smoothening with B-spline method, vertex detection technique by Gaussian filtering, accurate thresholding and morphological operation can subsequently be applied on flank face images of twist drill to obtain flank wear profile [74, 121, 122]. A level set based segmentation method is applied for accurate segmentation of wear profile in presence of smearing [74, 75].

### 13.4.1.4 Grinding

Grinding is a surface finish operation where a rotating abrasive wheel (grinding wheel) is rubbed on the semi-finished flat surface. Grain fracture, bond fracture and attrition wear are three types of pre-dominant wear in grinding wheel. Wear flats are developed on the grinding wheel surface due to attrition wear. Consequently, the increasing rate of wear flats area develops heat and burn the work-piece. But the automatic and precise segmentation of true wear flats from the wheel surface images are quite challenging task. An edge detection approach after thresholding were utilized to distinguish true wear flats from its background [123]. However, the accurate selection of intensity threshold and edge threshold was a difficult task. To overcome this problem, region growing based image segmentation can be utilized to segment true wear flats from its background [73]. Multi-scale wavelet transform analysis and classification technique can be applied for distinguishing the grains and cavities on the wheel surface [124]. Histogram based, GLCM based and fractal based texture analysis method can also be applied on the wheel surface image for describing the condition of grinding wheel surface [95].

**Table 13.1** Tool wear evaluation by image processing

Type of cutting tool	Type of tool wear	Application area	References
Cutting tool inserts (turning, milling)	Flank wear	Off machine tool inspection	[31, 37, 45, 52, 55, 72, 75, 79–81, 115, 116]
		On machine tool inspection	[33–35]
	Nose wear	Off machine tool inspection	[63]
		On machine tool inspection	[47, 48]
	Crater wear	Off machine tool inspection	[49, 50, 113, 117–120, 125]
	Drill bit	Flank wear	Off machine tool inspection
On machine tool inspection			[121]
Tool run-out and lip deviation		On machine tool inspection	[65]
		On-line inspection	[38]
Grinding wheel	Wear flats	On machine tool inspection	[73, 95, 123, 124]

Table 13.1 states the condition of the research on tool wear evaluation using digital image processing techniques.

A flow chart for determining the tool condition for flank wear is stated in Fig. 13.6. Capturing of flank face image using a diffused front lighting (preferably ring light) with CCD camera equipped with long working distance lens can be done for avoiding unwanted variation of brightness. Then, low pass filtering (LPF) for noise reduction, image enhancement for enhancing sharpness and image segmentation for segmenting worn area from unworn region are needed, subsequently. Reference line detection for measuring flank wear width can be done by using edge detection and Hough transform on segmented image. Then, different parameters of flank wear can automatically be measured from the thresholded and morphologically operated wear profile with the help of detected reference line.

So, in direct technique, condition monitoring is done by analyzing the change in geometry in the cutting tool. Automatic measurement of different cutting tool wears are measured directly from the cutting tool image in direct method. Chatter, vibration, cutting force change etc. are not taken into account with cutting tool observation whereas surface finish can emphasize those changes as well as change in tool geometry. Thus, processing of machined surface images is become a hot topic to monitor the tool condition, now-a-days.

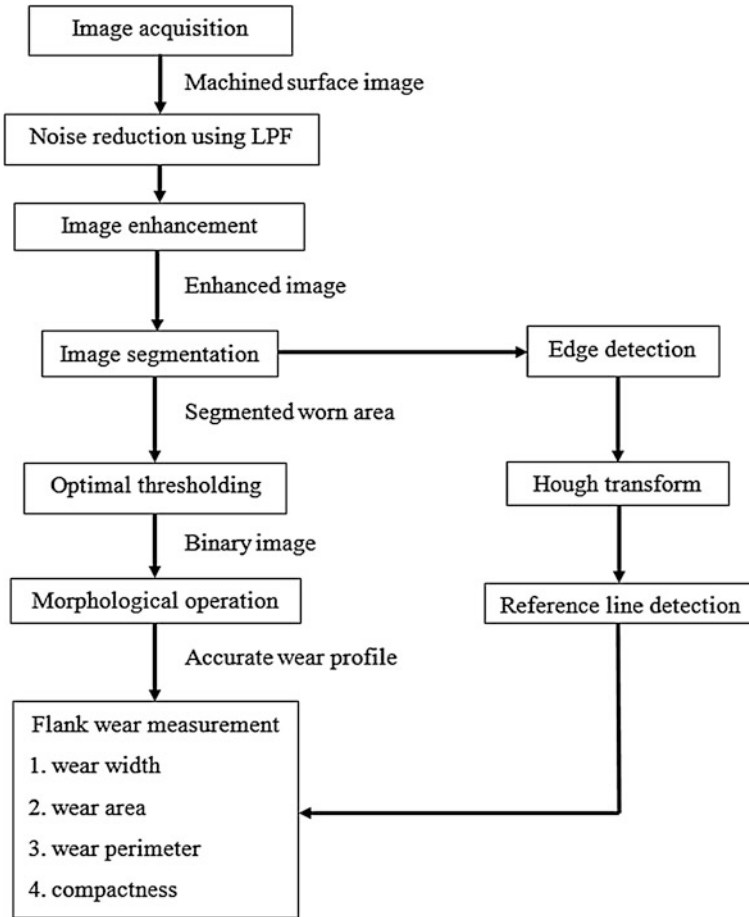


Fig. 13.6 Flow chart for tool flank wear evaluation using image processing

**13.4.2 Machined Part Evaluation**

Mechanical strength, wear resistance, bearing properties etc. are dependent on the surface finish of machined parts. The surface finish is dependent on the cutting tool geometry, machining conditions and machine tool condition. Geometrical and dimensional accuracy of machined part are also dependent on the above mentioned factors. Thus, product quality inspection involves inspection of surface finish by surface profiler or machine vision sensors. There are tactile and non-tactile techniques to assess the surface quality of the machined parts. In tactile techniques, surface roughness parameters are measured using a stylus instrument. But there is a chance of scratches on soft materials in tactile techniques due to the tracking of stylus on measurable surface. Thus, the non-tactile techniques using 3D surface

profiler or machine vision sensors play a very important role for soft materials. However, surface finish checking by 3D surface profiler is an offline technique which is accurate but time consuming. So, now-a-days, research is going on for surface finish inspection via machine vision sensor using digital image processing tool to achieve in-process surface finish inspection. While tactile techniques characterize a linear track over the surface of the part, the computer vision techniques allow to characterize whole areas of the surface of the machined part, providing more information [98, 126, 127]. Machine vision techniques also computationally very efficient day by day to apply the surface finish inspection online which can pave a way for online TCM. According to this, it is possible to apply these techniques for controlling the processes in real time on an autonomous manner. An exhaustive validity check can also be made to every single part produced for process layout. Continuous advances have been made in sensing technologies and, particularly, in the vision sensors that have been specially enhanced in capabilities with lower cost with increasing flexibility. In all, computer vision is a very useful non-invasive technique for the industrial environment. In this technique, just machined surface images are captured and processed by image texture analysis techniques to extract texture descriptors which are in turn describe the surface finish. Researches on different texture analysis methods for surface finish assessment by digital image processing are discussed in the following sections.

#### 13.4.2.1 First Order Statistical Texture Analysis Technique

Surface reflection obtained from the smoother and coarser surfaces are specular and diffused, respectively [83]. Consequently the shape of the image histogram changes. In rougher surface, more number of pixels consisting of high intensity value with a flatter peak histogram can be observed. Thus, the property of histogram is different for smoother and coarser surface image.  $G_a$ ,  $SD$ ,  $RMS$ ,  $R1$ ,  $R2$ , maximum value of gray level distribution (GLD), peak of GLD histogram, base width of histogram etc. are extracted as surface finish descriptors in first order statistical texture analysis [51, 84–89, 97–99].  $R1$  parameter is independent of light intensity variation and magnification [87]. Sarma et al. [128] achieved a linear correlation of measured  $R_a$  values and evaluated  $G_a$  values. Variance of frequency histogram of machined surface images is also used for surface roughness description [88].

Machining condition also plays an important role in average surface roughness. Thus, cutting speed, feed rate, depth of cut and vision based  $G_a$  values were used to estimate and predict average surface roughness for turning using polynomial network and adaptive neuro-fuzzy inference system (ANFIS) in [90, 91]. In the same direction, Lee et al. [94] fed  $G_a$ ,  $SD$  and spatial frequency as inputs to ANFIS for predicting average surface roughness.

Surface finish description is enhanced by magnifying and enhancing machined surface images using cubic convolution based interpolation technique and linear edge crispening method [42, 129]. A better prediction of  $R_a$  values is achieved

using calculated  $G_a$  values of magnified and enhanced images. A genetic algorithm based Evolvable Hardware (EHW) chip has also been used successfully in [130] for determining the  $G_a$  value of machined surface images.

Dhanasekar et al. [61] studied the blurring effect for capturing machined surface images of milled and ground specimens in moving condition. They deblurred those images using Richardson–Lucy restoration algorithm. Those deblurred images were pre-processed to compensate the inhomogeneous illumination. Afterwards, the spatial frequency, arithmetic mean value and standard deviations were extracted as texture features. An artificial neural network (ANN) was used with these three features as inputs to predict the vision based surface roughness.

In a very recent work, Al-Kindi and Zughaer [20] captured face-milled surface images by attaching two high resolution cameras with CNC machine spindle. The effect of tool imprint on the machined surface was emphasized in their study. They tried to find out vision based micro roughness by scanning a line along lay direction on the image. Also they have incorporated the roughness features into the G-code of CNC machine to control the machining condition.

However, the real-time control at the time of machining has not been encountered by them. First order statistical texture analysis technique is not capable to find the relation between the occurrence of pixels and also these techniques may suffer from variation of lighting conditions which can be overcome by using higher order statistical texture analyses.

However, the relationship among the pixel values cannot be evaluated in first order statistical texture analysis. This relationship can be obtained by using higher order statistical texture analysis technique.

#### 13.4.2.2 Higher Order Statistical Texture Analysis Technique

In higher order statistical texture analyses, the co-occurrence of pixel values or the relationship among the pixel values along a particular direction is evaluated by the help of higher order statistics. GLCM and RLS techniques are second order and higher order statistical texture analyses techniques, respectively, which are used to extract surface finish features from machined surface images.

Texture descriptors obtained from GLCM of ground, milled and shaped surface images can be utilized to classify the machining process [109]. Surface finish features can also be extracted from GLCM [21, 39, 104–107, 131]. Those features are also related with feed rate, machining speed and depth of cut for turning and milling [105–107]. However, pixel pair spacing (PPS) and angular direction of GLCM calculation have major effects on the result [21, 131]. With the variation of feed rate, the optimum PPS value is varied [21]. A periodic relation of extracted texture descriptor with PPS values can be obtained in this regard. Utilizing this phenomenon, a method reported in [104] can be adopted to obtain optimum pixel pair spacing and direction for progressive tool flank wear monitoring. However, there is a variation of GLCM features on the inclination of machined surface [132].

The classification of chatter rich and chatter free surface can also be described by GLCM technique [96].

Run length statistical analysis techniques has been applied to extract surface finish features in [66, 67, 109, 110]. Progressive tool flank wear monitoring has been done by RLS features in [58, 59, 100]. However, micro-texture analysis cannot be achieved using RLS technique.

#### 13.4.2.3 Geometrical Texture Analysis Technique

The feed marks become small and irregular as the cutting tool wear takes place. Also a more scattered angular distribution of feed marks on machined surface images has been observed with the increase of tool wear. This phenomena can be represented by Hough transform. Hough transform has been applied on machined surface texture images by Kassim et al. and Mannan et al. [68, 69]. They improved Hough transform technique for faster computation by using neighbourhood processing. A canny edge detection technique has been utilized by them before the Hough transform. Line length, fitting error and standard deviation of angle were the extracted features for surface finish description. However, in this process, the minute details of gray level variations in machined surface images are lost due to image binarization technique. But this process is a fast technique which may be applicable for online TCM purpose. A MLPNN technique has also been applied by them to predict the tool flank wear by feeding the values of those extracted texture descriptors. However, the error of prediction has not been reported by them.

Bamberger et al. [70] inspected the chatter marks produced on valve seats at the time of machining by using image processing. Hough transform based circle fitting method, shape analysis based circularity technique and GLCM based texture analysis technique to classify the accepted and rejected valve seats. However, the method of quality inspection was dependent on radial edge detection method based on the threshold value. Thus, an automatic threshold selection technique was required to minimise the operator's intervention for quality check.

#### 13.4.2.4 Model Based Texture Analysis Technique

Model based texture analysis technique can be able to quantify the essential qualities or complexities of texture. Fractal analysis technique is the most popular technique among them which can be utilized the self-similarity nature of natural surfaces [82]. Fractal analysis is also an image magnification invariant technique. Sometimes, fractal dimension can represent the roughness of a surface [111]. Self-affinity of an image is evaluated using fractal dimension. Two methods of finding fractal dimension by using fractional Brownian motion and box-counting have been used by Kassim et al. [133] for end-milled surface images. They classified four states of tool wear by utilizing the slope and intercept values (fractal index) of fractal plots along six directions for fractional Brownian motion and one slope and



intercept value for 3D box-counting technique using Hidden Markov model (HMM). However, any classification accuracy has not been reported by them. Kang et al. [112] determined the fractal dimension of end milled surface images. They determined the fractal dimension of the line scanned profile perpendicular to the lay direction and also determined the fractal dimension (FD) of gray level images using 3D box-counting. FD was in between 1–2 and 2–3 for first and later case, respectively. They found that the fractal dimension is increasing with the increase in tool flank wear and measured average surface roughness. However, any quantitative correlation study of FD with wear or surface roughness was not present in their work. Vesselenyi et al. [134] utilized 3D box-counting method for characterizing the fractal dimension. They fitted the fractal plots (log–log plot between size of box and number of box) in linear, second and third order polynomial from where nine coefficients of fitting equations have been extracted as texture features. Finally, they classified shaped, ground and polished surface images from those extracted features by employing fuzzy C-means clustering technique. However, any classification accuracy has not been reported in their work.

### 13.4.2.5 Frequency Domain Based Texture Analysis Techniques

In spatial domain, the images are dependent on the magnification and rotation. Also the periodic nature of turned and milled surfaces cannot be quantified directly in spatial domain. These drawbacks can be overcome by using frequency domain based texture analyses of machined surface images which is more suitable for online application. An image can be converted into frequency domain using Eq. (13.9). Then the squared value of  $F(u, v)$  is representing the power spectrum of an image. A normalized power spectrum,  $p(u, v)$  is then calculated as per Eq. (13.14) [40].

$$p(u, v) = \frac{P(u, v)}{\sum_{(u,v) \neq (0,0)} P(u, v)} \quad (13.14)$$

The following texture features are then extracted from normalized power spectrum.

1. Major peak frequency is the Euclidean distance of frequency co-ordinate of the maximum peak from the origin
2. Principal component magnitude squared is the maximum eigen value of covariance matrix of  $p(u, v)$ .
3. Average power spectrum is the average value of the power spectrum with respect to the image size.
4. Central power spectrum percentage is the percentage ratio of central frequency component and summation of power spectrum.
5. Ratio of major axis to minor axis is the ration of maximum and minimum eigen values of covariance matrix of  $p(u, v)$ .

Prediction of surface roughness of shaped and milled surface is done by feeding the above mentioned texture features to an artificial neural network [40] or an abductive network [93]. Major peak frequency is the best selected features among them [40]. With another approach, the images of speckle patterns created on ground and milled surfaces are pre-processed by butterworth filter and centralised FFT of those pre-processed images are determined [53]. Then the autocorrelation co-efficients in X, Y and diagonal directions are determined as surface finish measures.

A surface roughness prediction method has been proposed using major peak frequency, principal component magnitude square and  $G_a$  values of end milled surface images, cutting speed, feed rate and depth of cut as inputs to a back propagation neural network (BPNN) in [125].

Dhanasekar and Ramamoorthy [43] enhanced the resolution of the ground and milled surface image by using projection on convex sets (POCS) algorithm. They found that POCS algorithm has the largest signal to noise ratio compared to other super resolution image reconstruction algorithms for ground and milled surface images. Then they extracted major peak frequency, principal component magnitude square and standard deviation of gray level distribution from the reconstructed image. Finally, they predicted the surface roughness using group method of data handling (GMDH) and compared those predicted values with stylus based surface roughness. Highly improved correlation of vision based surface roughness value with stylus based average surface roughness has been observed in their work after applying super resolution reconstruction algorithm on both ground and milled surface images.

Liu et al. [27] critically captured the R-surface of a deep hole drilled valve using a camera equipped with long working distance lenses and a ring light illuminator. Principal component magnitude squared, average power spectrum, central power spectrum percentage and ratio of maximum and minimum Eigen values of the covariance matrix of normalized power spectrum were utilized to describe the surface finish of the R-surface images. The average surface roughness of R-surface has been predicted using a hybrid genetic algorithm and back propagation neural network. Predicted surface roughness has been compared with the average surface roughness measured using a three dimensional surface profiler.

However, Fourier transform are not able to consider the spatial information present in an image. Also multi-channel or multi-resolution representation of an image cannot be obtained using Fourier transform. Thus, Gabor filtering technique and wavelet based filtering approach come into the picture. Gabor and wavelet based methods are able to extract the information of spatial and spatial-frequency domains. They can also be able to represent an image in multi-resolution which is a mimic of human visual perception. Also, the band pass filtering can be obtained from those two approaches. Gabor filter is a sinusoidal filter using Gaussian function where frequency and orientation can be modulated. Tsai and Wu [135] automatically classified the defective and no-defective samples of milled specimens using the energy function evaluated from Gabor filtered images of those samples. Optimum parameters for Gabor filter were obtained when the energy of filtered defect free sample was zero or minimum. They distinguished the homogeneous and non-homogeneous surface texture also.

Wavelet analysis of machined surface image is also a very useful technique. Wavelet is produced from a function, called mother wavelet, by shifting and scaling. Wavelet is mainly useful due to its multi-resolution analysis capability. An image is decomposed into four images with low frequency components in both directions (LL), low frequency in X- and high frequency component in Y-directions (LH), high frequency in X- and low frequency component in Y-direction (HL) and high frequency components in both directions (HH). Thus the required spatio-frequency information can be obtained at different decomposition levels using wavelet transform. Thus, the multi-resolution information is obtained from machined surface images in spatial-frequency domain, whereas fast Fourier transform of an image cannot give any multi-resolution information. Haar, Morlet, Coiflet, Daubechies are some popular wavelet functions depending on their shape. Morlet and Daubechies wavelet are some common wavelets used to classify different types of surface images [56, 57]. Niola et al. [58] used Haar wavelet transform on milled, ground and polished surface images to reduce the inhomogeneous illumination problem. Coiflets are used successfully on one dimensional surface profile obtained from turned surface image [60]. In a defect detection and classification study on various types of surfaces, Zhang et al. [136] utilized shape features, Laws features extracted by convolving images using Laws kernels, discrete cosine transform (DCT) features extracted by convolving DCT kernels, features utilizing Gabor filter banks and statistical co-occurrence features extracted from GLCM for classification of 15 defects on engineered surfaces. Combination of Gabor filter bank features and GLCM features were the best selected feature inputs to a support vector machine algorithm for best classification.

#### 13.4.2.6 Quality Inspection Observing Burr Profiles

Plastic deformation produces unwanted burrs (a projected edge) in drilled and end milled products. Products with burrs are not acceptable in terms of quality. Thus, the automatic burr inspection is an essential part of product quality inspection using machine vision system. Nakao [54] utilized four 45° inclined mirrors surrounded the drilled hole for inspecting the side of the burrs, accurately. Noise reduction, component labelling and binarization have been applied prior to the burr thickness and burr height measurement with micron level accuracy. However, a three dimensional method is needed for determining more accurate burr height as the burr profile is not in a single plane. Yoon et al. [62] inspected the burr profile of drilled parts using shape from focus technique. They captured several images based on the best focus positions by systematically varying the object to camera distance. They used Laplacian mask for edge enhancement prior to measure burr height, burr width, hole location error and diameter deviation. Sharan and Onwubolu [78] measured burr height and burr width of end milling burr from the pre-processed image of end-milled product.

A summary of the application of digital image processing for machined product quality assessment is stated in Table 13.2. It can be depicted from Table 13.2 that

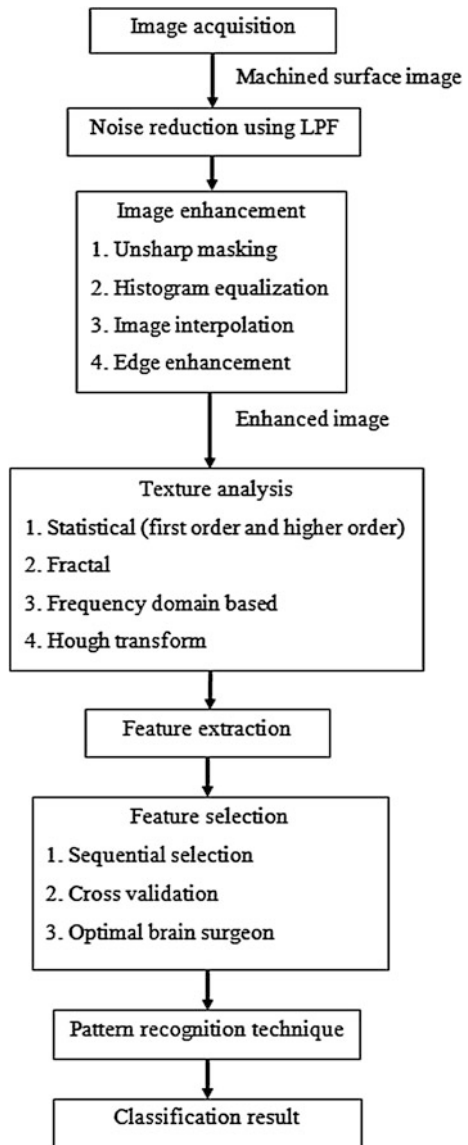
**Table 13.2** Machined product quality assessment using digital image processing

Type of machining	Quality assessment scope	Application area	References
Turning	Surface finish assessment and classification	Off-machine tool	[83–85, 88, 107, 128, 131]
		On-machine tool	[39, 60, 89, 93, 94, 98, 99]
	Progressive tool wear assessment through machined surface images	On-machine tool	[21, 66–69, 104, 110]
	Chatter identification	Off-machine tool	[70]
		On-machine tool	[96]
Milling	Surface finish assessment	Off-machine tool	[40, 42, 43, 51, 53, 56, 57, 87, 97, 105, 106, 109, 125, 129, 130, 132, 135]
		On-machine tool	[20, 62, 112]
	Progressive tool wear assessment through machined surface images	On-machine tool	[41, 68, 69, 133]
	Burr profile	Off-machine tool	[78]
Grinding	Surface finish	Off-machine tool	[42, 43, 51, 53, 86, 87, 109, 129, 134]
Shaping	Surface finish	Off-machine tool	[40, 42, 87, 109, 129, 132, 134]
Drilling	Surface finish	Off-machine tool	[27]
	Burr profile	Off-machine tool	[54, 62]

there is an absence of research on real time monitoring technique, even there are very few researches on on-machine tool monitoring of tool wear through processing of machined surface images.

A flow chart shown in Fig. 13.7 describes the method for tool condition classification by employing machine vision on machined surface images. After acquisition of machined surface images, an image enhancement method for enhancing the edge details can be applied. Then, one or more texture analyses techniques can be applied for extraction of texture features for describing tool flank wear. Images should be captured for progressive wear monitoring time to time. Feature selection or feature reduction technique is also needed for selecting well correlated features with the change of tool condition. Then a classification method employing any pattern recognition technique (SVM, MLPNN, ANFIS etc.) is needed to apply for classification of tool condition viz. tool wear states, tool breakage, tool fracture, tool chipping etc.

**Fig. 13.7** Flow chart for tool condition classification from machined surface images



### 13.5 Case Study

A case study for progressive tool wear detection using the extracted features from GLCM of machined surface images is presented in this chapter for a typical machining condition.

**Table 13.3** Experimental set up

Equipment	Specifications
CCD camera	Genie HM1024
Tele lens with polarizer	Zoom lens with macro zoom (18–108 mm)
Image acquisition software	SAPERA LT
Surface roughness tester	Mitutoyo Surftest SJ-201
Lighting system	DC regulated fibre optic guided light (Navitar)

### 13.5.1 Experimental Set Up

Turning of a C-50 steel work piece is done in a CNC lathe machine with 150 m/min cutting speed, 0.24 mm/rev feed rate and 1.5 mm depth of cut using an uncoated carbide cutting tool insert SNMG 120408-QM. Images of just machined surfaces are captured with a machine vision system mentioned in Table 13.3.

Average surface roughness ( $R_a$ ) of the corresponding surfaces are measured with the help of a stylus-based surface profiler. Simultaneously, the flank wear of the tool insert are measured with the help of Olympus BX51M microscope with 10× zoom equipped with image acquisition software, Olympus image analyzer. Experimental set up is shown in Fig. 13.8.

Each image is cropped to  $300 \times 300$  pixels before processing. All image processing are done in MATLAB<sup>®</sup> environment using standard toolboxes of MATLAB<sup>®</sup>.

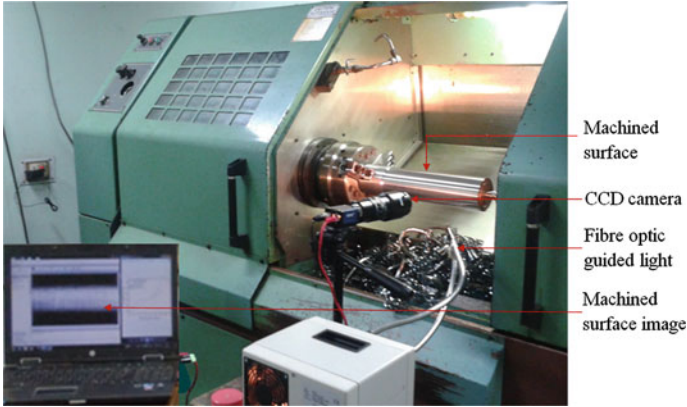
### 13.5.2 Image Processing

The image processing techniques applied on captured and cropped machined surface images are described stepwise.

Step 1: Image interpolation technique by using bicubic interpolation is applied on machined surface images to magnify them. Images of  $300 \times 300$  pixels size are magnified to the size of  $600 \times 600$  pixels. In bicubic interpolation, a two-dimensional bicubic interpolation kernel, stated in Eq. (13.15) is convolved with the original image matrix.

$$h_{bic} = h_{cub}(x).h_{cub}(y) \quad (13.15)$$

where,  $h_{cub}(x)$  and  $h_{cub}(y)$  are one-dimensional cubic interpolation kernel stated in Eq. (13.15a) along x and y axis, respectively.



**Fig. 13.8** Experimental set up

$$\begin{aligned}
 h_{cub}(x) &= |x|^3 - 2 \cdot |x|^2 + 1 && \text{for } 0 \leq |x| < 1 \\
 &= -|x|^3 + 5 \cdot |x|^2 - 8 \cdot |x| + 4 && \text{for } 1 \leq |x| < 2 \\
 &= 0 && \text{for } |x| \geq 2
 \end{aligned} \tag{13.15a}$$

Images before and after applying the interpolation are shown in Fig. 13.9a and b.

Step 2: After magnifying the image, histogram equalization technique is applied on magnified images to achieve a homogeneous distribution of illumination. Figure 13.10a is the histogram equalized image of Fig. 13.9b.

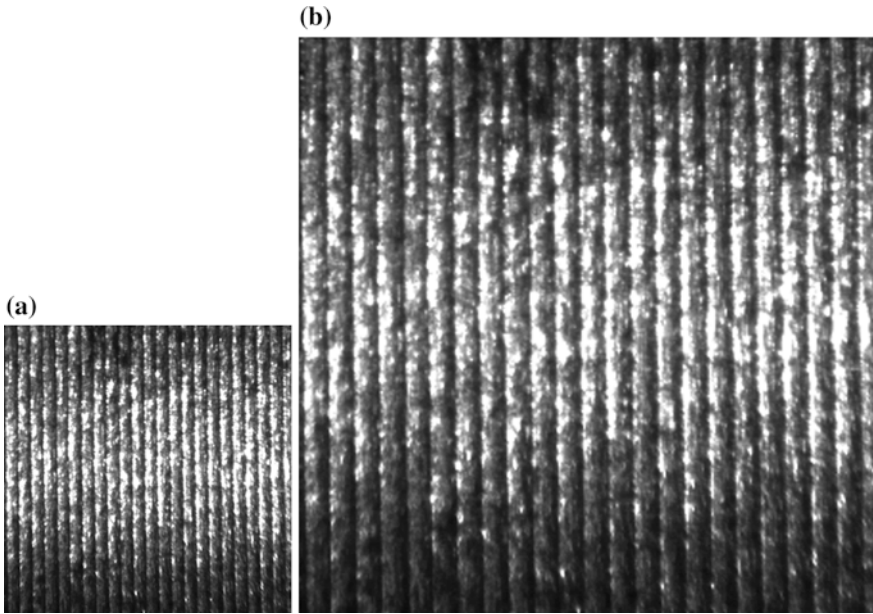
Step 3: Unsharp masking technique is applied on histogram equalized images to enhance the sharpness of image. This technique is mainly used for edge crispening. The sharpen image,  $f_{sharp}(x, y)$  of an original image,  $f(x, y)$  is stated in Eq. (13.16).

$$f_{sharp}(x, y) = f(x, y) + g(x, y) \tag{13.16}$$

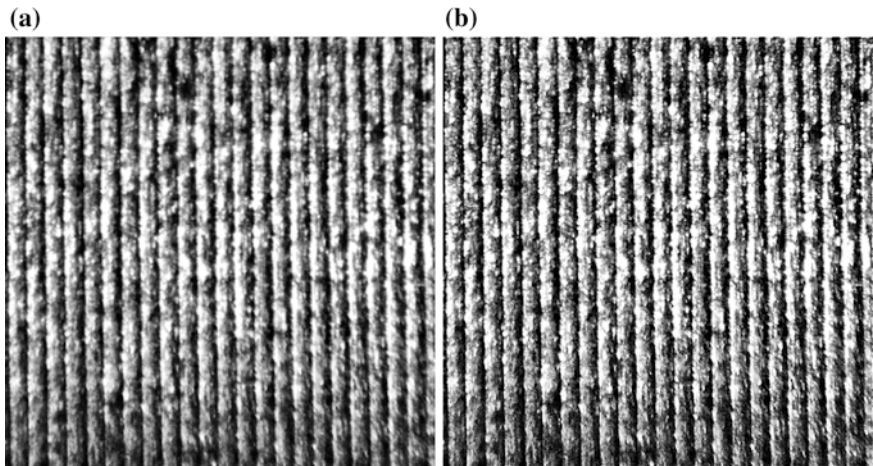
where,  $g(x, y) = f(x, y) - f_{smooth}(x, y)$  and  $f_{smooth}(x, y)$  is a low-pass filtered version by Laplacian filter of  $f(x, y)$ .

Figure 13.10b is the image after applying unsharp masking.

Step 4: After the image enhancement, gray level co-occurrence technique has been applied on all enhanced images. GLCM of rougher surface is off diagonal compared to the GLCM of smoother surface as reported in [21]. Thus, the extracted texture features of GLCM of surface images should be varied systematically as the machined surface texture become rougher or coarser with increase in tool wear. Optimization of PPS and direction parameters is done as described in [21]. The optimum PPS for this study is 13. Contrast, second diagonal moment, homogeneity, inverse difference moment (*IDM*), cluster shade (*CSH*) and cluster prominence (*CPM*)



**Fig. 13.9** Image **a** before applying and **b** after applying bicubic interpolation



**Fig. 13.10** Machined surface image **a** after applying histogram equalization and **b** after applying unsharp masking

features are extracted for tool wear assessment. Contrast and second diagonal moment (*SDM*) are the representative of coarseness in a texture image as stated in Eqs. (13.17) and (13.18), respectively.



$$Contrast = \sum_{i=1}^n \sum_{j=1}^n (i-j)^2 p(i,j) \quad (13.17)$$

$$SDM = \sum_{j=1}^n j \sum_{i=1}^n 0.5|i-j|p(i,j) \quad (13.18)$$

where,  $i$  and  $j$  are gray levels or GLCM matrix co-ordinate positions;  $n$  is the number of gray level and  $p(i,j)$  is the probability of co-occurrence between image pixels.

In the above three equations,  $|i-j|$  or  $(i-j)^2$ , which represent how much the matrix elements are diagonally distributed, are multiplied with probability of co-occurrence of image pixel. Thus, *Contrast* and *SDM* is expected to increase for more off-diagonal distribution of GLCM which is an obvious case for rougher surface.

Homogeneity and *IDM* are the measures of homogeneous or similar pixel occurrence present in a periodic manner in a surface texture. Homogeneity and *IDM* are higher for more homogeneous or smoother surfaces. Homogeneity and *IDM* are stated in Eqs. (13.19) and (13.20), respectively.

$$Homogeneity = \sum_{i=1}^n \sum_{j=1}^n \frac{p(i,j)}{1+|i-j|} \quad (13.19)$$

$$IDM = \sum_{i=1}^n \sum_{j=1}^n \frac{p(i,j)}{1+(i-j)^2} \quad (13.20)$$

Cluster shade (*CSH*) represents the skewness of GLCM histogram as stated in Eq. (13.21). As the surface deteriorated with increase in tool wear, *CSH* becomes negative and gives a decreasing trend.

$$CSH = \sum_{i=1}^n \sum_{j=1}^n p(i,j)[(i+j) - (\mu_x + \mu_y)]^3 \quad (13.21)$$

where,  $\mu_x$  and  $\mu_y$  are the means of every row and every column of GLCM.

Cluster prominence (*CPM*) represents the kurtosis or peakedness of GLCM histogram as stated in Eq. (13.22). As the surface deteriorated with increase in tool wear, *CPM* or the peakedness of GLCM histogram reduces and gives a decreasing trend.

$$CPM = \sum_{i=1}^n \sum_{j=1}^n p(i,j)[(i+j) - (\mu_x + \mu_y)]^4 \quad (13.22)$$

Measured  $VB_{average}$  is plotted with machining time in Fig. 13.11a. An increasing trend of tool flank wear is depicted in this plot. Measured average

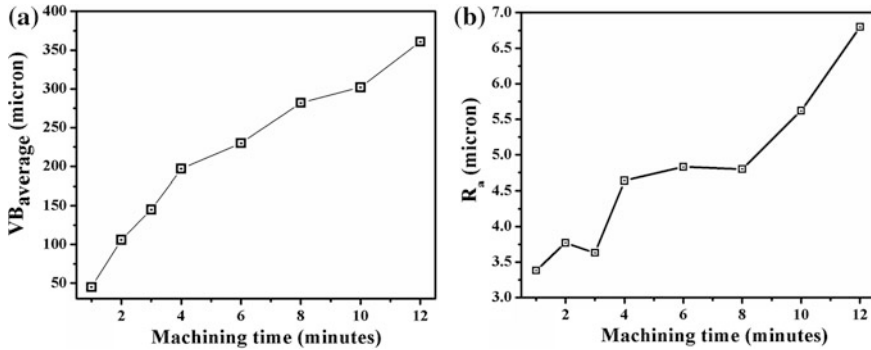


Fig. 13.11 Plot of a  $VB_{average}$  and b average surface roughness ( $R_a$ ) with machining time

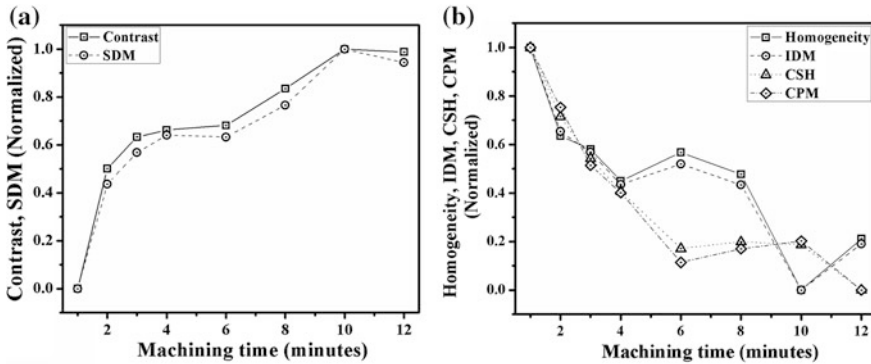
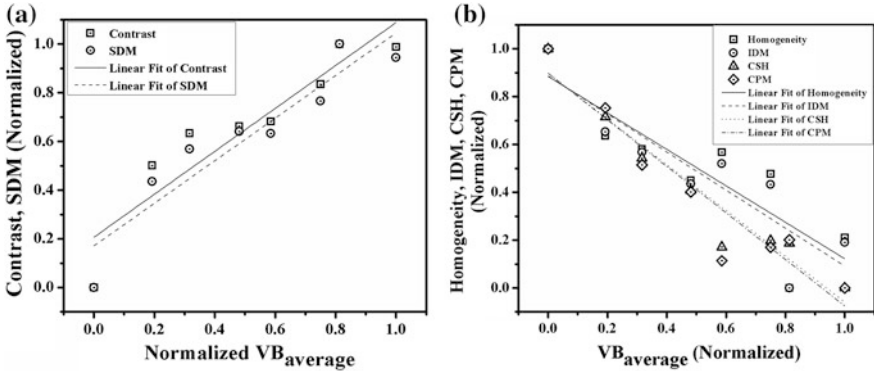


Fig. 13.12 Plots of a Contrast and *SDM* and b Homogeneity, *IDM*, *CSH* and *CPM* versus machining time

surface roughness ( $R_a$ ) by contact type surface profiler are plotted against machining time in Fig. 13.11b. An increasing trend of  $R_a$  with machining time is depicted in this plot.

All the just turned surface images are processed using gray level co-occurrence technique and the extracted features, namely, *Contrast* and *SDM* are plotted with machining time in Fig. 13.12a and other texture features, namely, Homogeneity, *IDM*, *CSH* and *CPM* are plotted with machining time in Fig. 13.12b. The trend of variation for *Contrast* and *SDM* with machining time is increasing, as depicted in Fig. 13.12a. Homogeneity, *IDM*, *CSH* and *CPM* have a decreasing trend with increasing machining time as per Fig. 13.12b.

However, all the extracted features are normalized in 0 to 1 for bringing them in the same scale. As the machining time increases, machined surface becomes rougher or coarser which can be depicted from Fig. 13.11b by observing the



**Fig. 13.13** Plots of **a** Contrast, *SDM* and **b** Homogeneity, *IDM*, *CSH* and *CPM* versus  $VB_{average}$

increasing trend of surface roughness. Thus the texture image also become coarser and distribution of GLCM elements become more scattered or off-diagonal with the increase in roughness. Consequently, Contrast and *SDM* have an increasing trend as explained in Eqs. (13.17) and (13.18) and Homogeneity and *IDM* have a decreasing trend as explained in Eqs. (13.19) and (13.20). As the GLCM becomes off-diagonal for coarser surface, the skewness and peakedness of GLCM histogram reduces. Thus, *CSH* and *CPM* reduces with the increase in surface roughness as explained in Eq. (13.21) and (13.22).

Also there is a need to find the correlation between extracted features and  $VB_{average}$  for determining the condition of the tool from the machined surface images. Thus, all the six extracted features are plotted against normalized values of  $VB_{average}$  and the plots are also fitted linearly. It is depicted from Fig. 13.13a that Contrast and *SDM* have an increasing trend with increase in  $VB_{average}$ . Figure 13.13b shows a decreasing trend of Homogeneity, *IDM*, *CSH* and *CPM* with increase in  $VB_{average}$ . As the tool imprints on the machined surface at the time of machining, the condition of the tool can be observable from the change of machined surface texture.

Values of linear fitting coefficients (or coefficient of determination) and slopes for quantifying the correlation of extracted features with tool flank wear are presented in Table 13.4.

Thus, high linear correlation of Contrast, *SDM*, *CSH* and *CPM* with tool flank wear can be depicted by observing the values of coefficient of determination and slope stated in Table 13.4. It can be seen from this case study that GLCM technique can be used well for depicting the condition of the cutting tool.

**Table 13.4** Coefficient of determination and slope for extracted features

Extracted features	Linear correlation with $VB_{\text{average}}$	
	Coefficient of determination	Slope
Contrast	0.86	0.88
<i>SDM</i>	0.88	0.87
Homogeneity	0.75	-0.76
<i>IDM</i>	0.80	-0.79
<i>CSH</i>	0.93	-0.96
<i>CPM</i>	0.90	-0.97

### 13.6 Concluding Remarks

This chapter discusses about the important aspects of application of digital image processing in conventional machining, specially about the tool condition monitoring. It also tells about the importance of major tool condition monitoring techniques used in industry. Advantages and disadvantages of machine vision techniques applied for tool condition monitoring and surface finish studies over the other techniques have also been presented in this chapter. A review of machine vision techniques applied in conventional machining has been addressed here. Authors have included an example for depicting the progressive tool condition monitoring applying digital image processing from their own research work.

Digital image processing techniques play an important role to determine the tool condition through a non-invasive way. Different states of tool wear, tool breakage, tool chipping and tool fracture can be classified by the image processing of machined surface texture images as well as tool images. Less costly image acquisition hardware is needed to determine the quality of machined surface texture and applied image processing algorithms are simpler for the same. However, capturing the turned surface image in moving condition is also a challenging task due to the high speed of work piece and the curvature effect of turned surface. Real-time tool condition monitoring system utilizing machine vision techniques has not been achieved till date. Main causes of this drawback are blurring effect produced for capturing the image of moving work piece and lower computational speed. Blurring effect can be overcome by using a more robust image acquisition system employing high speed camera, diffused front lighting with triggering capability and using image deblurring technique. However the image area for processing should be optimum with simpler algorithm to cope up with computational speed.

## References

1. Watanabe R, Mizuno Y, Kawasaki A (1999) Determination of non-uniform sintering shrinkage in MIM powder compacts by digital image processing. *Met Mater* 5:163–169
2. Fernandez C, Platero C, Campoy P, Aracil R (1993) Vision system for on-line surface inspection in aluminum casting process. In: Proceedings of IEEE international conference on industrial electronics, control, instrumentation and automation, vol 3, pp 1854–1859
3. Feiste, KL Reichert C, Reimche W, Stegemann D (1998) Process integrated detection and characterization of casting defects. *NDT.net*. <http://www.ndt.net/article/ecndt98/steel/348/348.htm>. Accessed 27 April 2013
4. Sinha P, Muthukumar S, Sivakumar R, Mukherjee SK (2008) Condition monitoring of first mode of metal transfer in friction stir welding by image processing techniques. *Int J Adv Manuf Technol* 36:484–489
5. Balfour C, Smith JS, Amin-Nejad S (2004) Feature correlation for weld image-processing applications. *Int J Prod Res* 42:975–995
6. Lee RS, Hsu QC (1994) Image-processing system for circular-grid analysis in sheet-metal forming. *Exp Mech* 34:108–115
7. O’Leary P (2005) Machine vision for feedback control in a steel rolling mill. *Comput Ind* 56:997–1004
8. Dutta S, Pal SK, Mukhopadhyay S, Sen R (2013) Application of digital image processing in tool condition monitoring: A review. *CIRP J Manuf Sci Technol*. <http://dx.doi.org/10.1016/j.cirpj.2013.02.005>
9. Smith GT (2008) *Cutting tool technology industrial handbook*. Springer, London
10. Teti R, Jemielniak K, O’Donnell G, Dornfeld D (2010) Advanced monitoring of machining operations. *CIRP Ann Manuf Technol* 59:717–739
11. Tool-life testing with single-point turning tools, ISO3685:1993
12. Abellan-Nebot JV, Subirón FR (2010) A review of machining monitoring systems based on artificial intelligence process models. *Int J Adv Manuf Technol* 47:237–257
13. Rehorn AG, Jiang J, Orban PE (2005) State-of-the-art methods and results in tool condition monitoring: a review. *Int J Adv Manuf Technol* 26:693–710
14. Roth JT, Djurdjanovic D, Yang X, Mears L, Kurfess T (2010) Quality and inspection of machining operations- tool condition monitoring. *J Manuf Sci Eng—Trans ASME* 132:041015–1–041015-16
15. Byrne G, Dornfeld D, Inasaki I, Ketteler G, König W, Teti R (1995) Tool condition monitoring (TCM)—the status of research and industrial application. *CIRP Ann Manuf Technol* 44:541–567
16. Prickett PW, Johns C (1999) An overview of approaches to end milling tool monitoring. *Int J Mach Tool Manuf* 39:105–122
17. Dimla E, Dimla S (2000) Sensor signals for tool-wear monitoring in metal cutting operations—a review of methods. *Int J Mach Tool Manuf* 40:1073–1098
18. Jemielniak K (1999) Commercial tool condition monitoring systems. *Int J Adv Manuf Technol* 15:711–721
19. Jemielniak K, Arrazola PJ (2008) Application of AE and cutting force signals in tool condition monitoring in micro-milling. *CIRP J Manuf Sci Technol* 1:97–102
20. Al-Kindi G, Zughaer H (2012) An approach to improved CNC machining using vision-based system. *Mater Manuf Process* 27:765–774
21. Dutta S, Datta A, Chakladar ND, Pal SK et al (2012) Detection of tool condition from the turned surface images using an accurate grey level co-occurrence technique. *Precis Eng* 36:458–466
22. Okada S, Imade M, Miyauchi H et al (1998) 3-D shape measurement of free-form machined surfaces by optical ring imaging system. *IEEE Xplore*. <http://ieeexplore.ieee.org/iel4/5851/15628/00722834.pdf>. Accessed 25 Feb 2013

23. McBride J, Maul C (2004) The 3D measurement and analysis of high precision surfaces using confocal optical methods. *IEICE Trans Electron* E87-C:1261–1267
24. Abouelatta OB (2010) 3D Surface roughness measurement using a light sectioning vision system. *World Congress Engineering*. [http://www.iaeng.org/publication/WCE2010/WCE2010\\_pp698-703.pdf](http://www.iaeng.org/publication/WCE2010/WCE2010_pp698-703.pdf). Accessed 25 Feb 2013
25. Demircioglu P, Durakbasa MN (2011) Investigations on machined metal surfaces through the stylus type and optical 3D instruments and their mathematical modelling with the help of statistical techniques. *Measurement* 44:611–619
26. Jantunen E (2002) A summary of methods applied to tool condition monitoring in drilling. *Int J Mach Tool Manuf* 42:997–1010
27. Liu W, Zheng X, Liu S, Jia Z (2012) A roughness measurement method based on genetic algorithm and neural network for microheterogeneous surface in deep-hole parts. *J Circuit Syst Comp* 21:1250005–1250019
28. Jähne B (2002) *Digital image processing*. Springer, Heidelberg
29. Burke MW (1996) *Image acquisition: handbook of machine vision engineering*. Chapman and Hall, London
30. Weis W (1993) Tool Wear Measurement on basis of optical sensors, vision systems and neuronal networks (application milling). *IEEE Xplore*. <http://ieeexplore.ieee.org>. Accessed 4 March 2010
31. Kurada S, Bradley C (1997) A machine vision system for tool wear assessment. *Tribol Int* 30:295–304
32. Kim JH, Moon DK, Lee DW et al (2002) Tool wear measuring technique on the machine using CCD and exclusive jig. *J Mater Proc Technol* 130–131:668–674
33. Wang WH, Hong GS, Wong YS (2005) Flank wear measurement by successive image analysis. *Comput Ind* 56:816–830
34. Wang WH, Hong GS, Wong YS (2006) Flank wear measurement by a threshold independent method with sub-pixel accuracy. *Int J Mach Tool Manuf* 46:199–207
35. Pfeifer T, Wieggers L (2000) Reliable tool wear monitoring by optimized image and illumination control in machine vision. *Measurement* 28:209–218
36. Jurkovic J, Korosec M, Kopac J (2005) New approach in tool wear measuring technique using CCD vision system. *Int J Mach Tool Manuf* 45:1023–1030
37. Kerr D, Pengilly J, Garwood R (2006) Assessment and visualisation of machine tool wear using computer vision. *Int J Adv Manuf Technol* 28:781–791
38. Makki H, Heinemann RK, Hinduja S, Owodunni OO (2009) Online determination of tool run-out and wear using machine vision and image processing techniques. In: *Innovative production machines and systems*. Cardiff University, Wales, UK
39. Alegre E, Barreiro J, Castejón M, Suarez S (2008) Computer vision and classification techniques on the surface finish control in machining processes. In: *Campilho A, Kamel M (eds) Image analysis and recognition (LNCS 5112)*. Springer, Berlin
40. Tsai DM, Chen JJ, Chert JF (1998) A vision system for surface roughness assessment using neural networks. *Int J Adv Manuf Technol* 14:412–422
41. Bradley C, Wong YS (2001) Surface texture indicators of tool wear—a machine vision approach. *Int J Adv Manuf Technol* 17:435–443
42. Kumar R, Kulasekar P, Dhanasekar B, Ramamoorthy B (2005) Application of digital image magnification for surface roughness evaluation using machine vision. *Int J Mach Tool Manuf* 45:228–234
43. Dhanasekar B, Ramamoorthy B (2008) Assessment of surface roughness based on super resolution reconstruction algorithm. *Int J Adv Manuf Technol* 35:1191–1205
44. Gonzalez RC, Woods RE (2002) *Digital image processing*. Prentice-Hall Inc, Upper Saddle River
45. Sortino M (2003) Application of statistical filtering for optical detection of tool wear. *Int J Mach Tool Manuf* 43:493–497
46. Alegre E, Barreiro J, Fernandez RA, Castejn M (2006) Design of a computer vision system to estimate tool wearing. *Mater Sci Forum* 526:61–66

47. Sahabi HH, Ratnam MM (2008) On-line monitoring of tool wear in turning operation in the presence of tool misalignment. *Int J Adv Manuf Technol* 38:718–727
48. Sahabi HH, Ratnam MM (2009) Assessment of flank wear and nose radius wear from workpiece roughness profile in turning operation using machine vision. *Int J Adv Manuf Technol* 43:11–21
49. Yang M, Kwon O (1996) Crater wear measurement using computer vision and automatic focusing. *J Mater Process Technol* 58:362–367
50. Yang M, Kwon O (1998) A tool condition recognition system using image processing. *Control Eng Pract* 6:1389–1395
51. Zhongxiang H, Lei Z, Jiayu T, Xuehong M, Xiaojun S (2009) Evaluation of three-dimensional surface roughness parameters based on digital image processing. *Int J Adv Manuf Technol* 40:342–348
52. Stemmer M, Pavim A, Adur M et al (2005) Machine vision and neural networks applied to wear classification on cutting tools. In: *Proceedings of the EOS conference on industrial imaging and machine vision*, Munich, Germany
53. Dhanasekar B, Krishna MN, Bhaduri B, Ramamoorthy B (2008) Evaluation of surface roughness based on monochromatic speckle correlation using image processing. *Precis Eng* 32:196–206
54. Nakao Y (2001) Measurement of drilling burr by image processing technique. ASPE. [http://www.aspe.net/publications/Annual\\_2001/PDF/POSTERS/METRO/FORM/1142.PDF](http://www.aspe.net/publications/Annual_2001/PDF/POSTERS/METRO/FORM/1142.PDF). Accessed 6 July 2010
55. Fadare DA, Oni AO (2009) Development and application of a machine vision system for measurement of tool wear. *ARPN J Eng Appl Sci* 4:42–49
56. Josso B, Burton DR, Lalor MJ (2001) Wavelet strategy for surface roughness analysis and characterisation. *Comput Methods Appl Mech Eng* 191:829–842
57. Josso B, Burton DR, Lalor MJ (2002) Frequency normalised wavelet transform for surface roughness analysis and characterisation. *Wear* 252:491–500
58. Niola VN, Quaremba G (2005) A problem of emphasizing features of a surface roughness by means the Discrete Wavelet Transform. *J Mater Process Technol* 164–165:1410–1415
59. Li PY, Hao CY, Zhu SW (2007) Machining tools wear condition detection based on wavelet packet. In: *Proceedings of the sixth international conference on machine learning and cybernetics*, Hong Kong
60. Zawada-Tomkiewicz A (2010) Estimation of surface roughness parameter based on machined surface image. *Metrol Meas Syst XVII*:493–504
61. Dhanasekar B, Ramamoorthy B (2010) Restoration of blurred images for surface roughness evaluation using machine vision. *Tribol Int* 43:268–276
62. Yoon H, Chung SC (2004) Vision inspection of micro-drilling processes on the machine tool. *Trans NAMRI/SME* 32:391–394
63. Liang YT, Chiou YC, Louh CJ (2005) Automatic wear measurement of Ti-based coatings milling via image registration. *IAPRConference*. <http://www.mva-org.jp/Proceedings/CommemorativeDVD/2005/papers/2005088.pdf>. Accessed 25 Feb 2013
64. Inoue S, Konishi M, Imai J (2009) Surface defect inspection of a cutting tool by image processing with neural networks. *Mem Fac Eng—Okayama Univ* 43:55–60
65. Atli AV, Urhan O, Ertürk S, Sönmez M (2006) A computer vision-based fast approach to drilling tool condition monitoring. *Proc Inst Mech Eng B—J Eng Manuf* 220:1409–1415
66. Kassim AA, Mannan MA, Jing M (2000) Machine tool condition monitoring using workpiece surface texture analysis. *Mach Vision Appl* 11:257–263
67. Mannan MA, Kassim AA, Jing M (2000) Application of image and sound analysis techniques to monitor the condition of cutting tools. *Pattern Recogn Lett* 21:969–979
68. Mannan MA, Mian Z, Kassim AA (2004) Tool wear monitoring using a fast Hough transform of images of machined surfaces. *Mach Vis Appl* 15:156–163
69. Kassim AA, Mian Z, Mannan MA (2004) Connectivity oriented fast Hough transform for tool wear monitoring. *Pattern Recogn* 37:1925–1933

70. Bamberger H, Ramachandran S, Hong E, Katz R (2011) Identification of machining chatter marks on surfaces of automotive valve seats. *J Manuf Sci Eng—Trans ASME* 133:041003-1–041003-7
71. Oguamanam DCD, Raafat IH, Taboun SM (1994) A machine vision system for wear monitoring and breakage detection of single-point cutting tools. *Comput Ind Eng* 26:575–598
72. Lanzetta M (2001) A new flexible high-resolution vision sensor for tool condition monitoring. *J Mater Process Technol* 119:73–82
73. Lachance S, Bauer R, Warkentin A (2004) Application of region growing method to evaluate the surface condition of grinding wheels. *Int J Mach Tool Manuf* 44:823–829
74. Duan G, Chen YW, Sukegawa T (2010) Automatic optical flank wear measurement of microdrills using level set for cutting plane segmentation. *Mach Vis Appl* 21:667–676
75. Xiong G, Liu J, Avila A (2011) Cutting tool wear measurement by using active contour model based image processing. In: *Proceedings of the 2011 IEEE international conference on mechatronics and automation*, Beijing, China
76. Galante G, Piacentini M, Ruisi VF (1991) Surface roughness detection by tool image processing. *Wear* 148:211–220
77. Park JJ, Ulsoy AJ (1993) On-line flank wear estimation using an adaptive observer and computer vision, Part 2: Experiment. *J Eng Ind—Trans ASME* 115:37–43
78. Sharan RV, Onwubolu GC (2011) Measurement of end-milling burr using image processing techniques. *Proc I Mech Eng B—J Eng Manuf* 225:448–452
79. Castejon M, Alegre E, Barreiro J, Hernandez LK (2007) On-line tool wear monitoring using geometric descriptors from digital images. *Int J Mach Tool Manuf* 47:1847–1853
80. Barreiro J, Castejon M, Alegre E, Hernandez LK (2008) Use of descriptors based on moments from digital images for tool wear monitoring. *Int J Mach Tool Manuf* 48:1005–1013
81. Alegre E, Rodríguez RA, Barreiro J, Ruiz J (2009) Use of contour signatures and classification methods to optimize the tool life in metal machining. *Estonian J Eng* 15:3–12
82. Tuceryan M, Jain AK (1998) Texture analysis. In: Chen CH, Pau LF, Wang PSP (eds) *The handbook of pattern recognition and computer vision*. World Scientific, Singapore
83. Damodarasamy S, Raman S (1991) Texture analysis using computer vision. *Comput Ind* 16:25–34
84. Hoy DEP, Yu F (1991) Surface quality assessment using computer vision methods. *J Mater Process Technol* 28:265–274
85. Al-Kindi GA, Banl RM, Gilt KF (1992) An application of machine vision in the automated inspection of engineering surface. *Int J Prod Res* 30:241–253
86. Cuthbert L, Huynh V (1992) Statistical analysis of optical Fourier transform patterns for surface texture measurement. *Meas Sci Technol* 3:740–745
87. Ramamoorthy B, Radhakrishnan V (1993) Statistical approaches to surface texture classification. *Wear* 167:155–161
88. Wong FS, Nee AFC, Li XQ, Reisdorj C (1997) Tool condition monitoring using laser scatter pattern. *J Mater Process Technol* 63:205–210
89. Gupta M, Raman S (2001) Machine vision assisted characterization of machined surfaces. *Int J Prod Res* 39:759–784
90. Lee BY, Tarn YS (2001) Surface roughness inspection by computer vision in turning operations. *Int J Mach Tool Manuf* 41:1251–1263
91. Ho SY, Lee KC, Chen SS, Ho SJ (2002) Accurate modeling and prediction of surface roughness by computer vision in turning operations using an adaptive neuro-fuzzy inference system. *Int J Mach Tool Manuf* 42:1441–1446
92. Lee BY, Juan H, Yu SF (2002) A study of computer vision for measuring surface roughness in the turning process. *Int J Adv Manuf Technol* 19:295–301
93. Lee BY, Yu SF, Juan H (2004) The model of surface roughness inspection by vision system in turning. *Mechatronics* 14:129–141



94. Lee KC, Ho SJ, Ho SY (2005) Accurate estimation of surface roughness from texture features of the surface image using an adaptive neuro-fuzzy inference system. *Precis Eng* 29:95–100
95. Arunachalam N, Ramamoorthy B (2007) Texture analysis for grinding wheel wear assessment using machine vision. *Proc Inst Mech Eng B—J Eng Manuf* 221:419–430
96. Khalifa OO, Densibali A, Faris W (2006) Image processing for chatter identification in machining processes. *Int J Adv Manuf Technol* 31:443–449
97. Akbari AA, Fard AM, Chegini AG (2006) An effective image based surface roughness estimation approach using neural network. *IEEEXplore*. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4259888&tag=1](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4259888&tag=1). Accessed 25 Feb 2013
98. Al-Kindi GA, Shirinzadeh B (2007) An evaluation of surface roughness parameters measurement using vision-based data. *Int J Mach Tool Manuf* 47:697–708
99. Elango V, Karunamoorthy L (2008) Effect of lighting conditions in the study of surface roughness by machine vision—an experimental design approach. *Int J Adv Manuf Technol* 37:92–103
100. Ravikumar S, Ramachandran KI, Sugumaran V (2011) Machine learning approach for automated visual inspection of machine components. *Exp Syst Appl* 38:3260–3266
101. Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. *IEEE Trans Syst SMC-3*:610–621
102. Jain R, Kasturi R, Schunck BG (1995) *Machine vision*. McGraw-Hill, London
103. Davies ER (2008) Introduction to texture analysis. In: Mirmedi M, Xie X, Suri J (eds) *Handbook of texture analysis*, 1st edn. Imperial College Press, London
104. Datta A, Dutta S, Pal SK, Sen R, Mukhopadhyay S (2012) Texture analysis of turned surfaces using grey level co-occurrence technique. *Adv Mater Res* 365:38–43
105. Gadelmawla ES, Eladawi AE, Abouelatta OB, Elewa IM (2008) Investigation of the cutting conditions in milling operations using image texture features. *Proc Inst Mech Eng B—J Eng Manuf* 222:1395–1404
106. Gadelmawla ES, Eladawi AE, Abouelatta OB, Elewa IM (2009) Application of computer vision for the prediction of cutting conditions in milling operations. *Proc Inst Mech Eng B—J Eng Manuf* 223:791–800
107. Gadelmawla ES (2011) Estimation of surface roughness for turning operations using image texture features. *Proc Inst Mech Eng B—J Eng* 225:1281–1292
108. Galloway MM (1975) Texture analysis using gray level run lengths. *Comput Graphics* 4:172–179
109. Ramana KV, Ramamoorthy B (1996) Statistical methods to compare the texture features of machined surfaces. *Pattern Recogn* 29:1447–1459
110. Kassim AA, Mannan MA, Mian Z (2007) Texture analysis methods for tool condition monitoring. *Image Vis Comput* 25:1080–1090
111. Russ JC (1998) Fractal dimension measurement of engineering surfaces. *Int J Mach Tool Manuf* 38:567–571
112. Kang MC, Kim JS, Kim KH (2005) Fractal dimension analysis of machined surface depending on coated tool wear. *Surf Coat Technol* 193:259–265
113. Karthik A, Chandra S, Ramamoorthy B, Das S (1997) 3D tool wear measurement and visualisation using stereo imaging. *Int J Mach Tool Manuf* 37:1573–1581
114. Prasad KN, Ramamoorthy B (2001) Tool wear evaluation by stereo vision and prediction by artificial neural network. *J Mater Process Technol* 112:43–52
115. Otto T, Kurik L, Papstel J (2003) A digital measuring module for tool wear estimation. In: Katalinic B (ed) *DAAAM international scientific book 2003*. DAAAM International, Vienna
116. Schmitt R, Hermes R, Stemmer M et al (2005) Machine vision prototype for flank wear measurement on milling tools. In: 38th CIRP manufacturing systems seminar, Florianopolis, Brazil
117. Wang WH, Wong YS, Hong GS (2006) 3D measurement of crater wear by phase shifting method. *Wear* 261:164–171

118. Devillez A, Lesko S, Mozerc W (2004) Cutting tool crater wear measurement with white light interferometry. *Wear* 256:56–65
119. Dawson TG, Kurfess TR (2005) Quantification of tool wear using white light interferometry and three-dimensional computational metrology. *Int J Mach Tool Manuf* 45:591–596
120. Ng KW, Moon KS (2001) Measurement of 3-D tool wear based on focus error and micro-coordinate measuring system. [http://aspe.net/publications/Annual\\_2001/pdf/posters/metro/instrum/1207.pdf](http://aspe.net/publications/Annual_2001/pdf/posters/metro/instrum/1207.pdf). Accessed on 27 Feb 2013
121. Liang YT, Chiou YC (2006) An effective drilling wear measurement based on visual inspection technique. [www.atlantis-press.com/php/download\\_paper.php?id=267](http://www.atlantis-press.com/php/download_paper.php?id=267). Accessed 26 Feb 2013
122. Su JC, Huang CK, Tarng YS (2006) An automated flank wear measurement of microdrills using machine vision. *J Mater Process Technol* 180:328–335
123. Yasui H, Haraki Y, Sakata M (2001) Development of automatic image processing system for evaluation of wheel surface condition in ultra-smoothness grinding. [http://aspe.net/publications/Annual\\_2001/pdf/papers/pfab/1191.pdf](http://aspe.net/publications/Annual_2001/pdf/papers/pfab/1191.pdf). Accessed 28 Feb 2013
124. Heger T, Pandit M (2004) Optical wear assessment for grinding tools. *J Electron Imag* 13:450–461
125. Palani S, Natarajan U (2011) Prediction of surface roughness in CNC end milling by machine vision system using artificial neural network based on 2D Fourier transform. *Int J Adv Manuf Technol* 54:1033–1042
126. Schmähling J, Hamprecht FA, Hoffmann DMP (2006) A three-dimensional measure of surface roughness based on mathematical morphology. *Int J Mach Tool Manuf* 46:1764–1769
127. Senin N, Ziliotti M, Groppetti R (2007) Three-dimensional surface topography segmentation through clustering. *Wear* 262:395–410
128. Sarma PMMS, Karunamoorthy L, Palanikumar K (2009) Surface roughness parameters evaluation in machining GFRP composites by PCD tool using digital image processing. *J Reinf Plast Comp* 28:1567–1585
129. Dhanasekar B, Ramamoorthy B (2006) Evaluation of surface roughness using a image processing and machine vision system. *MAPAN—J Metrol Soc I* 21:9–15
130. Narayanan MR, Gowri S, Krishna MM (2007) On line surface roughness measurement using image processing and machine vision. [http://www.iaeng.org/publication/WCE2007/WCE2007\\_pp645-647.pdf](http://www.iaeng.org/publication/WCE2007/WCE2007_pp645-647.pdf). Accessed 28 Feb 2013
131. Gadelmawla ES (2004) A vision system for surface roughness characterization using the gray level co-occurrence matrix. *NDT E Int* 37:577–588
132. Priya P, Ramamoorthy B (2010) Machine vision for surface roughness assessment of inclined components. *Key Eng Mater* 437:141–144
133. Kassim AA, Mian Z, Mannan MA (2006) Tool condition classification using Hidden Markov Model based on fractal analysis of machined surface textures. *Mach Vis Appl* 17:327–336
134. Vesselenyi T, Dzitac I, Dzitac S, Vaida V (2008) Surface roughness image analysis using quasi-fractal characteristics and fuzzy clustering methods. *Int J Comput Comm Control* 3:304–316
135. Tsai DM, Wu SK (2000) Automated surface inspection using Gabor filters. *Int J Adv Manuf Technol* 16:474–482
136. Zhang X, Krewet C, Kuhlenkötter B (2006) Automatic classification of defects on the product surface in grinding and polishing. *Int J Mach Tool Manuf* 46:59–69

# Chapter 14

## Formability and Simulative Tests in Modern Sheet Metal Forming Education

P. A. F. Martins, L. Montanari, V. A. Cristino and M. B. Silva

**Abstract** This chapter provides the theoretical and experimental background to the utilization of circle grid analysis in conjunction with formability and simulative laboratory tests at undergraduate and graduate courses of metal forming with the objective of exposing students to machine-tools, tool systems and practical problems that are commonly found in daily production of sheet metal parts. The presentation starts with a description of strain measures and measurements techniques, extends this background to plastic instability and fracture, introduces the forming limit diagrams and systematizes the associated formability tests, and reviews the process simulation tests aimed at characterizing bendability, stretchability and drawability of blanks. Then, presentation proceed with a comprehensive description of a flexible, versatile and robust laboratory tool system that was developed by the authors to support education and basic research activities of sheet metal forming. The last part of the chapter presents the fundamentals of a simple finite element computer program developed by the authors that gives support to sheet metal forming laboratory tests and provides students with a strong link between experimentation and numerical simulation.

### 14.1 Introduction

Despite its importance to society in general and engineering in particular, sheet metal forming developed throughout the years as an art with major techniques and procedures being transmitted from craftsmen to apprentices as professional secrets with no scientific background.

---

P. A. F. Martins (✉) · V. A. Cristino · M. B. Silva  
Instituto Superior Técnico, University of Lisbon, Av. Rovisco Pais,  
Lisbon 1049-001, Portugal  
e-mail: pmartins@ist.utl.pt

L. Montanari  
Escola de Engenharia de São Carlos, University of São Paulo, Av. Trabalhador  
São-Carlense, 400, São Carlos, SP CEP: 13566-590, Brasil

The widespread introduction of manufacturing courses at the graduate and post-graduate levels of mechanical, materials and industrial engineering curricula during the second half of the twentieth century stimulate the understanding of sheet metal forming process in the light of existing scientific knowledge. Although some courses were designed to be very descriptive and mainly focused on materials, equipment and products because of the wrong assumption that further knowledge should be acquired on the job training, there were other courses that successfully integrate knowledge on materials science, solid mechanics and thermodynamics to provide students with a deep understanding on the fundamentals of sheet metal forming.

The last vision prevailed over the purely descriptive courses and modern engineering curricula aim to prepare students to understand the major operating parameters, the formability constraints and to utilize sophisticated analytical and numerical models to estimate the loads, the pressures and the major field variables associated with the production of sheet metal parts. However, the integration of basic scientific knowledge in sheet metal forming courses is only part of the solution to train young engineers capable of solving real problems. The hopeful assumption that young engineers, well versed in the theoretical fundamentals would readily acquire technical know-how to project, design and fabricate tooling systems and to solve daily production issues only works if the young engineers are naturally gifted. The remaining young engineers will normally encompass various types of painful learning experiences before acquiring the required technical skills.

One solution to overcome this problem is to strengthen the sheet metal forming courses with laboratory training focused on the practical aspects of tool design and on the analysis and evaluation of the interrelationship of material, process and shape in the formability of sheet metal (hereafter designated as 'blank').

This chapter is focused on the importance of hands-on laboratory tests in sheet metal forming education. Presentation combines the fundamental concepts of strain measures, strain measurement, forming limits and finite element modelling with laboratory tests for evaluating sheet metal formability by means of experimental methods based on the forming limit diagrams and based on simulative operations aimed at characterizing bendability, stretchability and drawability of the blanks. A flexible and versatile tool system developed by the authors that is capable of being used in both education and research activities is comprehensively described for the benefit of those readers who want to design a similar tool for supporting their teaching activities or for those who simply want to understand how this solution can be done well at a low cost.

## **14.2 Strain Measures and Measurement**

One of the main difficulties of the graduate students attending sheet metal forming courses is the understanding of the physics behind strain measures that are commonly utilized in large elasto-plastic deformations. This section reviews these

strain measures, describes the experimental techniques for their measurement and graphical representation in the principal strain space (also known as the forming limit diagram), and introduces the formability limits by necking and fracture.

### 14.2.1 Strain Measures

Deformation in sheet metal forming generally comprises of stretch, rigid body rotation and translation. Stretch provides the changes in shape and the modifications in the state of stress, whereas rigid body rotation and translation do not provide variations in shape and internal stresses. Figure 14.1 presents the three basic modes of deformation applied to a one-dimensional rod.

Appropriate strain measures are necessary to evaluate the stress response, to predict the formability limits and to determine the force and energy requirements of the sheet metal forming process independently of rigid body rotation and translation. In principle, the stretch ratio  $\lambda$  between the final  $l$  and the initial  $l_0$  lengths of the one-dimensional rod pictured in Fig. 14.1 could be utilized as a strain measure for large elasto-plastic deformations,

$$\lambda = \frac{l}{l_0} \quad (1)$$

However, the application of a strain measure exclusively based in the stretch ratio  $\lambda$  would lead to values  $\lambda \approx 1$  (with major differences in the second and third decimal places) in case of metallic materials undergoing small to medium deformations and to values  $\lambda > 3$  in case of elastomers experiencing very large deformations. In addition, the absence of deformation would be characterized by a value of strain  $\lambda = 1$  instead of  $\lambda = 0$ , as one would normally expect from a physical point of view.

As a result of this, the strain measures commonly utilized in large deformations are preferentially derived from finite strain tensors that are defined in an element of material undergoing finite deformation between the original  $\mathbf{X}$  (or undeformed) and the final  $\mathbf{x}$  (or deformed) configurations (Fig. 14.2). The coordinates  $\mathbf{X}$  and  $\mathbf{x}$  allow mapping the present position  $\mathbf{x}$  as a function of the original position  $\mathbf{X}$  as follows,

$$\mathbf{x} = \mathbf{x}(\mathbf{X}, t) \quad (2)$$

and the components  $d\mathbf{x}$  of the final differential line element  $P'Q'$  are obtained from the components  $d\mathbf{X}$  of the original differential line element  $PQ$  embedded in the material by means of the deformation gradient tensor  $\mathbf{F}$ ,

$$d\mathbf{x} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}} d\mathbf{X} \quad \Rightarrow \quad \mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}} \quad (3)$$

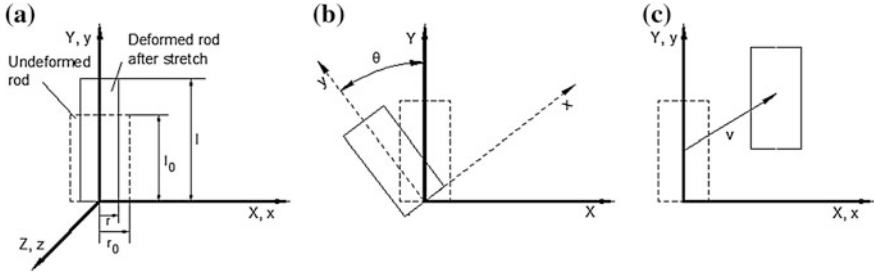
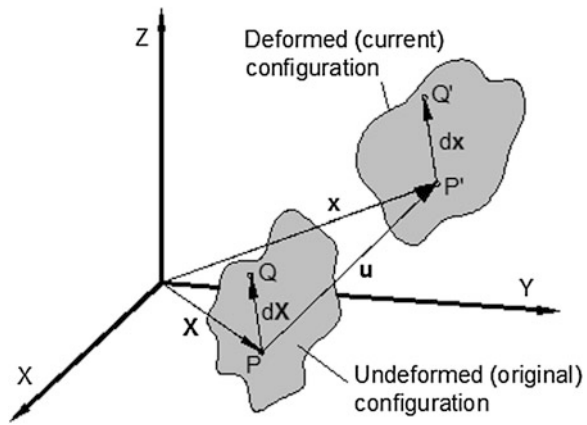


Fig. 14.1 One-dimensional rod undergoing, a stretching, b rigid body rotation, c translation

Fig. 14.2 Finite deformation between the original and deformed configurations



The deformation gradient  $\mathbf{F}$  is a non-symmetric tensor that provides a complete description of the deformation. Stretching is included as well as rigid body rotation but translation will not be addressed in the presentation.

The most commonly utilized strain measures in large deformation are built from the deformation gradient tensor  $\mathbf{F}$  as follows,

$$\begin{aligned}
 \text{Green-Lagrange strain } \mathbf{E} &= \frac{1}{2}(\mathbf{C} - \mathbf{I}) \\
 \text{Almansi strain } \mathbf{e} &= \frac{1}{2}(\mathbf{I} - \mathbf{B}^{-1}) \\
 \text{Logarithmic strain } \boldsymbol{\varepsilon} &= -\frac{1}{2} \ln(\mathbf{B}^{-1})
 \end{aligned}
 \tag{4}$$

The symbol  $\mathbf{I}$  denotes the identity tensor,  $\mathbf{B} = \mathbf{F}\mathbf{F}^T$  is the left Cauchy-Green tensor and  $\mathbf{C} = \mathbf{F}^T\mathbf{F}$  is the right Cauchy-Green tensor. The strain measures in Eq. (4) are appropriate for large deformations because the finite strain tensors  $\mathbf{B}$  and  $\mathbf{C}$  are symmetric and result in the identity tensor  $\mathbf{I}$  thereby, providing zero strains, in case of rigid body rotation without stretch.

**Table 14.1** One-dimensional definition of strain measures

n	Strain measure	Definition
-2	Green–Lagrange	$\frac{1}{2} \left( \frac{l^2}{l_0^2} - 1 \right)$
-1	Engineering	$\left( \frac{l-l_0}{l_0} \right)$
0	Logarithmic	$\lim_{n \rightarrow 0} \left( \frac{1}{n} \left( 1 - \left( \frac{l_0}{l} \right)^n \right) \right) = \ln \left( \frac{l}{l_0} \right)$
1	Swainger	$\left( \frac{l-l_0}{l} \right)$
2	Almansi	$\frac{1}{2} \left( 1 - \frac{l_0^2}{l^2} \right)$

Further understanding of the strain measures in Eq. (4) is given by considering the one-dimensional rod undergoing pure stretch in the absence of rigid body rotation that is illustrated in Fig. 14.1a. In such a case, because there is no rotation, the co-rotational reference system  $xyz$ , which rotates with the material in case of rigid body rotation, is coincident with the material reference system  $XYZ$  and, therefore, the deformation gradient  $\mathbf{F}$  is identical to the stretch ratio  $\lambda$  between the final  $l$  and the initial  $l_0$  lengths of the rod.

The corresponding left  $\mathbf{B}$  and the right  $\mathbf{C}$  Cauchy–Green strain tensors are equal to  $\lambda^2 = l^2/l_0^2$  and the strain measures in Eq. (4) can be simply written as follows,

$$\begin{aligned} \text{Green–Lagrange strain } E &= \frac{1}{2} \left( \frac{l^2}{l_0^2} - 1 \right) \\ \text{Almansi strain } e &= \frac{1}{2} \left( 1 - \frac{l_0^2}{l^2} \right) \\ \text{Logarithmic strain } \varepsilon &= \ln \left( \frac{l}{l_0} \right) \end{aligned} \quad (5)$$

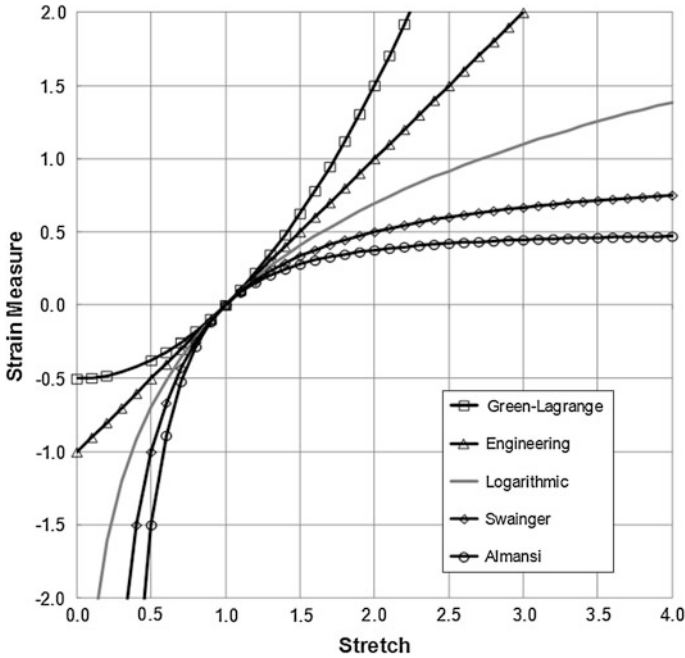
The generalization of Eq. (5) into a single equation from where all known strain measures may be derived results in the following algebraic expression,

$$\text{Strain} = \frac{1}{n} \left( 1 - \left( \frac{l_0}{l} \right)^n \right) \quad (6)$$

where  $n$  is an integer taking values in accordance with Table 14.1.

Equation (6) allows the Green–Lagrange, logarithmic and Almansi strain measures to be listed along with the classical engineering and Swainger strain measures. The classical engineering strain is commonly applied in linear small deformations (familiar to all graduate students) and the Swainger strain is a non-linear approximation of the Almansi strain. However, what distinguishes the Green–Lagrange, logarithmic and Almansi strains from the engineering and Swainger strains is their tensorial character.

Figure 14.3 provides a graphical representation of the five strain measures that are listed in Table 14.1. As seen, only the logarithmic strain has the natural, or intuitive, property of becoming infinite as stretch approaches infinity or becoming



**Fig. 14.3** Comparison between various strain measures in a rod undergoing pure stretch in the absence of rigid body rotation

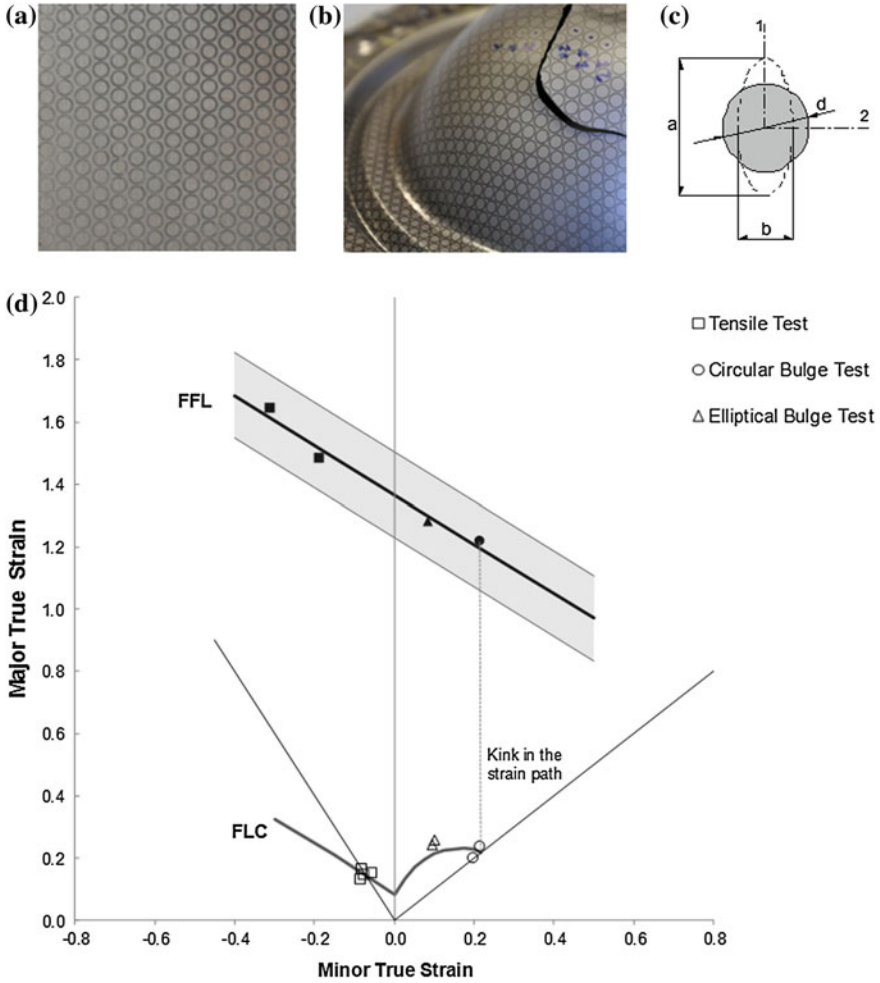
negative infinite as stretch goes to zero. The Almansi strain asymptotes to one-half in tension, while the Green–Lagrange strain asymptotes to negative one-half in compression. Both the Almansi and Green–Lagrange strains are approximately equal to the engineering strain when deformation is small.

In what follows, logarithmic strains will be preferentially utilized.

### 14.2.2 Strain Measurement by Circle-Grid Analysis

The experimental procedure that is commonly utilized for measuring strains in sheet metal forming was originally proposed by Keeler and Goodwin [1, 2]. The procedure is designated as circle-grid analysis and involves etching or imprinting a grid of circles (usually with 2–5 mm in diameter  $d$ ) on the surface of the blanks before forming and measuring the major  $a$  and minor  $b$  axes of the ellipses that result from the plastic deformation of the circles during sheet metal forming tests (Fig. 14.4). The values of the principal logarithmic strains at the surface of the deformed blanks are calculated from,

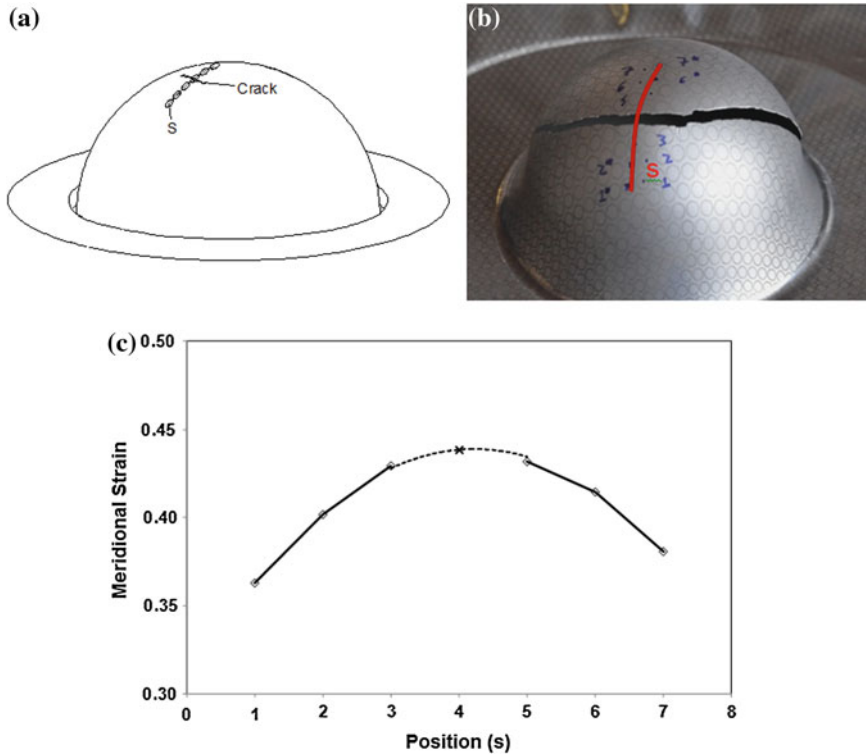




**Fig. 14.4** Fundamentals of strain measurement by *circle grid* analysis; **a** original *circle grid* etched on the surface of the blank made from aluminum AA1050-H111, **b** deformed blank shows *circles* distorted into *ellipses* as a result of the hydraulic bulge formability test, **c** the major and minor axis of the *ellipse* are measured to determine the principal logarithmic strains at the surface of the deformed blank and **d** forming limit curve (FLC) and fracture forming limit (FFL) derived from *circle grid* analysis

$$\epsilon_1 = \ln\left(\frac{a}{d}\right), \epsilon_2 = \ln\left(\frac{b}{d}\right) \tag{7}$$

and can be graphically represented in the forming limit diagram (Fig. 14.4d). Measurements of the critical major and minor strains from the deformed circles that have been previously etched or imprinted over the surfaces of the blanks



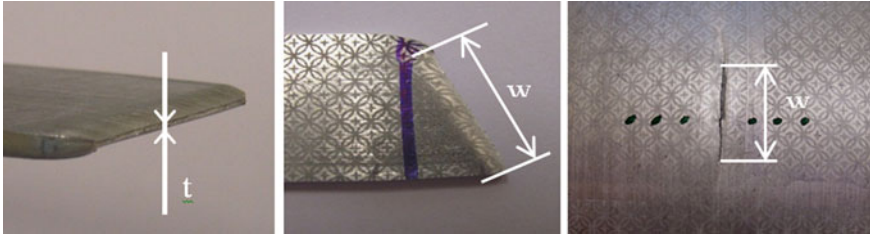
**Fig. 14.5** Zurich no. 5 procedure for determining the strains at necking by means of the strains retrieved from measurements in adjacent deformed *circles* along a direction perpendicular to the crack. **a** schematic drawing, **b** test specimen **c** interpolation procedure

utilized in different formability tests allow determining the forming limit curve (FLC), the fracture forming limit (FFL) and the influence of process operative conditions in the strain loading paths and associated failure limits.

The FLC represents the boundary between the strain combinations which produce material failure by necking (plastic instability) and those that are admissible in sheet forming. Its location and shape is constructed by measuring the principal strains, at critical areas of the blanks, from deformed circles placed just outside the neck (that is adjacent to the region of intense localization). This is because these circles represent the condition of the uniformly thinned sheet just before necking occurs [3]. The procedure is graphically illustrated in Fig. 14.5.

The intersection of the FLC with the major strain axis is found to occur at values in fair agreement with the value of the strain hardening exponent  $n$  of the stress-strain curve  $\sigma = K \epsilon^n$  obtained by means of uniaxial tension testing.

The FFL is more difficult to determine than the FLC. Application of grids even with very small circles in order to obtain strains in the necking region after it forms and, therefore, close to the fracture provides strain values that cannot be



**Fig. 14.6** Experimental procedures that are utilized for obtaining the experimental values of strain along the **a** thickness and the **b** width directions at the onset of fracture (FFL) for specimens utilized in tensile and **c** bulge formability tests

considered the fracture strains. This is because such grids create measurement problems and suffer from sensitivity to the initial size of the circles used in the grids due to the inhomogeneous deformation in the neighbourhood of the crack.

As a result of this, the experimental procedure for constructing the FFL requires measuring of thickness before and after fracture at several places along the crack in order to obtain the ‘gauge length’ strains. The strain in the width direction is obtained differently as it is illustrated in Fig. 14.6 for the case of tensile and bulge formability tests.

As seen in Fig. 14.6, in case of uniaxial tension tests measurements are directly taken from the width of the specimens whereas in case of hydraulic bulge tests, measurements require the utilization of the imprinted grid of circles in order to obtain the initial and deformed reference lengths. The third fracture strain component, in the plane of the sheet with direction perpendicular to the crack, is determined by volume constancy knowing the two other strains [4].

The strains at the onset of fracture (FFL) are commonly approximated by a straight line  $\varepsilon_1 + \varepsilon_2 = c$  falling from left to right and in good agreement with the condition of constant thickness strain at fracture (given by a slope of ‘-1’) that was proposed by Atkins [5].

The large distance between neck formation (FLC) and collapse by fracture (FFL) that is observed in Fig. 14.4d indicates that AA1050-H111 is a very ductile material that allows a considerable through thickness strain between neck initiation and fracture. As seen in Fig. 14.4d, at the onset of local instability implying transition from the FLC representing necking towards the FFL, a sharp bend occurs in the strain path when testing is done with hydraulic bulge tests. The strain paths of hydraulic bulge tests show a kink after neck initiation towards vertical direction, corresponding to plane strain conditions, as schematically plotted by the grey dashed line. The strain paths of tensile formability tests also undergo a significant change of strain ratio from slope -2 to a steeper one although not to vertical direction. The absence of a sharp kink of the strain path into vertical direction, and a less abrupt bend instead, is due to the fact that major and minor strains after the onset of necking do not coincide with the original pulling direction. A comprehensive analysis on the direction of the strain paths in the second quadrant of the forming limit diagram can be found in the work of Atkins [5, 6].

### 14.2.3 Strain Measurement by Square-Grid Analysis

Circle-grid analysis has proved successful in both research and press shop applications because it is an easy and efficient experimental technique to evaluate the strain loading paths, to determine the material formability limits and to visualize the critical areas of sheet metal parts. However, the method should not be employed in the analysis of non-homogeneous deformations (where a circle does not deform into an ellipse), in the analysis of non-coaxial strain paths (where the principal strain rate directions do not rotate with respect to the material), or in situations where it is difficult to distinguish the major and minor diameters of the deformed circles [7, 8]. Still, these problems are often neglected because it is generally not possible from a homogeneously deformed circle, to deduce whether its deformation has taken place along a coaxial or a non-coaxial strain path.

Square-grid analysis for sheet metal forming was developed by Sowerby et al. [9] with the objective of avoiding the aforementioned problems. The fundamentals of the technique are illustrated in Fig. 14.7 and the following presentation is limited to two-dimensional (plane strain or axisymmetric) or simple three-dimensional cross section symmetric surfaces.

The first step in measuring the strains is to divide the square-grid into triangles so that strains can be determined by evaluating the final configuration  $OP'Q'$  with reference to its original (or undeformed) configuration  $OPQ$ . Assuming that the nodal coordinates after deformation can be related with the original coordinates by means of the continuous mapping transformation provided by Eq. (2), the variation in the side-length of each triangle can be computed from the deformation gradient tensor  $\mathbf{F}$  (Eq. 3) that characterizes the local deformation at each nodal point of the square-grid.

The evaluation of the deformation gradient tensor  $\mathbf{F}$  from the initial and final nodal coordinates of the side lengths  $OP$  and  $OQ$  is performed by inverting the following system of equations,

$$\begin{cases} d\mathbf{x}_p = \mathbf{F} d\mathbf{X}_p \\ d\mathbf{x}_q = \mathbf{F} d\mathbf{X}_q \end{cases} \Rightarrow \begin{bmatrix} x_p \\ x_q \\ y_p \\ y_q \end{bmatrix} = \begin{bmatrix} X_p & Y_p & 0 & 0 \\ X_q & Y_q & 0 & 0 \\ 0 & 0 & X_p & Y_p \\ 0 & 0 & X_q & Y_q \end{bmatrix} \begin{bmatrix} F_{11} \\ F_{12} \\ F_{21} \\ F_{22} \end{bmatrix} \quad (8)$$

in order to obtain the four coefficients of  $\mathbf{F}$ ,

$$\begin{bmatrix} F_{11} \\ F_{12} \\ F_{21} \\ F_{22} \end{bmatrix} = \frac{1}{(X_p Y_q - X_q Y_p)} \begin{bmatrix} Y_q & -Y_p & 0 & 0 \\ -X_q & X_p & 0 & 0 \\ 0 & 0 & Y_q & -Y_p \\ 0 & 0 & -X_q & X_p \end{bmatrix} \begin{bmatrix} x_p \\ x_q \\ y_p \\ y_q \end{bmatrix} \quad (9)$$

The Green-Lagrange strain measure (Eq. 4) is then computed from the right Cauchy-Green tensor  $\mathbf{C} = \mathbf{F}^T \mathbf{F}$ ,

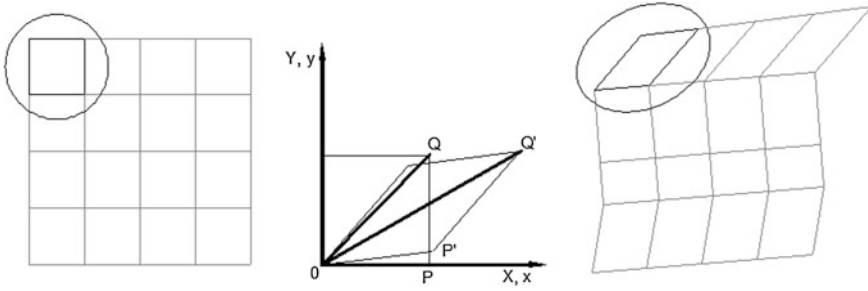


Fig. 14.7 Determination of the principal strains from a square-grid

$$\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \mathbf{F}^T \mathbf{F} = \begin{bmatrix} F_{11}^2 + F_{21}^2 & F_{11}F_{12} + F_{21}F_{22} \\ F_{12}F_{11} + F_{22}F_{21} & F_{12}^2 + F_{22}^2 \end{bmatrix} \quad (10)$$

to obtain  $\mathbf{E} = \frac{1}{2}(\mathbf{C} - \mathbf{I})$ , and the correspondent principal Green–Lagrange strains are given by,

$$E_1, E_2 = \frac{E_{11} + E_{22}}{2} \pm \sqrt{\left(\frac{E_{11} - E_{22}}{2}\right)^2 + E_{12}^2} \quad (11)$$

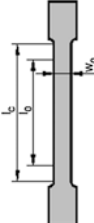
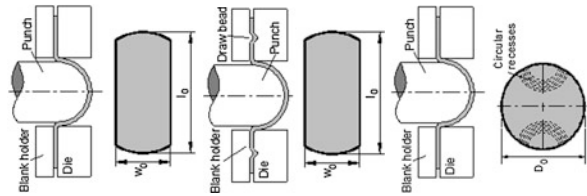
The principal logarithmic strains in each triangle of the square-grid are then determined from the relation  $\varepsilon = \ln \sqrt{1 + 2E}$  between the two strain measures that can be obtained from Table 14.1, in conjunction with the incompressibility condition,

$$\varepsilon_1 = \ln \sqrt{1 + 2E_1} \quad \varepsilon_2 = \ln \sqrt{1 + 2E_2} \quad \varepsilon_3 = -(\varepsilon_1 + \varepsilon_2) \quad (12)$$

Despite the advantages claimed for the utilization of square grids, its implementation in computer aided strain measurement and analysis systems may experience difficulties in providing accurate measurements in complex three dimensional surfaces of real stamped or deep drawn parts. This is because the utilization of square grids requires mapping the elements on the deformed parts back into a plane representing the initial blank by means of geometric modelling, which is dependent on the assumptions and strategies incorporated into the mapping strategy [9]. In some gross approximations, for example, the area of the quadrilaterals is assumed not to change and, therefore, variation of thickness will not be taken into account.

However, the main reason for not using square grids in this presentation is due to the fact that students should be preferentially exposed to manual and semi-automatic measuring procedures where square-grid analysis is impossible to be utilized.

**Table 14.2** Formability test methods based on forming limit diagrams

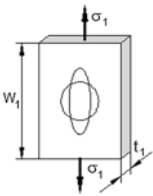
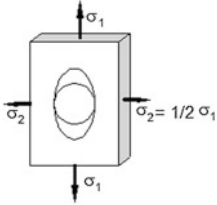
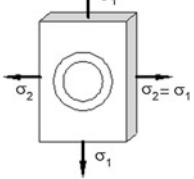
Deformation mode	Test	State of strain	State of stress	Schematic drawing
Uniaxial	Tensile test	$\epsilon_1 > 0$ $\epsilon_2 = \epsilon_3 < 0$ $\epsilon_2 = \epsilon_3 = -\epsilon_1/2$	$\sigma_1 > 0$ $\sigma_2 = \sigma_3 = 0$	
	Plane strain (and the transition region between uniaxial and biaxial deformation modes)	$ \epsilon_1 > 0 $ $-\epsilon_1/2 < \epsilon_2 < \epsilon_1$ $\epsilon_3 = -(\epsilon_1 + \epsilon_2)$	$\sigma_1 > 0$ $\sigma_1 > \sigma_2 > 0$ $\sigma_3 = 0$	
	Hecker test [11] (variant to Nakazima's test with draw beads to avoid drawing-in)			
	Hasek test [12] (variant to Nakazima test employing circular blanks with recesses)			

(continued)

Table 14.2 (continued)

Deformation mode	Test	State of strain	State of stress	Schematic drawing
Biaxial	Hemispherical dome test	$\epsilon_1 = \epsilon_2 > 0$ $\epsilon_3 < 0$ $\epsilon_1 = \epsilon_2 = -\epsilon_3/2$	$\sigma_1 = \sigma_2 > 0$ $\sigma_3 = 0$	
	Hydraulic bulge test			
	Marciniak test [13]			

**Table 14.3** Failure by plastic instability as a function of the deformation mode

Deformation mode	State of strain	State of stress	Schematic drawing
Uniaxial	Diffuse necking: $\epsilon_1 = n$ $\epsilon_2 = \epsilon_3 = -\frac{n}{2}$ $\bar{\epsilon} = n$	$\sigma_1 > 0$ $\sigma_2 = \sigma_3 = 0$	
	Localized necking: $\epsilon_1 = 2n$ $\epsilon_2 = \epsilon_3 = -n$ $\bar{\epsilon} = 2n$		
Plane strain	$\epsilon_1 = n$ $\epsilon_2 = 0$ $\epsilon_3 = -n$ $\bar{\epsilon} = \frac{2n}{\sqrt{3}}$	$\sigma_1 > 0$ $\sigma_2 = \frac{\sigma_1}{2}$ $\sigma_3 = 0$	
Biaxial	$\epsilon_1 = \epsilon_2 = n$ $\epsilon_3 = -2n$ $\bar{\epsilon} = 2n$	$\sigma_1 = \sigma_2 > 0$ $\sigma_3 = 0$	

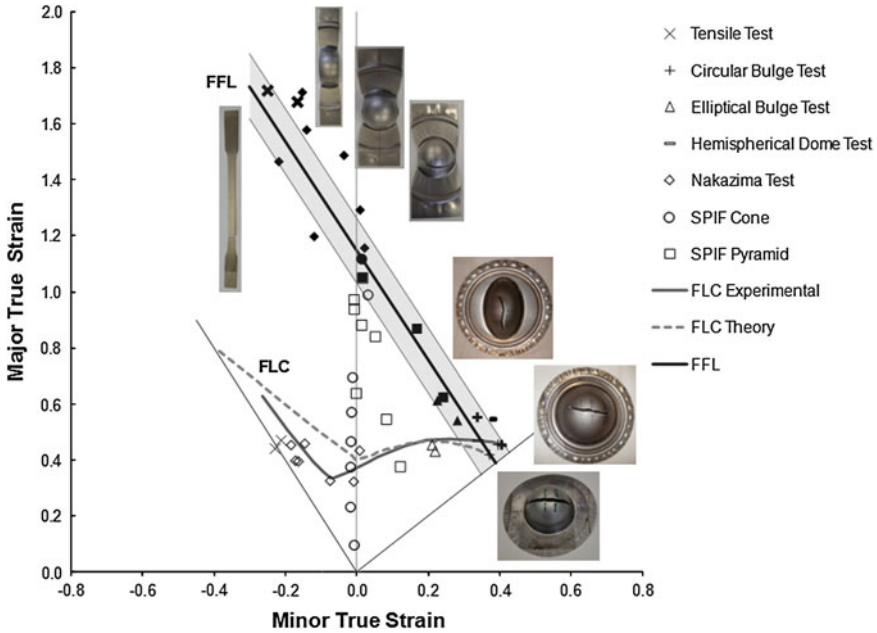
### 14.3 Evaluating Formability by Tests Based on Forming Limit Diagrams

The evaluation of sheet metal formability by means of laboratory test methods based on forming limit diagrams is aimed to experimentally determine the critical strains at failure by necking (plastic instability, FLC) and fracture (FFL). The formability testing methods based on forming limit diagrams can be classified as uniaxial, plane-strain and biaxial formability tests as a function of the deformation mode (that is, the strain loading path) that is characteristic of each specific test (Table 14.2).

Figure 14.8 shows experimental results of the formability tests listed in Table 14.2 that were obtained by the authors for AISI 304L stainless steel sheets with 0.5 mm thickness. By plotting the measured strains it was possible to construct the FLC and the FFL that are shown in the Fig. 14.8, the open markers plotted refer to the experimental strains at necking and the solid markers refer to the experimental strains at failure.

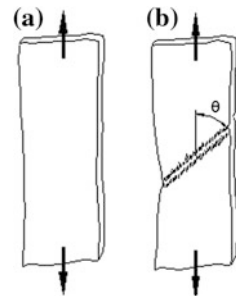
Failure by plastic instability occurs by the development of a diffuse or localized neck (Fig. 14.9).





**Fig. 14.8** Forming limit diagram of the AISI 304L stainless steel constructed from several formability tests of Table 14.2. The enclosed photographs show several tests specimens at failure

**Fig. 14.9** **a** Diffuse necking and **b** localized necking in uniaxial tension. The angle  $\theta = \arctan(\sqrt{(R+1)}/R)$  of the localized neck depends upon the coefficient of anisotropy  $R$ , which is defined as the ratio of plastic strains  $\epsilon_2/\epsilon_1$  during the uniaxial tension test



Diffuse necking is accompanied by contraction strains (non-uniform deformation) in both the width  $w$  and thickness  $t$  directions of the specimens and leads to continuation of deformation under unstable conditions of failing load (or pressure). This type of failure occurs for tests providing strain paths in both first and second quadrants of the forming limit diagram. Localized necking is characterized by a progressive development of the neck at an angle  $\theta$  to the loading axis and by a substantial amount of straining after the onset of diffuse necking. This type of failure should not be confused with diffuse necking, which precedes it.

The general approach for determining the state of strain at the onset of diffuse necking requires calculating the maximum force. In case, for example, of biaxial symmetric loading conditions the applied forces are  $F_1 = F_2 > 0$ ,  $F_3 = 0$  (Table 14.3),

$$\begin{aligned} F_1 &= \sigma_1 A_1 \\ F_2 &= \sigma_2 A_2 \end{aligned} \quad (13)$$

The principal stresses  $\sigma_1 = \sigma_2$  for identical cross sectional areas  $A_1 = A_2$  at the current (or deformed) configuration and the plastic instability condition requires  $dF = 0$ ,

$$dF_1 = d\sigma_1 A_1 + \sigma_1 dA_1 = 0 \quad (14)$$

or,

$$\frac{d\sigma_1}{\sigma_1} = -\frac{dA_1}{A_1} = d\varepsilon_1 \quad (15)$$

Because the effective stress  $\bar{\sigma} = \sigma_1$  and the effective strain  $\bar{\varepsilon} = 2\varepsilon_1$  for biaxial symmetric loading conditions under the von Mises yield criterion, Eq. (15) can be rewritten as,

$$\frac{d\bar{\sigma}}{d\bar{\varepsilon}} = \frac{\bar{\sigma}}{2} \quad (16)$$

and the effective strain at the onset of diffuse necking for a material following a stress–strain relationship  $\sigma = K \varepsilon^n$  is given by,

$$\bar{\varepsilon} = 2n \quad (17)$$

The principal strains  $\varepsilon_1$ ,  $\varepsilon_2$ ,  $\varepsilon_3$  at the onset of diffuse necking are provided in Table 14.3 together with similar results obtained for uniaxial, plane strain and biaxial stretching conditions. The principal strains at the onset of localized necking in uniaxial tension tests that are included in Table 14.3 are determined under the assumption of plane strain conditions along the localized neck (Fig. 14.9b).

Figure 14.8 also allows comparing the experimental FLD with the theoretical FLD proposed by Swift [14] after combining the critical strains at localized and diffuse necking in case of materials with a stress–strain relationship  $\sigma = K \varepsilon^n$ ,

$$\begin{aligned} \text{Localized necking (second quadrant)} \quad \varepsilon_1 &= \frac{n}{1 + \rho} \\ \text{Diffuse necking (first quadrant)} \quad \varepsilon_1 &= \frac{2n (1 + \rho + \rho^2)}{(1 + \rho) (2\rho^2 - \rho + 2)} \end{aligned} \quad (18)$$

where  $\rho = \varepsilon_2/\varepsilon_1$  is assumed to be constant during loading.

To conclude this section on formability tests based on forming limit diagrams it is worth noticing that complex sheet metal forming encompass variations in the strain loading paths that may have a profound effect on the measured strains and,

consequently, on the formability limits by necking (FLC). On the contrary, the formability limits by fracture (FFL) are supposed to be material dependent and, therefore, insensitive to variations in the strain paths.

In the majority of the formability tests that are listed in Table 14.2 it is reasonable to assume that the ratio of the major to the minor strain is constant (or near constant) throughout the deformation and, therefore, the strain loading paths are said to be proportional (or linear). Somewhat non-linear strain paths that occur in some of the formability tests (e.g. hemispherical dome test with friction), are found to be slightly concave towards plane strain conditions (see also Fig. 14.8).

Additional influence on the formability limits may also arise from variations in the thickness of the blanks (producing changes in the FLC) and in the metallurgical conditions (producing changes in both the FLC and the FFL), among other variables.

## 14.4 Evaluating Formability by Simulative Tests

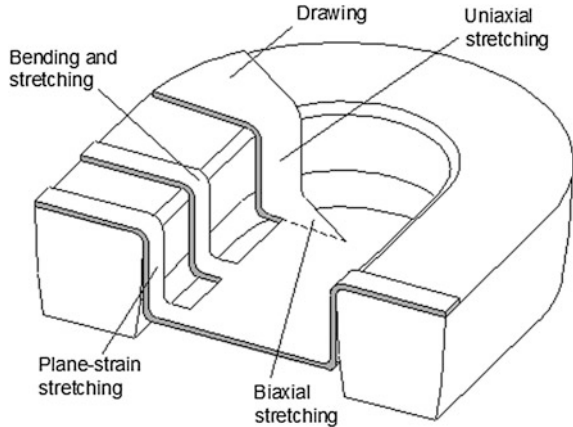
No matter how complex sheet metal forming operations may seem, they all consist on combinations or sequences of three basic deformation patterns; bending, stretching and drawing. Contrary to formability tests aimed at determining the failure limits by necking and fracture (FLC and FFL) that were comprehensively described in the previous section, simulative tests are designed with the purpose of replicating the states of strain and stress at different locations of the blanks during real sheet metal forming operations (Fig. 14.10).

Bending usually appears associated with stretching at the top and bottom of the walls, stretching occurs at the walls (uniaxial or plane-strain conditions) and at the planar bottom (bi-axial conditions) and drawing arises at the corners of the sheet metal parts.

In pure bending the outer fibers of the blank are in tension while the inner fibers are in compression. The critical parameter in bending is the ratio  $r/t$  of the bend radius  $r$  to the thickness  $t$  of the sheet. In fact, as  $r/t$  decreases, the tensile strain at the outer fibers increase and material may crack. Because in many sheet metal forming operations, the blanks are pulled over a radius (die  $r_{cd}$  and punch  $r_{cp}$  corner radii) it is important to evaluate the sheet metal formability by means of stretch-bending simulative tests with wedge-shaped punches, under plane strain conditions (Table 14.4). Formability is quantified by measuring the punch depth at the maximum load or by determining the minimum bend radius at which a crack appears at the outer surface of the blank. The stretch-bending test with wedge-shaped punches can also be utilized for analysing elastic recovery (springback) upon removal of the forming load.

In case of sheet metal forming operations that produce flanges to strengthen the edge of the holes, to improve its appearance or to provide additional support for joining other sheet or tubular metal parts it is recommended to characterize

**Fig. 14.10** Simplified scheme showing typical deformation patterns of sheet metal forming operations



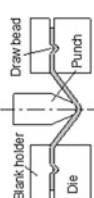
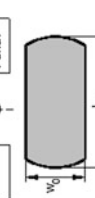
formability by means of hole-flanging simulative tests, under combined stretching and bending deformation patterns (Table 14.4).

In pure stretching, the blank is clamped at its outer periphery and in-plane tensile stresses cause the blank to change shape at the expense of sheet thickness. Ball punch simulative tests, such as the Erichsen [15] and Olsen [16] cup tests are commonly utilized to evaluate formability by measuring the height of the cup at the onset of failure. However, the Erichsen and Olsen cup tests are not pure stretch simulative tests because they allow some drawing-in and have low accuracy due to the small size of the tools and the impossibility of ensuring a sound fixing of the blanks. This justifies the utilization of the hemispherical dome test which is performed with larger diameter blanks and makes use of draw beads to avoid any drawing-in (Table 14.4). The formability is also quantified by measuring the height of the cup at the onset of failure. Gosh [17] proposed a variant of the hemispherical dome test performed with strip specimens of various widths and quantified formability by means of the limiting dome height (LDH) curve, a graphical representation of the heights of the height of the deformed blanks at the onset of failure as a function of the minimum in-plane strain  $\epsilon_2$ .

In case of sheet metal forming operations that require stretching a pre-cut hole in a blank it is recommended to utilize a hole-expansion simulative test (Table 14.4) that quantifies formability by means of ratio of the maximum diameter of the deformed hole to the initial diameter of the hole.

Pure drawing of a blank causes elongation in one direction by the tensile stresses produced by the punch and compression in the perpendicular direction as its diameter decreases. The most commonly used test for evaluating formability in deep drawing is the Swift cup test [18, 19] in which circular blanks of various diameters are deep drawn into cups by a flat-bottomed cylindrical punch (Table 14.4). Formability is quantified by the limiting draw ratio (LDR) defined as the ratio of the diameter of the largest blank that can be successfully drawn to the diameter of the punch (Fig. 14.11).

**Table 14.4** Simulative tests as a function of the metal forming deformation patterns

Deformation pattern	Test	Schematic drawing	Formability measure
Stretching and bending	Stretch-bending test (with wedge-shaped punches)		<ul style="list-style-type: none"> <li>• Punch depth at the maximum load (onset of failure)</li> <li>• Minimum bend radius at which a crack appears</li> <li>• Bend angle before and after unloading (elastic recovery)</li> </ul>
Hole-flanging test	Hole-flanging test		<ul style="list-style-type: none"> <li>• Limiting forming ratio defined as the ratio of the maximum diameter of the finished flange to the initial diameter of the hole <math>LFR = \frac{D_{max}}{D_0}</math></li> </ul>

(continued)

Table 14.4 (continued)

Deformation pattern	Test	Schematic drawing	Formability measure
Stretching	Hemispherical dome test		<ul style="list-style-type: none"> <li>• Punch depth at the maximum force (onset of failure)</li> </ul>
Limiting dome test (Drewes and Gosh variant to hemispherical dome test)			<ul style="list-style-type: none"> <li>• Limit dome height curve presenting the punch depth at the maximum force as a function of the in-plane strain <math>\epsilon_2</math> at the onset of failure</li> </ul>
Hole expansion test			<ul style="list-style-type: none"> <li>• Limiting forming ratio defined as the ratio of the maximum diameter of the deformed hole to the initial diameter of the hole <math>LFR = \frac{D_{max}}{D_1}</math></li> </ul>

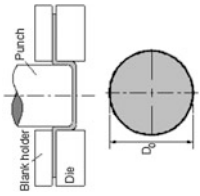
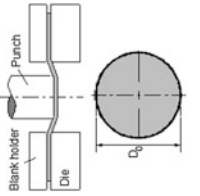
(continued)

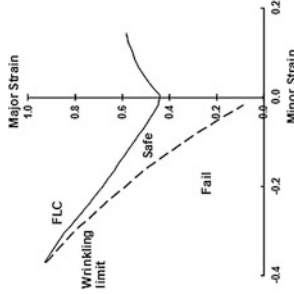
Table 14.4 (continued)

Deformation pattern	Test	Schematic drawing	Formability measure
Stretching and Drawing	Siebel-Pomp test (stretch-drawing variant of the hole expansion test)		<ul style="list-style-type: none"> <li>Limiting forming ratio defined as the ratio of the maximum diameter of the deformed hole to the initial diameter of the hole <math>LFR = \frac{D_{max}}{D_0}</math></li> </ul>
	Conical cup test (also known as the Fukui test)		<ul style="list-style-type: none"> <li>Limiting forming ratio defined as the ratio of the diameter of the base of the conical cup at the onset of failure to the diameter of the original blank</li> </ul>

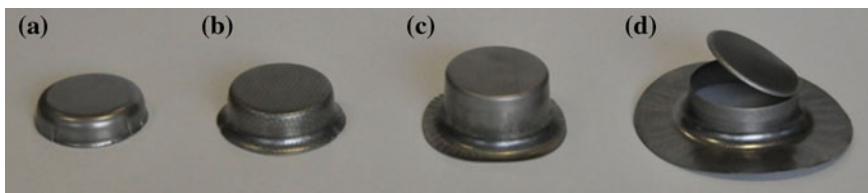
(continued)

**Table 14.4** (continued)

Deformation pattern	Test	Schematic drawing	Formability measure
Cylindrical cup test (also known as the Swift test)			<ul style="list-style-type: none"> <li>Limiting draw ratio defined as the ratio of the diameter of the largest blank that can be successfully drawn to the diameter of the punch <math>LDR = \frac{D_{max}}{D_p}</math></li> </ul>
Wrinkling test (variant to cylindrical cup test that uses a punch with a diameter much smaller than the die opening)			<ul style="list-style-type: none"> <li>The critical wrinkling strains plotted on the forming limit diagram</li> </ul>

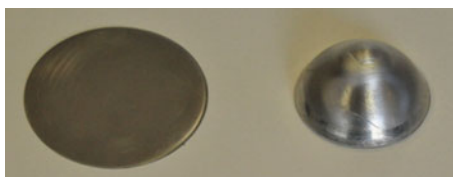






**Fig. 14.11** Experimental Swift cups obtained for AISI 304L stainless steel for several diameter ratios  $D_0/D_p$  **a**  $D_0/D_p = 1.42$  **b**  $D_0/D_p = 1.62$  **c**  $D_0/D_p = 1.98$  **d**  $D_0/D_p = 2.3$

**Fig. 14.12** Experimental Fukui specimen before and after the test from aluminium AA1050-H111



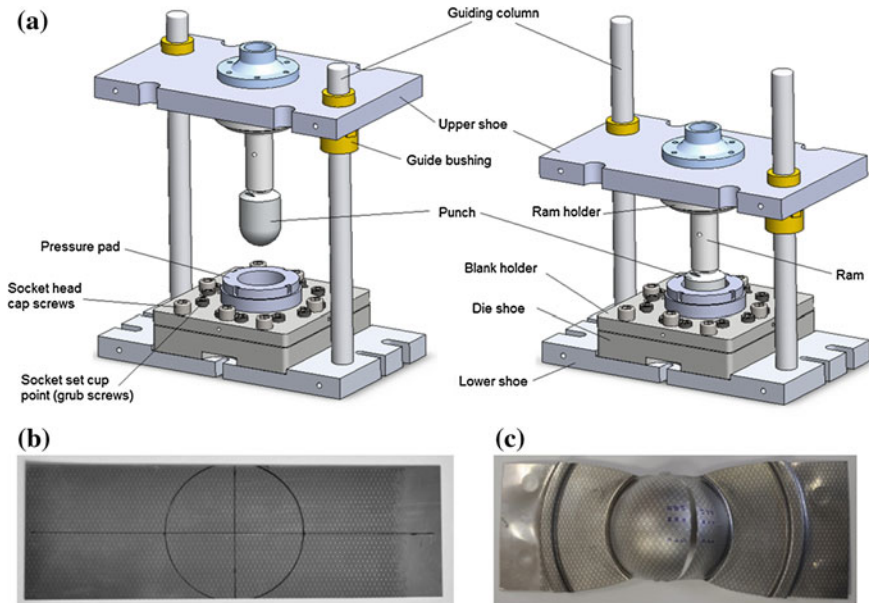
Because in drawing operations the length of the unsupported wall is significant and can lead to wrinkling it is sometimes convenient to perform a simulative wrinkling test (Table 14.4). The simulative wrinkling test is a variant of the Swift cup test that uses a punch with a diameter much smaller than the die opening with the objective of artificially stimulating an unsupported wall where wrinkles may be easily triggered. The test allows determining the wrinkling failure curve in the forming limit diagram.

In case of metal forming operations involving stretching and drawing it is recommended to perform the Fukui conical cup test [20] (Table 14.4 and Fig. 14.12). In this test formability is quantified by the ratio of the diameter of the base of the conical cup at the onset of failure to the diameter of the original blank.

The aforementioned tests are classified as a function of the deformation pattern and are comprehensively listed in Table 14.4.

## 14.5 Basic Laboratory Testing Equipment

This section starts by describing a flexible, versatile and robust laboratory tool system that was developed by the authors to support education and basic research activities in sheet metal forming and finishes with a brief overview of other equipment that are also necessary to allow students to engage in a wide range of hands-on experimental tests.



**Fig. 14.13** a Schematic representation of the flexible laboratory tool system with a punch-die set for performing the hemispherical dome test (*open* and *closed* positions) and photographs of the **b** initial and **c** deformed test blank

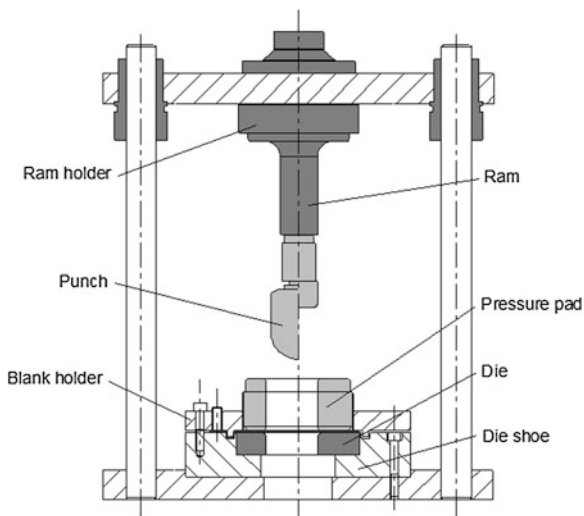
### 14.5.1 Flexible Laboratory Tool System

The flexible laboratory tool system (hereafter named ‘tool system’ or simply ‘tool’) must be preferentially installed in a hydraulic universal testing machine or, if not available, in a low capacity hydraulic press. In case of installing the tool system in a hydraulic press it is necessary to account for the need to integrate load cells and displacement transducers to acquire the experimental evolution of the punch load with displacement for subsequent treatment and analysis.

In what follows, authors consider the tool system to be installed in a hydraulic universal testing machine and for the purpose of presentation, and although not corresponding to what readers directly observe in Fig. 14.13, the tool will be split into the following three different groups of components; structural, process and specific active tool parts.

Structural parts comprise components such as the guiding columns, the ram holder and the upper and lower shoes that are independent of the formability tests (based on forming limit diagrams or based on simulating operations) to be performed. The overall length of the guiding columns and ram were chosen so that insertion of blanks and extraction of deformed blanks can be easily performed for all test cases.

**Fig. 14.14** Two dimensional drawing of the tool system shown in Fig. 14.13 at the beginning and end of stroke for the hemispherical dome test (*left*) and the cylindrical (Swift) cup test (*right*)



The die shoe, the ram, the pressure pad and the blank holder are typical process parts whereas the punches and the dies are specific active tool parts that are dependent on the geometry of the test to be performed (see also Fig. 14.14). The die shoe carries the different die rings utilized in the tests and is fixed to the lower shoe. Alternative die shoes may be designed for special purpose specific sheet metal forming tests. The ram carries the various punches that are utilized in the tests and is attached to the upper shoe by means of the ram holder.

The blank holder is built upon different components that increase effectiveness and versatility to its main purpose of preventing wrinkling. Removable draw beads with rectangular cross section can be attached to the blank holder in order to control or restrict (e.g. Hecker's variant to Nakazima tests) the flow of the blank into the die opening. The draw beads are placed away from the die ring and the grooves are machined in the top of the die shoe, close to the outside diameter of the die ring. Moving the draw beads away from the die ring adds flexibility to the tool system because different tests can be performed with a single die. The location of the draw beads also reduces the possibility of the die cracking at the edges. A screw-loaded pressure pad (with 1 mm pitch) enables different pressures to be applied on the blank placed on top of the die ring, thereby controlling wrinkling and drawing-in. 12 M20 cup point grub screws placed outside the diameter of the draw beads are available to extreme testing conditions where the blanks need to be hold-down without any drawing-in.

The punches and die rings are the active tool components. Each punch-die set is characterized by its profile, corner radius and clearance between the punch and die opening radius. Table 14.5 provides details of three different punch-die sets that allow performing various types of sheet metal formability tests. The main advantage of the proposed tool system is the flexibility to increase the number of

**Table 14.5** Details of the various laboratory formability tests that can be performed with four basic punch-die sets

Punch and die set	Drawing	Different tests	Geometry of the blanks
Punch with hemispherical profile		<p>Nakazima test</p> <p>Hecker's variation of Nakazima test</p> <p>Hemispherical dome test</p> <p>Limiting dome test</p> <p>Hemispherical cup test (similar to the normalized Erichsen/Olson ball punch tests)</p>	

(continued)

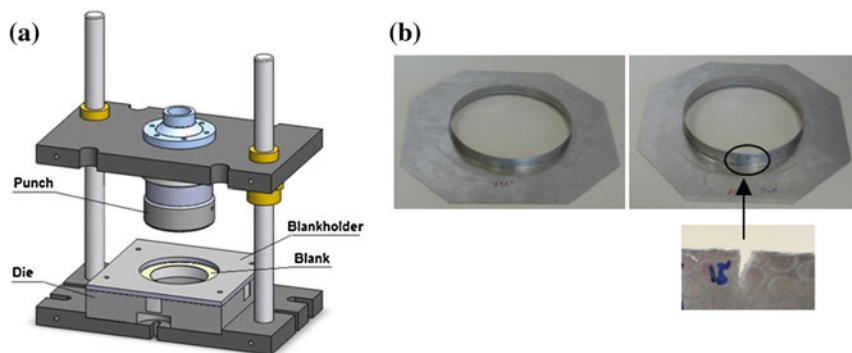
**Table 14.5** (continued)

Punch and die set	Drawing	Different tests	Geometry of the blanks
Punch with prismatic profile		Stretch-bending test	
Die with conical profile		Conical cup test (Fukui cup test)	

(continued)

**Table 14.5** (continued)

Punch and die set	Drawing	Different tests	Geometry of the blanks								
Punch with cylindrical profile		Marciniak test									
	<table border="1"> <thead> <tr> <th>Reference</th> <th>R<sub>90</sub></th> </tr> </thead> <tbody> <tr> <td>PS 5/0.2</td> <td>2</td> </tr> <tr> <td>PS 5/0.5</td> <td>5</td> </tr> <tr> <td>PS 5/0.8</td> <td>8</td> </tr> </tbody> </table>	Reference	R <sub>90</sub>	PS 5/0.2	2	PS 5/0.5	5	PS 5/0.8	8	Cylindrical cup test (Swift cup test)	
Reference	R <sub>90</sub>										
PS 5/0.2	2										
PS 5/0.5	5										
PS 5/0.8	8										
		Hole expansion test									
		Siebel-Pomp test									
		Hole flanging test									
		Wrinkling test									



**Fig. 14.15** **a** Schematic representation of a modified set-up of the laboratory tool system that allows performing hole flanging tests in blanks with 250 mm diameter and **b** Successful and unsuccessful hole-flanges produced in blanks of aluminium AA1050-H111

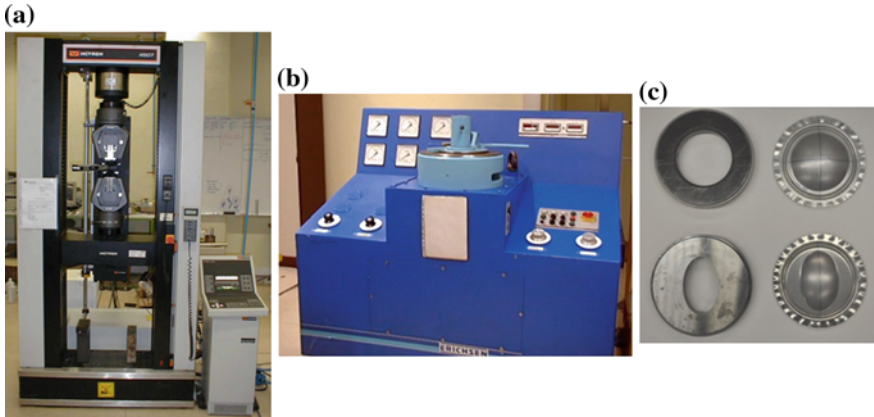
punch-die sets to perform other tests (e.g. rectangular cup tests not addressed in the presentation) and the possibility of easily replacing the ram and the die shoe to allow the utilization of punch and die sets with larger geometries (Fig. 14.15).

### 14.5.2 Other Equipment

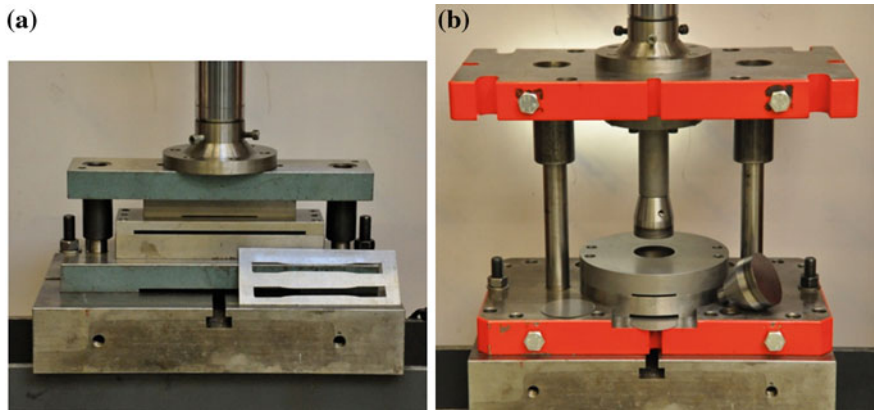
There are other equipment that are necessary to fulfil the aims and objective of engaging sheet metal forming students in a wide range of hands-on experimental tests. Besides the necessity of having a universal testing machine for performing the uniaxial tension tests it is always good, although not strictly necessary, to have an electro-hydraulic universal sheet metal testing machine to perform biaxial circular or elliptical bulge tests, among others. Figure 14.16 presents photographs of two of these types of machines and details of the circular and elliptical dies that are commonly utilized in bulge tests.

The blanks utilized in the laboratory tests are usually cut out of the metal sheets and stacked to form a shape with an appropriate aspect ratio to be precisely milled to the desired geometry in a conventional or CNC milling machine. The procedure requires no investment in equipment, because it normally makes use of already existing machines, but it is time consuming and requires specialists for operating the milling machines.

An alternative is to design and fabricate a blanking tool that allows installing different punch-die sets for cutting out the blanks from the metal sheets already with the desired geometry. However, because the blanking operation damages the edge of the specimens (e.g. produces burrs, cracks and changes the stress response due to cold strain hardening caused by shearing) it is necessary, in case of the specimens for uniaxial tension tests or in case of the blanks for hole-expansion or hole-flanging tests, to perform a subsequent grinding operation with sand paper to



**Fig. 14.16** a Universal testing machine, b electro-hydraulic universal sheet metal testing machine and c photographs of *circular* and *elliptical* dies and specimens utilized in hydraulic bulge tests

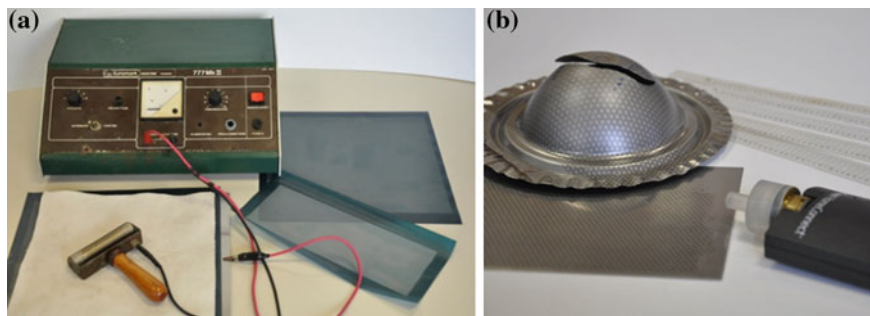


**Fig. 14.17** a Special purpose blanking tool for producing specimens for uniaxial tension tests and b flexible blanking tool for producing *circular blanks* for different sheet metal forming tests

smooth and make the edges perpendicular to the plane of the sheet and to eliminate existing burrs and cracks. The overall procedure is able to produce specimens and blanks with milled compliant edges. Figure 14.17 presents photographs of two different blanking tools that are currently utilized to produce specimens for uniaxial tension tests and blanks for sheet metal forming tests.

To conclude, it is important to make reference to the equipment for electroetching circular (or rectangular) grids on the surfaces of the specimens and blanks and for measuring the deformed circles. Although there are nowadays expensive and sophisticated equipment for performing both tasks (encompassing the utilization of strain sensible paints and laser based measuring systems), Fig. 14.18 only shows classical solutions that are adequate for education purposes.





**Fig. 14.18** **a** Electro-chemical metal marking equipment for imprinting *circular* or *rectangular grids* on the surface of the blanks with detail of marking stencils and **b** flexible rulers and optical PC based strain measuring system

## 14.6 Numerical Simulation of Laboratory Tests

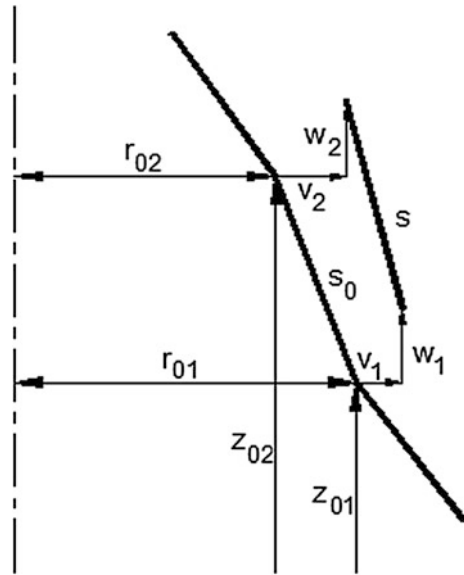
Tekkaya and Martins [21] recently provide industrial, education and academic users of computer programs a basic overview of finite elements in metal forming that will enable them to recognize the pitfalls of the existing formulations, identify the possible sources of errors and understand the routes for validating their numerical results. The overview draws from the fundamentals of the finite elements, plasticity and material science to aspects of computer implementation, modelling, accuracy, reliability and validation. The presentation is illustrated and enriched with selected examples obtained from research and industrial metal forming applications.

Under these circumstances, the aim and objective of this section is much less ambitious and entirely focused in presenting the fundamentals of a simple finite element computer program (SLabs) that gives support to some of the sheet metal forming laboratory tests that were described in previous sections. The presentation is supported by an example and will be carried out without resorting to complicated theoretical and numerical topics that may go beyond what is often lectured in basic metal forming courses at undergraduate and graduate levels.

### *14.6.1 Fundamentals of the Rigid Plastic Finite Element Formulation*

The rigid plastic functional that gives support to the finite element computer program SLabs for two-dimensional (axisymmetric or plane strain) forming of thin sheets is given by (body forces are neglected) [22],

**Fig. 14.19** Discretization of a blank by linear elements



$$\Pi = \int_V \bar{\sigma} d\bar{E} dV + \int_V \frac{1}{2} h d\bar{E}^2 dV - \int_S (T + dT) u dS \quad (19)$$

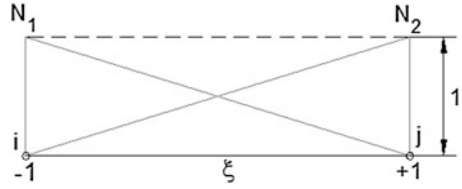
where  $\bar{\sigma}$  is the effective stress,  $d\bar{E}$  is the increment of the effective Green–Lagrange strain,  $h = \partial\bar{\sigma}/\partial\bar{\epsilon}$  is the slope of the stress–strain curve,  $T$  is the surface traction applied on the boundary,  $u$  is the increment of displacement, and  $V$  and  $S$  are the volume and surface of the reference configuration, respectively. The first term in Eq. (19) corresponds to the volumetric strain energy of deformation, the second term represents the energy involved with strain hardening and the third term represents the work of the external surface tractions.

Considering blanks to be made of (or discretized by) a series of linear elements (Fig. 14.19) and assuming that circumferential and meridian directions are the principle in-plane directions (and thickness direction is the third principal direction), the Green–Lagrange strain may be approximated, for simplicity, by the logarithmic strain during a small time increment (refer also to Fig. 14.3),

$$d\mathbf{E} = \begin{Bmatrix} dE_1 \\ dE_2 \end{Bmatrix} = \begin{Bmatrix} dE_s \\ dE_\theta \end{Bmatrix} = \begin{Bmatrix} \ln(s/s_0) \\ \ln(r/r_0) \end{Bmatrix} \quad (20)$$

where  $s_0$  (undeformed length),  $s$  (deformed length),  $r_0$  (original radial location) and  $r$  (deformed radial location) are given by,

**Fig. 14.20** Shape functions utilized in the interpolation of the incremental displacement field inside a linear element connecting nodes ‘i’ and ‘j’



$$\begin{aligned}
 s_0 &= \sqrt{(r_{02} - r_{01})^2 + (z_{02} - z_{01})^2} \\
 s &= \sqrt{(r_{02} - r_{01} + v_2 - v_1)^2 + (z_{02} - z_{01} + w_2 - w_1)^2} \\
 r_0 &= \frac{1}{2} (r_{02} + r_{01}) \quad r = r_0 + \frac{1}{2} (v_2 + v_1)
 \end{aligned} \tag{21}$$

The symbols  $v$  and  $w$  denote the increments of displacement  $u$  in the  $r$  and  $z$  directions.

The values of the effective stress and of the effective strain increment may be written as a function of their circumferential and meridian components as follows,

$$\begin{aligned}
 \bar{\sigma} &= \sqrt{\sigma_s^2 - \frac{2R}{1+R} \sigma_s \sigma_\theta + \sigma_\theta^2} \\
 d\bar{E} &= \frac{1+R}{\sqrt{1+2R}} \sqrt{dE_s^2 + \frac{2R}{1+R} dE_s dE_\theta + dE_\theta^2} = \sqrt{\frac{2}{3}} [d\mathbf{E}^T D d\mathbf{E}]^{1/2}
 \end{aligned} \tag{22}$$

where  $R$  is the coefficient of anisotropy (planar), which was previously defined as the ratio of the logarithmic width strain to the logarithmic thickness strain in uniaxial tension, and  $D$  is given by,

$$D = \frac{3(1+R)}{2(1+2R)} \begin{bmatrix} 1+R & R \\ R & 1+R \end{bmatrix} \tag{23}$$

By replacing (22) and (23) into the functional (19) and following the aforementioned discretization procedure, the contribution of each individual finite element ‘m’ is obtained from,

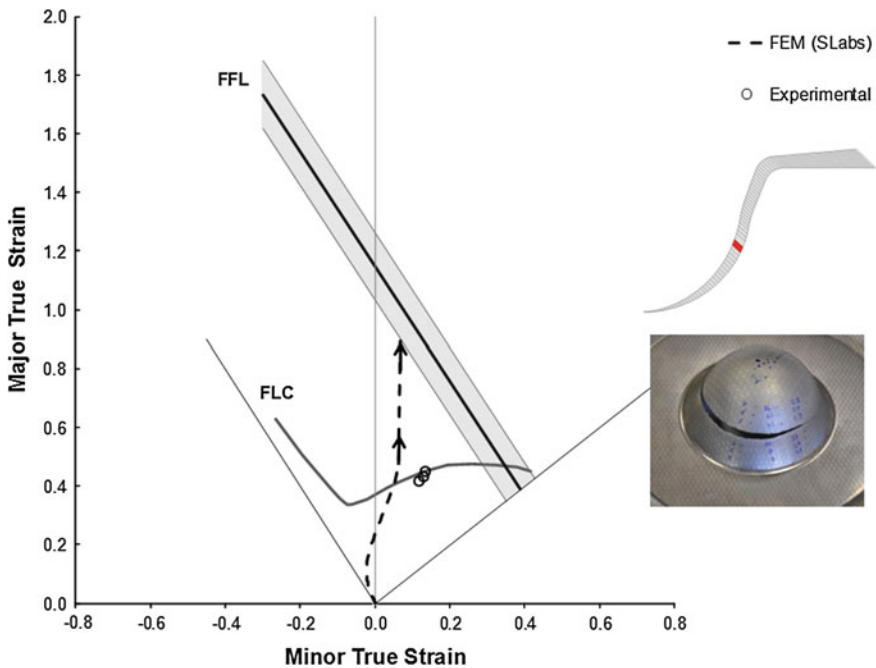
$$\Pi^m = \int_V \bar{\sigma} \sqrt{\frac{2}{3}} [d\mathbf{E}^T D d\mathbf{E}]^{1/2} dV + \frac{1}{2} \int_V h \left( \frac{2}{3} \right) [d\mathbf{E}^T D d\mathbf{E}] dV - \int_S \mathbf{T}^T N \mathbf{u} dS \tag{24}$$

where  $\mathbf{u} = \{v, w\}$  is the incremental displacement vector at elemental level and  $N$  are the shape functions of the linear element (Fig. 14.20),

$$\mathbf{u} = \begin{Bmatrix} v \\ w \end{Bmatrix} = \begin{bmatrix} \frac{1+\xi}{2} & 0 & \frac{1+\xi}{2} & 0 \\ 0 & \frac{1+\xi}{2} & 0 & \frac{1+\xi}{2} \end{bmatrix} \begin{Bmatrix} v_1 \\ w_1 \\ v_2 \\ w_2 \end{Bmatrix} \tag{25}$$

**Table 14.6** The main experimental parameters

Material	Stainless steel AISI 304L
Stress–strain curve	$\sigma = 1307\epsilon^{0.39}$ MPa
Anisotropy coefficient	0.88
Diameter of the blank	175 mm
Thickness of the blank	0.5 mm
Radius of the punch	37.5 mm
Opening diameter of the die	84.75 mm
Corner radius of the die	7 mm



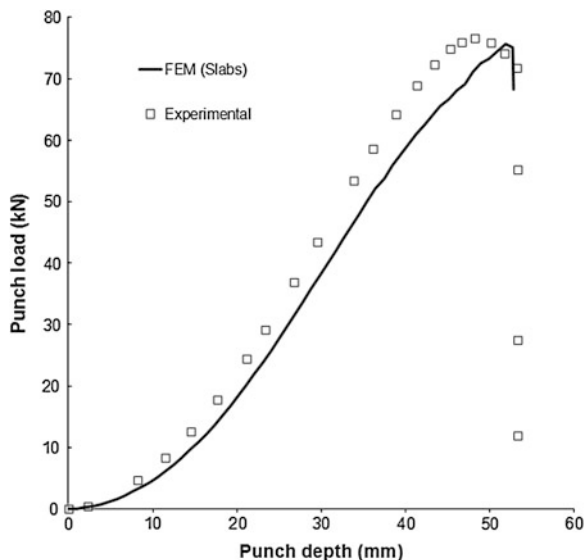
**Fig. 14.21** Experimental and finite element predicted strain loading paths for the hemispherical dome test with friction (AISI 304L)

Summing up the contributions of each element  $\Pi \cong \sum \Pi^m$  by assembling all the elements in the finite element scheme, the following nonlinear system of equations is obtained,

$$\frac{\partial \Pi}{\partial \mathbf{u}} = \sum \frac{\partial \Pi^m}{\partial \mathbf{u}^m} = 0 \tag{26}$$

from where the increments of displacements of each nodal point  $\mathbf{u}_i^m = \{v_i, w_i\}$  are obtained. Further details in the computer implementation of the rigid plastic formulation applied to sheet metal forming is provided elsewhere [22, 23].

**Fig. 14.22** Experimental and finite element predicted evolution of the punch load with displacement for the hemispherical dome test with friction (AISI 304L)



### 14.6.2 Application Example

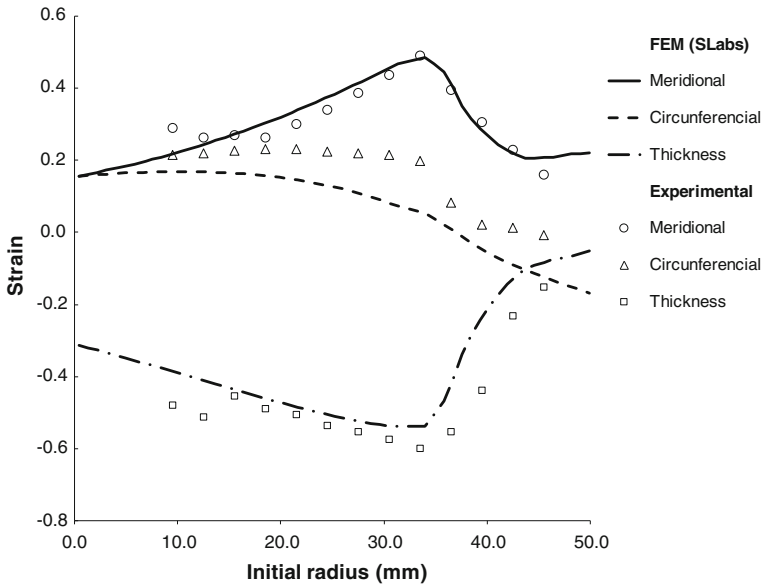
The selected application example demonstrates the utilization of the finite element computer program SLabs in conjunction with a hemispherical dome test with friction performed in AISI 304L blanks (Table 14.6).

The blank was discretized by means of 100 linear elements and several values of the coefficient of friction  $\mu$  were utilized in the simulations because the exact value of the friction coefficient was not known. Figure 14.21 shows the experimental and finite element predicted strain loading paths for the hemispherical dome test with friction performed in AISI 304L blanks in the forming limit diagram. The failure limits by necking (FLC) and fracture (FFL) were obtained by means of the experimental procedures that were previously addressed in Sect. 3 ('Evaluating Formability by Tests Based on Forming Limit Diagrams').

As seen in the Fig. 14.21, the agreement between experimental and finite element predicted strain paths is good for a coefficient of friction  $\mu = 0.23$ . Results also show numerical evidence of the experimental kink in the strain path, after neck initiation, towards the vertical direction.

Figure 14.22 shows a comparison between the experimental and the finite element predicted evolution of the punch load with displacement. The agreement is good and the drop in the punch load due to instability is well predicted by finite elements.

However, friction affects the distribution of thickness in the deformed blank namely, the location of maximum thinning which moves from the pole (leftmost position in Fig. 14.23) to the region where the blank detaches from the punch.



**Fig. 14.23** Experimental and finite element predicted evolution of thickness for the hemispherical dome test with friction (AISI 304L)

The application example shows the advantage of combining a simple finite element computer program such as SLabs with experimental laboratory work to help students easily understanding the influence of the main operating conditions (material, thickness and friction, among others) in the overall response of the sheet metal forming tests. The program may help students checking if their experimental data is compatible with what is expected from each type of sheet metal forming test. In this sense, finite element computer programs may also be regarded as good complements to education through hands-on laboratory tests.

## 14.7 Conclusions

This chapter presented the theoretical and experimental background for developing laboratory training activities in the modern education of sheet metal forming at graduate and post-graduate levels. A new flexible laboratory tool system developed by the authors, which is capable of performing the majority of the formability tests that are utilized for setting-up the failure limits by necking and fracture and for reproducing the deformation patterns that are commonly found in complex sheet metal forming operations, was comprehensively described.

The integration of finite element modeling with laboratory training activities was also addressed with the purpose of demonstrating how a simple, in-house, finite element computer program can help students in assessing the accuracy, reliability and validity of their experimental data. The program can also be utilized

as a student tool box for facilitating the study of sheet metal forming namely, for better understanding the influence of the major operating conditions that were investigated during the laboratory training activities.

**Acknowledgments** Luciana Montanari would like to acknowledge FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) for the financial support of the research work. The work of André Teodora and Pedro Pardal is also greatly acknowledged.

## References

1. Keeler SP (1968) Circular grid system—a valuable aid for evaluating sheet metal formability. SAE Technical Paper 680092
2. Goodwin G (1968) Application of strain analysis to sheet metal forming problems in the press shop. SAE Technical Paper 680093
3. Rossard C (1976) Mise en forme de métaux et alliages. CNRS, Paris
4. Silva MB, Nielsen PS, Bay N, Martins PAF (2011) Failure mechanisms in single-point incremental forming of metals. *Int J Adv Manuf Technol* 56:893–903
5. Atkins AG (1997) Fracture mechanics and metal forming: damage mechanics and the local approach of yesterday and today. In: Rossmann HP (ed) *Fracture research in retrospect*. Balkema, Rotterdam, pp 327–350
6. Atkins AG (1996) Fracture in forming. *J Mater Process Technol* 56:609–618
7. Hsu TC (1974) Present scope and future trend of sheet metal forming research. *Int J Prod Res* 12:99–115
8. Danckert J, Wanheim T (1979) The use of a square grid as an alternative to a circular grid in the determination of strains. *J Mech Working Technol* 3:5–15
9. Sowerby R, Duncan JL, Chu E (1986) The modeling of sheet metal stampings. *Int J Mech Sci* 28:415–430
10. Nakazima K, Kikuma T, Hasuka K (1968) Study on the formability of sheet steels. Technical Report No 264, Yawata Iron and Steel Company, pp 141–154
11. Hecker SS (1974) A cup test for assessing stretchability. *Met Eng Q* 14:30–36
12. Hasek V (1973) On the strain and stress states in drawing of large unregular sheet metal components (in German). *Berichte aus dem Institute für Umformtechnik, Universität Stuttgart*, No 25. Girardet, Essen
13. Marciniak Z (1971) Limits of sheet metal formability (in Polish). WNT, Warsaw
14. Swift HW (1952) Plastic instability under plane stress. *J Mech Phys Solids* 1:1–18
15. Erichsen AM (1914) A new test for thin sheets (in German). *Stahl und Eisen* 34:879–882
16. Olsen TY (1920) Machines for ductility testing. *Proceedings of the American society of materials*, vol 20. pp 398–403
17. Gosh AK (1975) The effect of lateral drawing-in on stretch formability. *Met Eng Q* 15:53–64
18. Chung SY, Swift HW (1951) Cup-drawing from a blank. *Proceedings of the institution of mechanical engineers*, vol 165. pp 199–223
19. Swift HW (1954) The mechanism of a simple deep-drawing operation. *Sheet Met Ind* 31:817–828
20. Fukui S (1939) Researches on the deep-drawing process. *Scientific papers of the institute of physical and chemical research*, pp 1422–1527
21. Tekkaya AE, Martins PAF (2009) Accuracy, reliability and validity of finite element analysis in metal forming: a user's perspective. *Eng Comput* 26:1026–1055
22. Kobayashi S, Oh SI, Altan T (1989) *Metal forming and the finite element method*. Oxford University Press, New York
23. Martins PAF, Barata Marques MJM (1993) Plane strain rigid plastic finite element formulation for sheet metal forming processes. *J Eng Manuf* 207:167–171

# Chapter 15

## Multimedia Resources in Engineering Education

**B. F. Yousif, Marita Basson and Carola Hobohm**

**Abstract** In this study, the impact of employing multimedia resources in the teaching of a common first year engineering course, Engineering Materials, was investigated. The aim was to determine if there was an improvement in students' reported understanding and evidence of increased student engagement in the online discussion forums as a result of using the resources. Threshold concepts and the related notion of troublesome knowledge has resurfaced in research as well as in current thinking about learning and teaching in higher education. There is no clear indication in engineering education literature on how multimedia resources could influence student understanding of threshold and complex concepts. The multimedia resources used in the course were developed as part of a framework of resources intended to guide students through identified threshold concepts. Animation software was used, supplemented by video recordings and inked annotations of each lecture. An online survey was conducted over three semesters utilizing an inbuilt facility in the online learning environment. The initial results showed that implementing the multimedia resources significantly improved the student engagement in the online discussion forums? The perceived quality of teaching and students' understanding of complex topics also improved as was revealed by results and students report. Seventy three per cent of the students stated their preference toward teaching methods using multimedia resources and inking. This preference was more pronounced in the off campus or online student cohort.

---

B. F. Yousif (✉) · M. Basson · C. Hobohm  
Faculty of Engineering and Surveying, University of Southern Queensland,  
Toowoomba, QLD 4350, Australia  
e-mail: Belal.Yousif@usq.edu.au



## 15.1 Introduction

The first aim of this longitudinal study was to determine if there was an improvement in student understanding of the concepts presented in an engineering education context, namely in an Engineering Materials course. The Engineering Materials course is a foundation course for engineering students in the first year of study. The course provided the ideal opportunity for this study due to the large class size when given the typical low response rate to surveys, and the presence of certain traditionally complex concepts. These threshold and problematic Engineering Materials concepts were presented in a visually enhanced or multimedia format.

*Multimedia learning* can be defined as learning from words and pictures and *multimedia instruction* as using words and pictures that are intended to foster learning, [1]. Words can be narrated or printed whilst pictures can be static or animated. In [1], Mayer and Moreno warn that their research has repeatedly revealed cognitive overload of learners when multimedia instruction was used.

The following part of this section shall firstly deal with threshold concepts and troublesome knowledge, and subsequently with the use of multimedia enrichment of course content to elicit higher order thinking, as that is the level of thinking that the authors deem is required to comprehend problematic concepts. This section will conclude with a discussion of the impact of multimedia enhanced presentation of threshold concepts and troublesome knowledge on student engagement.

### 15.1.1 Threshold Concepts and Troublesome Knowledge

It is argued in the literature that first year engineering courses provide the foundation for engineering students in terms of critical knowledge and skills that would benefit them in their later courses, [2–6]. The term *threshold concepts* derives from UK education theory and it describes those concepts (critical knowledge) that are essential to knowledge and understanding within specific disciplines, in this case engineering, [7]. In [8], Meyer and Land hold the visual metaphor of a threshold concept represented by a portal as valuable, as it explains how the portal comes into view, how it is approached, negotiated and transitioned. Critical knowledge or knowledge of threshold concepts enables students to more successfully engage with complex knowledge or troublesome concepts, thus progressing to higher levels of learning, [7]. A focus on threshold concepts enables educators to focus on concepts that are fundamental to their subjects or courses as opposed to bulking out their course content, [9].

First year engineering courses provide an opportunity to present threshold concepts in such a way that students can successfully master those concepts before moving on to more advanced material. In order to do that, the educator needs to determine the threshold concepts that should be embedded in his/her course content. In [8], Meyer and Land present five characteristics of threshold concepts:

1. Grasping a threshold concept is transformative because it involves ontological and conceptual shifts.
2. Grasping a threshold concept is irreversible and the learner is unlikely to forget it.
3. A threshold concept is integrative as it exposes the obscured interrelatedness of a phenomenon.
4. Threshold concepts are bounded in a specific discipline, it bind a subject together.
5. A threshold concept is likely to contain some *troublesome knowledge*.

Troublesome knowledge is defined as knowledge that appears counter-intuitive, alien or incoherent [9], or beyond understanding. This ‘trouble-someness’ can prevent mastery of threshold concepts as students’ intuitive understanding of such a concept can prevail [7]. Troublesome knowledge, although obviously important and maybe even exciting, can seem very alien and it might be difficult to get students to think about it, [8]. Threshold concepts are inherently problematic as it requires the learner to integrate ideas and accept the transformation of their own thinking and in their own learner-identity, [10].

An additional concern mentioned in the literature is that certain concepts within disciplinary fields seem to be more troublesome to students than others, [10]. Data collected from four international universities (University Southern Queensland, Nottingham University, Multimedia University, Al-Anbar University), confirmed that a high percentage of engineering students face problems in understanding certain topics in the Engineering Materials course, for example *atomic bonds*, the *binary phase diagram*, and the *microstructure of materials*. These concepts as well as other were presented using a multimedia approach whilst being careful to avoid cognitive overload by, for example, using voice-over techniques rather than transcripts or subtitles when presenting visual content, [1].

A method suggested by Land et al. [10] to deal with threshold concepts is seeing the process of the students’ learning as a journey or excursion in which the students would encounter threshold concepts as part of a framework designed to assist them to cope with these concepts. Threshold concepts thus have the potential to be powerful, transformative points and are viewed by some as the ‘jewels in the curriculum’, [10]. They are the crucial points in the curriculum that the educator can use to construct opportunities for the learners to gain crucial conceptual understanding, [10]. The authors of this chapter argue that technologically enhanced ways of introducing threshold concepts and troublesome knowledge through the use of multimedia resources and visualization tools (animations, videos, photos, inking), could enhance students’ conceptual understanding and offer an opportunity to cross the threshold to a new learner identity. The authors thus hold that higher order thinking skills and comprehension of complex knowledge, can be elicited through the inclusion and use of technologically enhanced material to present threshold concepts in first year engineering courses. Establishing the ability for higher levels of cognition in their first year, enables students to progress through their programs with greater ease.

### ***15.1.2 Technology-Enriched Environments and Higher Order Thinking***

Innovations in technology, the user-friendly nature and the low cost of multimedia production provide educators with unprecedented freedom to explore suitable alternatives to traditional education and tools like inking, recording lecture. The development of problem solving, critical thinking, and higher-order thinking skills are crucial to future success in a technology age, [11]. Educators therefore need to endeavour to regularly provide experiences that engage students at higher levels in Bloom's Taxonomy: the analysis, synthesis and evaluation levels, [11]. Moving to higher levels of learning can be an immediate process or it can take place over longer periods of time [7]. The transformative nature of the process can present challenges to students [7]. Higher order thinking has been defined in the literature as taking responsibility and control of our own mind [12] and at that level of description ties in with the process that learners might go through when mastering threshold concepts.

There is no clear indication in engineering education literature on how multimedia resources could influence student understanding of threshold and complex concepts [13, 14]. The research reported on in this chapter. The authors attempt to partly address that gap by reporting on a longitudinal study which evaluates the influence and value of embedded multimedia resources on reported student understanding of typically complex concepts in the Engineering Materials course.

Fostering higher order thinking in first year engineering students can highly enrich their learning experience. If students have the ability to think at higher levels, the lecturer is able to present more challenging material. Hence, comprehending more complex material becomes possible, because higher order thinking leads to greater understanding of critical and complex engineering problems, [2–4]. The question can thus be asked what educators in these situations can do to help students overcome the barriers to learning such as threshold knowledge and troublesome concepts, [10].

Introducing an idea in visual form has been found to be an important tool in the teaching and understanding of threshold concepts. Many authors have pointed out the importance of students' visualization in understanding complex topics [15, 16]. Different methods have been used over the years with different age groups in order to enhance students' imagination or visualization of problematic or complex problems. A more recent educational tool is the use of cartoons as visual tools that combine humour, exaggeration and symbols, thereby addressing various elements which have been proven to improve learning. Cartoons can be used effectively in the teaching process as they provide information with regards to concretely instructive objects, [3]. Using cartoons has been found to be a successful technique to transmit information to pre and primary school students in order to improve their higher order thinking and comprehension of critical topics, such as micro-structure of materials, failure mode in materials and formation of alloys, [2–6]. In engineering tertiary education, it is less appropriate to introduce techniques like

cartoons, especially in some engineering courses such as Materials, Statics, Dynamics, Design and Heat Transfer that typically have high incidences of complex concepts or troublesome knowledge. Real life applications are always preferred for engineering students to enhance their visualization and understanding of theories.

The analyses of the content of online discussions may reveal evidence of higher order thinking, [17]. Very few studies have evaluated the usefulness of online forums for eliciting higher order thinking [17] and the nature of the course, the type of task, and the wording of the initial prompt are all potential influences on the attainment of higher order thinking in an online environment.

### ***15.1.3 Student Engagement***

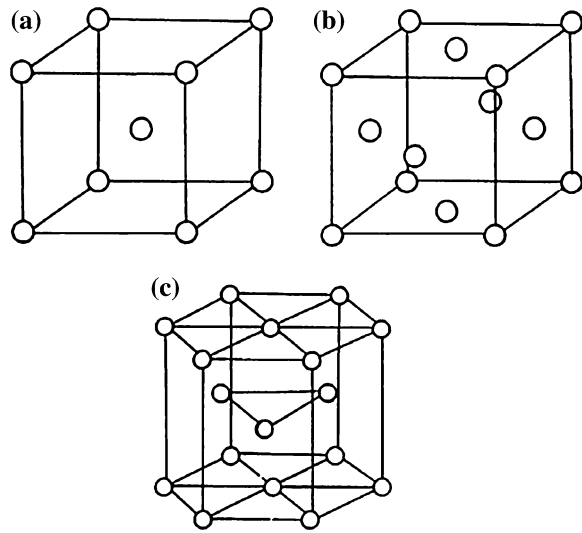
Many studies point to the need for active student engagement with the conceptual material and manipulation thereof [10]. The question is therefore what the framework of engagement and the forms of engagement would be that will lead to transformative learning [10]. In [9], Cousins emphasises the value of curriculum design that makes provision for the mastery of threshold concepts. One of the crucial aspects mentioned, is that of *listening* for understanding on the part of the educator. The educator should hear what students' misunderstandings and uncertainties are. This is in line with the aims of this study, which are to determine if there was an improvement in students' *reported* understanding and *evidence* of students' engagement as a result of using embedded multimedia resources. Looking for evidence of higher student engagement is deemed to be another form of 'listening' to what the students are reporting "Threshold concepts lead not only to transfigured thoughts but to a transfiguration of identity and adoption of extended or elaborated discourse" ([8], p. 21).

## **15.2 Methodology**

### ***15.2.1 Evaluation of Student Understanding***

The study originated from reported student difficulties in visualizing and comprehending complex concepts in the course. This point has been drawn from the feedback and comments from students during and at the end of semester. For example, students often struggle to understand and visualize the Metal Crystalline Patterns as shown in Fig. 15.1.

**Fig. 15.1** Unit cells for **a** body-centered cubic (bcc), **b** face-centered cubic (fcc), and **c** hexagonal close-packed structures (hcp)

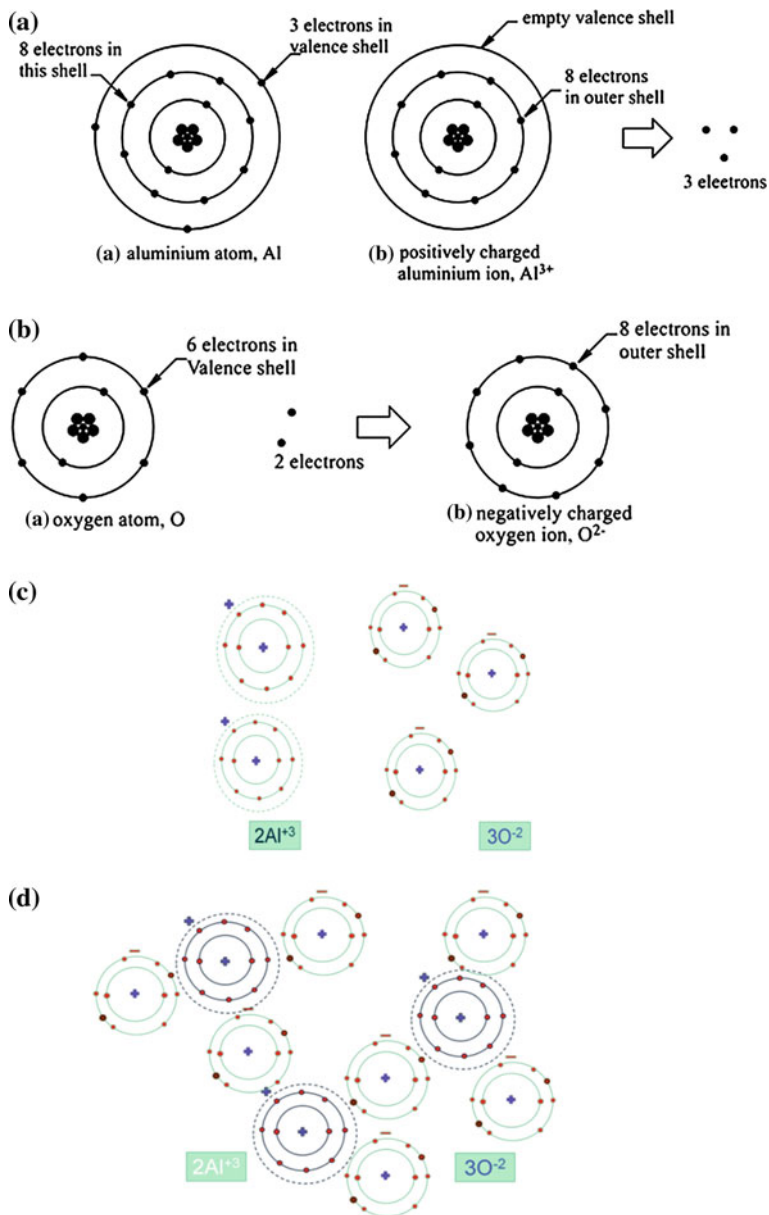


In addition, the course is offered to both on-campus and distance students providing an opportunity to determine whether there is a greater need for multimedia resources in the distance group due to lack of face-to-face contact with the lecturer.

### 15.2.2 Developing Multimedia Resources

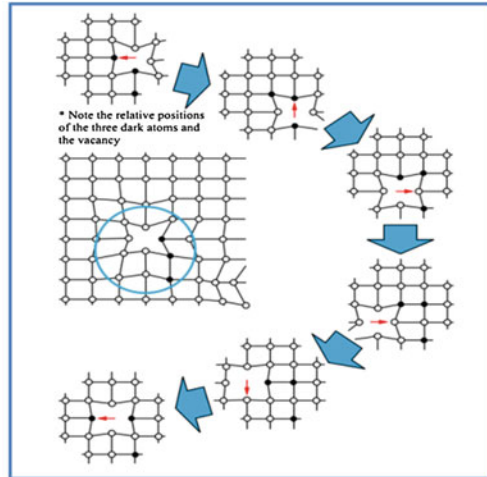
For each complex concept, the best fit in terms of technologically appropriate material was determined. For threshold concepts (for example Mechanical Testing), laboratory video footage was recorded with comprehensive verbal and inking explanations on how the tests should be performed, how data should be collected, and how the data should be interpreted. For Atomic Structure, animations were created and embedded in the lecture to show the movement of electrons around the nucleus. A similar method was implemented in presenting the Covenant and Ionic Bonds ( $\text{Al}_2\text{O}_3$ ), Tensile Behaviour of Metals (stretching in metallic bonds), Formation of Eutectic and Eutectoid in Binary Phase Diagrams. Animated flash images were developed to illustrate the concept of Grain Formation to be used in Equilibrium Phase Diagrams, Slip Line in Polymers and Corrosion in Metals. Figures 15.2, 15.3 and 15.4 display examples of traditional illustrations that were converted into animated ones.

For the benefit of external students, the lectures were recorded using Camtasia Relay software ([www.Techsmith.com](http://www.Techsmith.com)). The recorded lectures contained digital annotations which were used to augment the illustrations, animations and videos with explanations.

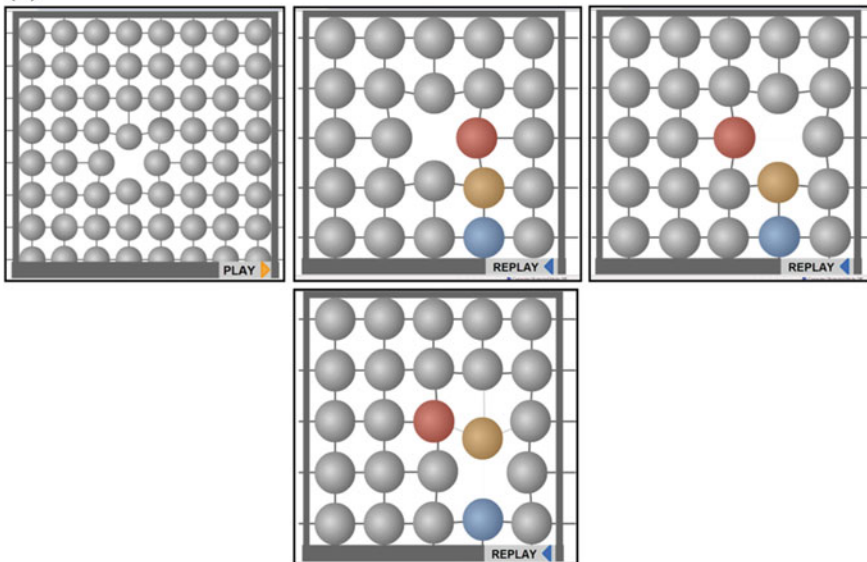


**Fig. 15.2** Covalent and ionic bonds in  $Al_2O_3$ : **a** and **b** without animation **c** first movement in the animated bonds, **d** continuous movement of the atoms

(a)



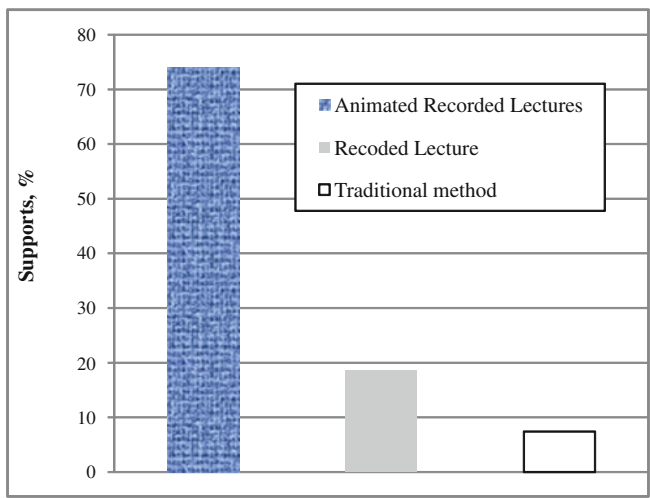
(b)



**Fig. 15.3** Diffusion in metal crystalline: **a** without animation **b** after animating the diffusion at different movements

### 15.3 Implementing Multimedia Resources

To evaluate the perceived effectiveness and impact of the animations embedded in the course content, three complex knowledge topics were selected to be presented in three different ways: as static PowerPoint slides, PowerPoint slides with inking, and lastly PowerPoint slides with inking and embedded multimedia resources. The



**Fig. 15.4** Formation of grain in binary phase diagram: **a** without animation **b** animation of grain formation at different stage

three topics were: binary phase diagram, microstructure of metals and tensile behaviour of metals. The presented topics were recorded using Camtasia Relay software and posted on the electronic publishing platform for use by both the on-campus and distance students.

An electronic survey was made available to 300 on-campus and distance students each semester for three consecutive semesters. The response rate was between 65 and 70 %. Feedback was collected through the e-mail and online survey facilities available in the online study environment, Moodle. The collected qualitative and quantitative data was analysed and plotted using Excel software.

### 15.4 Findings

The survey was categorized into two overarching sections, namely, the qualitative feedback from students regarding their preference of multimedia inclusions as a teaching technique, and the impact on the students’ perceived understanding and interaction with the teaching team.

#### 15.4.1 Inclusion of Multimedia in Content

Figure 15.4 shows the survey results relating to students’ preferences toward teaching with or without the use of multimedia components in lecture content. The



responses were divided into three categories of presentation techniques: (1) static PowerPoint slides, (2) PowerPoint slides with inking, and (3) PowerPoint slides with inking and embedded multimedia resources. Eighty per cent (80 %) of the students preferred the approach with both multimedia resources and inking embedded in the lecture slides. This result ties in with the finding that 60 % of students in a 4,000-student survey were sensory learners as evaluated on the Myers-Briggs Type indicators [18]. Students in this category reported an increased ability to understand the engineering threshold concepts and complex concepts. Beerman [18] reported that the use of multimedia technologies helps average and below average students in particular to learn more effectively. Moreover, presenting threshold concepts and complex knowledge in a technologically enhanced manner assisted these students in solving more complex problems in their assignments and final examinations. In [18], Beerman reported  $p < 0.05$  improvement in test scores where multimedia has been used in instruction. This was highly pronounced in the results of the off campus students. The percentage of failure has dropped to less than 5 % over the three semesters of the research project whilst maintaining the same difficulty level of questions in the final exam. This drop in failure was reported by [18] as early as 1996. In spite of these early results that indicate that there multimedia enhancements enable improved comprehension of abstract or complex material, there are still a lack of understanding of the fact that multimedia offers lecturers greater control over how and at what rate material is presented to students, [18]. It seems the deep understanding of the complex topics assisted the students to solve problems with the same level of difficulties to the introduced topics. At high levels of difficulties (e.g. interpreting complicated phase diagrams), about 65 % of the students achieved 60–70 % of the total marks. On the other hand, the students' feedback reflect their interest and understanding (e.g. "*even though I performed not well, but I know what are my mistakes*").

### ***15.4.2 Student Engagement and Interaction***

In previous offerings of this course, the students were provided with static study materials and recorded lectures without any multimedia resources (traditional method). When the traditional teaching content and methods were used, the interaction between the students was not as significant as during the application of the new method when the posted forums are compared between the previous and new semesters. The volume of interaction was determined through generating, two reports from the online learning environment, Figs. 15.5 and 15.6. The two reports show the number of posts which have been discussed in the online forums and the responses. The online learning environment is available to both on campus and distance students. It appears that multimedia enhancements of the lecture content have resulted in higher levels of interaction among the students and the teaching team compared to semesters in which no multimedia resources were utilised.

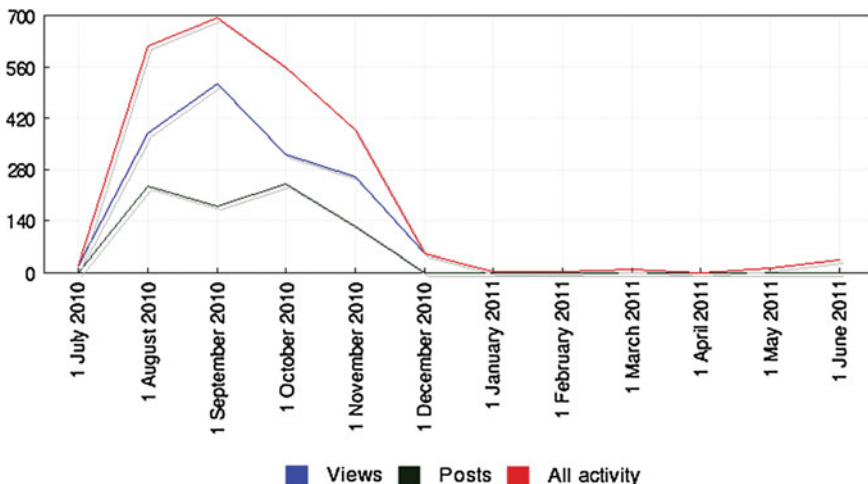


Fig. 15.5 Statistics of the activities by the students and the teaching team for the first semester of implementing the new style of teaching

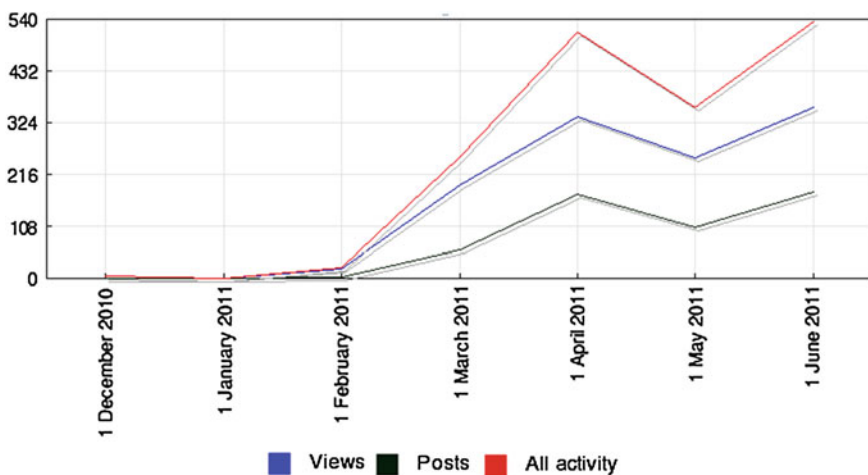


Fig. 15.6 Statistics of the activities by the students and the teaching team for the second semester of implementing the new style of teaching

Further research is required to extrapolate if this could be ascribed to the multimedia resources per se.

At the end of each semester, students are surveyed to determine their opinion of courses via a student questionnaire on the online learning environment system. One of the survey questions was “What did you find were the most helpful/effective aspects of this course?” 90 % of the students reported the recorded

lectures as the most helpful aspect in this course. This ties in with the finding of Hao [3] that multimedia enhances performance when students need more clues, that it lessens overload and helps students focus and that it is motivating.

In order to gain a more comprehensive view on the impact of embedded multimedia teaching resources on the higher order thinking of students, further research is required. If results are confirmed in further research projects, further embedding of technologically enriched content in this and other engineering courses could be recommended, particularly for enhancing understanding of threshold and complex concepts by distance education students.

## 15.5 Conclusion

This study investigated the impact of multimedia resources on the students' understanding when used in teaching an engineering materials course. This research concludes that there is a proven need for greater use of multimedia resources such as videos, animations, and inking resources in teaching threshold concepts and troublesome knowledge in engineering courses. The majority of the students reported on in this paper, indicated a preference for lectures with embedded multimedia resources, in particular animation and inking. Moreover, the survey conducted indicated that 73 % of the students reported an enhanced understanding of complex concepts in the Engineering Materials course. The students ascribed this enhanced comprehension to the embedded technological content and the inking explanations. To gain a comprehensive understanding of the impact of the multimedia resources on higher order thinking pertaining to Engineering Materials problem concepts, further research is recommended.

## References

1. Mayer RE, Moreno R (2003) Nine ways to reduce cognitive load in multimedia learning. *Educ Psychol* 38(1):43–52
2. Dalacosta K et al (2009) Multimedia application with animated cartoons for teaching science in elementary education. *Comput Educ* 52(4):741–748
3. Hao Y (2010) Does multimedia help students answer test items? *Comput Hum Behav* 26(5):1149–1157
4. Kaveevitichai C et al (2009) Enhancing nursing students' skills in vital signs assessment by using multimedia computer-assisted learning with integrated content of anatomy and physiology. *Nurse Educ Today* 29(1):65–72
5. Johnson TE, Archibald TN, Tenenbaum G (2010) Individual and team annotation effects on students' reading comprehension, critical thinking, and meta-cognitive skills. *Comput Hum Behav* 26(6):1496–1507
6. Holt JE, Radcliffe DF, School D (1985) Design or problem solving—a critical choice for the engineering profession. *Des Stud* 6(2):107–110

7. Worsley, S., M. Bulmer, and M. O'Brien. *Threshold concepts and troublesome knowledge in a second-level mathematics course*. in *Symposium Proceedings: Visualisation and Concept Development, UniServe Science, The University of Sydney*. 2008
8. Meyer JH, Land R (eds) (2006) Overcoming barriers to student learning. Threshold concepts and troublesome knowledge, p 19
9. Cousin G (2006) An introduction to threshold concepts. Planet 17:4–5 Special Issue on Threshold Concepts and Troublesome Knowledge
10. Land R. et al (2005) Threshold concepts and troublesome knowledge (3): implications for course design and evaluation. Improving Student Learning—equality and diversity. OCSLD, Oxford
11. Hopson MH, Simms RL, Knezek GA (2002) Using a technology-enriched environment to improve higher-order thinking skills. *J Res Technol Educ* 34(2):109–120
12. Miri B, David B-C, Uri Z (2007) Purposely teaching for the promotion of higher-order thinking skills: a case of critical thinking. *Res Sci Educ* 37(4):353–369
13. Austin KA (2009) Multimedia learning: cognitive individual differences and display design techniques predict transfer learning with multimedia learning modules. *Comput Educ* 53(4):1339–1354
14. Patterson DA (2011) Impact of a multimedia laboratory manual: Investigating the influence of student learning styles on laboratory preparation and performance over one semester. *Educ Chem Eng* 6(1):e10–e30
15. Ali J (2009) Cognitive support through visualization and focus specification for understanding large class libraries. *J Vis Lang Comput* 20(1):50–59
16. Pao Y-H, Meng Z (1998) Visualization and the understanding of multidimensional data. *Eng Appl Artif Intell* 11(5):659–667
17. McLoughlin D, Mynard J (2009) An analysis of higher order thinking in online discussions. *Innov Educ Teach Int* 46(2):147–160
18. Beerman KA (1996) Computer-based multimedia: new directions in teaching and learning. *J Nutr Educ* 28(1):15–18

# Index

## A

Active devices, 91, 104  
Advanced manufacturing, 6  
Anelastic displacement fields model, 97  
Annealing, 125, 126, 135, 137–139  
Artificial intelligence, 169, 177  
Artificial neural network (ANN), 177, 188  
Austenitization, 135  
Automated, 147, 149, 355, 336, 338

## B

Bingham plastic model, 46, 50, 51  
Bi-viscous model, 50  
Burr, 327, 331, 335, 341, 343–347, 356, 358, 360

## C

Camera, 372–375, 384, 386, 388, 395, 404  
Carbonyl iron nano-particles, 57, 58  
Chemical synthesis, 65, 69  
Chip formation, 287, 291, 294, 297, 300, 302, 303, 306, 315, 317, 329, 334, 336–338, 342, 352, 353  
CNC programming, 210  
Coatings, 332–334  
Complex mode analysis, 99  
Complex modulus, 92, 96, 98, 99  
Computer aided manufacturing (CAM), 242, 243, 246, 247, 258, 260, 271, 281  
Control strategy, 148, 149, 152, 161  
Cooling tests, 141  
Cutting edge radius, 327–329, 332, 335, 338, 342, 345, 352  
Cutting force, 332, 334, 335, 338, 341, 345, 352  
Cutting mechanics, 288, 292

## D

Deposited film, 70, 72, 78  
Design for environment, 7, 13  
Design of experiments (DOE)  
Digital factory, 219, 223  
Digital image, 367, 372, 373, 376, 390, 395, 404  
Dilatometer simulation, 135  
Direct frequency analysis, 98  
Drilling, 169, 184, 198, 368, 379, 381, 384, 387, 396

## E

Education, 412, 433, 441, 446  
Efficiency of the still, 20  
Electromagnetic circuit, 55  
Energy, 16–18, 20, 21, 25, 31, 33  
Engineering analysis, 136  
Engineering education, 449, 450, 452  
Environmentally conscious quality function deployment (ECQFD), 8, 13  
Environmentally impact assessment, 6, 10, 13  
Exergy, 25

## F

FEM modelling, 302, 310, 313, 314, 316  
Finishing, 243, 244, 248, 250, 252, 254, 256, 262, 265, 269, 271, 281  
Finite element method (FEM), 191  
Fluid modeling, 50  
Fluid properties, 56  
Fluid technology, 44, 59  
Forging, 169, 190–192, 196, 197  
Formability, 412, 413, 417–420, 424, 426–428, 434, 435, 445, 446  
Forming limit diagrams, 412, 424, 426, 434

- Fractional derivative model, 97, 98  
 Friction modelling, 311–313  
 Fuzzy, 177, 179, 180
- G**  
 Gaseous phase, 65, 69  
 Golla-Hughes-McTavish model, 96, 97  
 Grain, 66, 67, 69, 82, 111–113, 117, 118, 122, 125, 126, 131–134, 142, 143, 298, 300, 310, 316, 328, 332, 342, 343, 355, 360, 454  
 Grinding, 387, 388, 396
- H**  
 Hardware, 156, 157  
 Heat transfer modes, 20, 21  
 Heat treatments, 109, 124, 125, 128, 132, 138, 139, 143  
 Herschel–Bulkley model, 50–52  
 Heuristic optimization, 183  
 Higher order thinking, 450–452, 459, 460  
 Hydrothermal, 67
- I**  
 Illumination system, 373–375  
 Image acquisition system, 373, 404  
 Image processing methods, 376  
 Inorganic compound, 63, 72, 74, 75, 78, 84
- J**  
 Johnson, Mehl, Avrami and Kolmogorov (JMAK), 118, 120, 133
- L**  
 Laboratory tests, 412, 439, 441  
 Life cycle assessment (LCA), 9, 13  
 Linear regression models, 172, 186  
 Liquid-solid phase, 112  
 Lubricating oils, 64, 65, 70–72, 74, 75, 78, 80–83  
 Lubrication mechanism, 69, 70
- M**  
 Machining, 368, 369, 371–373, 376, 379, 386, 389, 391, 396, 401, 402, 404  
 Machining experiments, 241, 246, 248, 261, 264, 265, 267, 268, 270, 281, 298  
 Machining modelling, 242, 254, 288, 289, 301, 303, 304, 309  
 Machining sequences, 242, 244, 272, 277, 278  
 Machining setup, 243, 260, 271, 281  
 Machining strategy, 243, 246, 248, 250, 269–271  
 Magnetic circuit, 54, 55  
 Magnetic materials, 44, 52  
 Magneto-rheological (MR), 44, 45, 47, 48, 52, 56, 59  
 Management, 147, 148, 152, 159, 160, 166  
 Material damping, 90, 91  
 Material modelling, 130, 304, 307, 310, 311  
 Measurement, 412, 413, 416, 419–421  
 Mechanical engineering, 1, 3, 13  
 Mechanical engineering education, 1, 13  
 Mechanical micromachining, 285–289, 296–298, 300, 303, 305, 306, 310, 313, 316  
 Mechanical processes, 169, 181  
 Mechanical systems, 169, 180, 181  
 Meshing, 302, 303, 306, 307, 313  
 Metal forming, 138  
 Metals and alloys, 110, 114, 122  
 Micro cutting tools, 325, 327, 328, 332–334, 338, 343  
 Microemulsion, 65–67, 75  
 Micro machine tools, 334  
 Micromachining, 285–289, 297–300, 303, 305, 307–310, 313–316  
 Micromilling, 286, 298, 342  
 Micromilling mechanisms, 325, 334, 358  
 Microstructural information, 130–133  
 Microstructure, 110, 111, 116, 122–127, 130–133, 135, 143, 310, 314, 325, 342, 345, 452, 457  
 Milling, 201, 203, 204, 208, 211–214, 223–225, 227, 229–231, 233, 368, 375, 381, 385, 386, 388, 391, 396  
 Minimum chip thickness, 285, 287, 297, 298, 303, 316, 325, 334–338, 345, 348, 353–355  
 Modal strain energy, 99  
 Model geometry, 303  
 Modelling, 169–172, 177, 179, 180, 184, 190, 192, 195  
 Multimedia resources, 44–52, 54–59, 449, 451–454, 456–459  
 Multi-objective, 169, 181, 359  
 Multiphase steels, 109, 110, 131, 138, 143  
 Multi-process, 199, 211, 215, 216, 221, 227–237  
 Multi scale models, 139, 140–142

**N**

Nanoparticles, 65–67, 69–78, 82  
 Neural networks, 177  
 Neuro-fuzzy systems, 177, 179  
 Nonlinear regression models, 176  
 Numerical controller (NC), 199  
 Numerical integration, 98  
 Numerical simulation, 90, 93, 95, 441  
 Numeric optimization, 182, 183

**O**

Objective function, 254, 263  
 Oleylamine, 57, 58  
 Optimal real-time, 147  
 Optimization, 100–103, 105, 169, 181–184, 190, 196, 197, 199, 204, 209–211, 214, 215, 218–220, 226–232, 234, 235, 237  
 Optimization algorithm, 154–156, 160

**P**

Passive devices, 90, 91  
 Pentacarbonyl compound, 57  
 Permeation, 70, 83  
 Phase change material (PCM), 24–27, 30, 33  
 Phase change material storage  
 Phase transformation, 110–112, 116, 118, 121, 123, 131, 133, 134, 142  
 Plant setup, 148  
 PLC network, 147, 151, 152, 156, 157  
 Precipitation, 67  
 Probabilistic methods, 177, 180  
 Production lines, 147, 148, 158, 160, 164, 166  
 Productivity, 28, 30, 31, 33, 35, 37, 38  
 Programmable logic controllers (PLCs), 147–149, 151, 156, 159, 166  
 Prototypes, 203, 204, 224

**R**

Rare-earth, 64–69, 72, 74, 81–83  
 Rare-earth additives, 63, 74  
 Rare-earth borate, 78, 79  
 Rare-earth dialkylidithiophosphate, 79, 80, 81, 83  
 Rare-earth fluoride, 75, 76, 78  
 Rare-earth naphthenates, 74, 82  
 Rolling ball, 71  
 Roughing, 242–244, 248, 249, 254–257, 259–264, 269

**S**

SCADA Supervisor, 150, 166  
 Sculptured surface machining (SSM), 241–243, 246, 248, 250, 251, 267, 281  
 Shear mode, 47–49  
 Sheet metal forming, 411–413, 416, 420, 426–428, 433, 435, 439, 441, 446  
 Simulation, 199, 200, 202, 207, 208, 210, 213–215, 220, 221, 227, 228, 231–235, 237  
 Simulative tests, 427, 428  
 Single objective, 181  
 Size effect, 327, 334–336, 342, 345, 352, 353, 360  
 Software, 155–157  
 Solar distillation, 17, 18, 27  
 Solar drying technologies, 17  
 Solar still, 24, 25, 27, 28  
 Sol-gel, 65, 66  
 Solid phase, 65, 69  
 SPAIM (STEP-MC platform for advanced and intelligent manufacturing), 199, 204, 205, 207–216, 216–219, 223, 224, 226, 229, 230, 234, 236, 237  
 Squeeze mode, 49  
 STEP-NC, 199–219, 221–237  
 Strain measures, 412–415, 421  
 Structural dynamics, 89, 90, 104  
 Student engagement, 449, 453, 458  
 Supercritical fluid, 68  
 Supervisory control and data acquisition (SCADA), 147–149, 151–156, 161, 164, 166  
 Surface chemical modification, 72, 73  
 Surface physical modification, 72  
 Surface quality, 327, 332–337, 344, 353–357  
 Surface roughness, 169, 184, 198  
 Surface texture, 369, 375, 379, 382, 392, 399, 401, 404  
 Sustainability, 1–3, 5–7, 11–13  
 Sustainable materials, 3, 4, 11  
 Sustainable processes, 3, 5  
 Sustainable products, 3, 5  
 Synthetic oil, 59

**T**

Technological constraints, 241, 244, 248, 258, 281  
 Temperature, 329, 342, 345, 352, 356, 357  
 Thermodynamics, 110, 130, 142

Thermomechanical modeling, 143  
Threshold concepts, 449–454, 457, 460  
Tool condition monitoring (TCM), 368, 369, 372, 373, 384, 390, 404  
Tool life, 327–331, 333, 335, 344, 347, 348, 354, 356, 358  
Tool wear, 368, 369, 371, 372  
Transformation kinetics, 118, 121, 123, 132, 133, 142  
Tribochemical reaction, 70, 72, 75, 77–79, 83  
Tribological behavior, 65, 68, 80, 82  
Troublesome knowledge, 449–451, 453, 460  
Turning, 368, 385, 386, 388, 390, 391, 396, 398

**V**

Valve mode, 47, 48  
Viscoelastic constitutive model, 93, 95  
Viscoelastic damping treatments, 91, 93, 100, 104  
Viscoelastic materials, 91, 92

**W**

Water spray quenching, 129  
Workpiece materials