G. Pahl
W. Beitz
J. Feldhusen
K.H. Grote
Ken Wallace · Luciënne Blessing
Translators and Editors

Engineering Design

A Systematic Approach

Third Edition

Springer

Engineering Design

A Systematic Approach

G. Pahl and W. Beitz
J. Feldhusen and K.-H. Grote

Engineering Design

A Systematic Approach

Third Edition

Ken Wallace and Luciënne Blessing
Translators and Editors

Springer

Gerhard Pahl, em. Prof. Dr. h.c. mult.
Dr.-Ing. E.h. Dr.-Ing.
Fachbereich Maschinenbau
Technische Universität Darmstadt
Magdalenstrasse 4
64298 Darmstadt
Germany

† Wolfgang Beitz, Prof. Dr.-Ing.
E.h. Dr.-Ing.
1935–1998

Jörg Feldhusen, Prof. Dr.-Ing.
Institut für Allgemeine
Konstruktionslehre
des Maschinenbaus
Rheinisch Westfälische
Technische Hochschule
Aachen
Steinbachstrasse 54B
52074 Aachen
Germany

Karl-Heinrich Grote, Prof. Dr.-Ing.
Institut für Maschinenkonstruktion
Otto-von-Guericke-Universität
Magdeburg
Universitätsplatz 2
39106 Magdeburg
Germany

British Library Cataloguing in Publication Data
Engineering design : a systematic approach. – 3rd ed.
1. Engineering design
I. Pahl, G. (Gerhard), 1925– II. Wallace, Ken
620'.0042
ISBN-10: 1846283183

Library of Congress Control Number: 2006938893

ISBN 978-1-84628-318-5 3rd edition e-ISBN 978-1-84628-319-2 3rd edition
Printed on acid-free paper
ISBN 3-540-19917-9 2nd edition

3rd English edition, Springer 2007
2nd English edition, Springer 1996
1st English edition published by The Design Council, London, UK (ISBN 085072239X)

9 8 7 6 5 4 3 2 1

Springer Science+Business Media
springer.com

Preface

Sadly, just one year after the publication of the fourth German edition in 1997, my co-author Wolfgang Beitz died after a short but severe illness. His many outstanding contributions to engineering design, including his contribution to this book, were honoured in a memorial colloquium held in Berlin. It would have made me very happy if he had been able to see the continuing success of our book, including its translation into Portuguese. Our collaboration was a perfect one—always fruitful, always beneficial. I am deeply grateful to him.

The book, "Pahl/Beitz—Konstruktionslehre", has now been translated into eight languages and recognised as an international reference text. For reasons of continuity, our publisher Springer wanted to publish a fifth German edition of the book. To assist with this task two former students of Wolfgang Beitz became involved: Professor Dr.-Ing. Jörg Feldhusen and Professor Dr.-Ing. Karl-Heinrich Grote, both of whom have continually promoted and expanded his ideas. Professor Feldhusen worked for many years as a senior designer in the automotive industry and is now at RWTH Aachen University, succeeding Professor Dr.-Ing. R. Koller. Professor Grote has considerable experience of teaching design and running projects as a Professor in the USA, and is now at the Otto-von-Guericke University in Magdeburg. He succeeded Professor Beitz as the Editor of the Dubbel Handbook for Mechanical Engineering.

Gerhard Pahl
Darmstadt

Authors' Forewords

Sixth German Edition

The fifth German edition, which was published in March 2003, was so well received that just a year later a sixth German edition was required. The opportunity was taken to add some new developments to the chapter on size ranges and modular products.

The authors would like to reiterate their thanks to all those involved in both editions.

G. Pahl, J. Feldhusen and K.-H. Grote
Darmstadt, Aachen and Magdeburg, April 2004

Fifth German Edition

For the fifth German edition we have retained the well-established pattern of the previous editions, but updated it with new material. Because of its widespread use, the basics of electronic data processing*, including CAD, have been moved into the chapter on fundamentals. The chapter on the product development process has been expanded and strengthened by adding new perspectives. As a result, Chapters 1–4 now fully represent the necessary basic knowledge, including cognitive aspects, needed to underpin a systematic approach to engineering design. Chapters 5–8 describe the application of this basic knowledge to product development from the task clarification phase, through conceptual design up to the final embodiment and detail design* phases, supported by many detailed examples. Chapter 9 describes some important generic solutions including composite structures*, mechatronics and adaptronics. Basic knowledge about machine elements is, as always, assumed. Chapter 10 covers, as in previous editions, the development of size ranges and modular products. The increasing importance of achieving high quality is reflected by additions to

* The starred topics do not appear in this third English edition and as a consequence some chapter numbers have changed—see Editors' Foreword.

Chapter 11. The important theme of estimating costs can be found, as before, in Chapter 12. Because the basics of data processing technology have now been included in the chapter on fundamentals, Chapter 13 focuses on general recommendations for designing with CAD*. Chapter 14 provides an overview of the recommended methods, and reports on experiences of using the approach in industrial practice. The book closes with a definition of terms* as they have been used in this book. The index supports a rapid search for specific themes.

In this way, the systematic approach to engineering design has been brought to a level that provides a basis for successful product development. Throughout, fundamentals have been emphasised and short-term trends avoided. The approach described also provides a sound basis for design education courses that help students move into design practice. The literature has been updated, offering those who are interested in more detail or in the historical background a rich source of information.

The authors have to thank many individuals. Frau Professor Dr.-Ing. L. Blessing, successor to Professor Wolfgang Beitz, kept the original figures and made them available to us. Professor Dr.-Ing. K. Landau, TU Darmstadt, helped us update the literature on design for ergonomics. Professors Dr.-Ing. B. Breuer, Dr.-Ing. H. Hanselka, Dr.-Ing. R. Isermann and Dr.-Ing. R. Nordmann, all from TU Darmstadt, contributed to the sections on mechatronics and adaptronics with suggestions, examples and figures. In this connection we also thank Dr.-Ing. M. Semsch for his contribution. Emeritus Professor Dr.-Ing. M. Flemming, ETH Zurich, greatly supported us with suggestions and figures on the themes of composite construction* and structronics. Last but not least, we thank all those hardworking assistants, such as Frau B. Frehse at the Institut für Maschinenkonstruktion-Konstruktionstechnik, Universität Magdeburg, who prepared and reworked the electronic transformation of the text and figures. Finally we warmly thank our publisher Springer, in particular Dr. Riedesel, Frau Hestermann-Beyerle, Frau Rossow and Herr Schoenefeldt for their continuous support and for the excellent printing of the text and figures.

G. Pahl, J. Feldhusen and K.-H. Grote
Darmstadt, Aachen and Magdeburg, June 2002

Fourth German Edition

The third edition of our book proved to be so popular that after a relatively short time a further edition was required. A reprint was not considered appropriate as several important new concepts and methods for the product development process had emerged, and these could not be ignored. Furthermore recently published findings needed to be taken into account.

The structure and content of the third edition forms the basis of the fourth edition. The topic of product planning has been extended through the integration of methods such as portfolio analysis and scenario planning. New sections have been introduced on effective organisation structures, on applying simultaneous engineering, on leadership and on team behaviour. The increasing importance of quality assurance has reinforced the need to adopt systematic engineering design as a primary measure. This should be extended through the application of secondary measures, such as Quality Function Deployment (QFD) using the House of Quality. Developments in the area of sustainability have led to modifications in the section on design for recycling. Because of its general technical and economic importance, a new section on design to minimise wear has been introduced. The method of target costing has been included in the chapter on design for minimum cost. Finally, the chapter on CAD required updating*.

The third edition, slightly abridged, has been translated into English, *Engineering Design: A Systematic Approach* (2nd Edition, Springer-Verlag, London), under the leadership of Ken Wallace, who was supported by Luciënne Blessing and Frank Bauert. We thank them warmly. A Japanese translation has also been published, and a translation into Korean is in progress. These translations significantly increase the international influence of *Konstruktionslehre*.

The employees of both our institutes have again supported our work on the fourth edition in their usual trusted and willing way. For their help we are deeply grateful. Our publishers have again to be thanked for the excellent advice we have received, as well as for their careful realisation of the book. Finally, we thank our wives for their continuous understanding, for without their support this book would never have been possible.

G. Pahl and W. Beitz
Darmstadt and Berlin, January 1997

Editors' Foreword

Background

The first German edition of *Konstruktionslehre* was published in 1977. The first English edition entitled *Engineering Design* was published in 1984 and was a full translation of the German text. Both the German and the English editions of the book rapidly became established as important references on systematic engineering design in industry, research and education. International interest in engineering design grew rapidly during the 1980s and many developments took place. To keep up-to-date with the changes, a second German edition was published in 1986. It was too soon after the publication of the first English edition to consider a second edition. However, since the translation was being extensively used to support engineering design teaching, a slightly abridged student edition entitled *Engineering Design – A Systematic Approach* was published in 1988.

When preparing the student edition, the opportunity was taken to review the translation and the contents of the first edition. No changes in terminology were thought necessary and the contents were the same as the first English edition except for the removal of two chapters.

The first chapter to be removed was the short chapter on detail design. It must be emphasised that this does not mean that detail design is considered unimportant or lacking in intellectual challenge. Quite the reverse is true. Detail design is far too broad and complex a subject to be covered in a general text. There are many excellent books covering the detail design of specific technical systems and machine elements. For these reasons, the German editions did not discuss technical aspects of detail design, but only dealt with the preparation of production documents and the numbering techniques required to keep track of them.

The second chapter to be removed dealt with computer support for design, including CAD. Again, this chapter was clearly not removed because the topic is unimportant. Computer support systems are used universally and develop rapidly. Many specialist texts are available.

In 1993 an updated and extended third German edition of *Konstruktionslehre* was published. It was considered timely to produce a second English edition to bring the translation into step with the latest thinking. The new layout of the German edition was incorporated, along with the important discussions of psychology and recycling. The new chapters on design for quality and design for minimum cost were included, but, for the reasons given above, the chapters on detail design and computer support were again omitted.

The third German edition also contained a new chapter that described selected standard solutions (machine elements, drives and controls) in line with the systematic approach and concepts presented in the book. This knowledge is covered comprehensively in the translation of the German *Dubbel* [*Dubbel Handbook for Mechanical Engineering*, Springer-Verlag, London, 1994]. This chapter was therefore also omitted.

There are now six German editions of Pahl/Beitz (4th 1997; 5th 2003; 6th 2005)—so it is timely to produce a third English edition. The structure has changed compared to the previous English edition and is described below.

Structure of the Third English Edition

Introduction—Chapter 1

The book starts with the historical background to modern systematic design thinking in Germany. The work of influential design researchers and practitioners is reviewed briefly.

Fundamentals—Chapter 2

This chapter discusses the fundamentals of technical systems and of the systematic approach, including cognitive aspects. The fundamentals of the use of computers to support product development were omitted for the reasons mentioned above.

Product Planning, Solution Finding and Evaluation—Chapter 3

In this chapter the flow of work during the process of planning is described, see Figure 3.2, along with general methods for finding and evaluating solutions that can be used not only for planning but also throughout the product development process. These methods are not linked to any specific design phase or type of product and include a range of intuitive and discursive methods.

Product Development Process—Chapter 4

This chapter presents the flow of work during the product development process and describes the main phases: Task Clarification; Conceptual Design; Embodiment Design; and Detail Design. The authors' overall model is shown in Figure 4.3. New to this edition is a discussion about the effective management and organisation of the design process.

Task Clarification—Chapter 5

This phase involves identifying and formulating the general and task-specific requirements and constraints, and setting up a requirements list (design specification). The steps of this phase are shown in Figure 5.1.

Conceptual Design—Chapter 6

This phase involves (see Figure 6.1):

- abstracting to find the essential problems
- establishing function structures
- searching for working principles
- combining working principles into working structures
- selecting a suitable working structure and firming it up into a principle solution (concept).

This chapter concludes with two detailed examples of applying the proposed methods to the design of a single-handed water mixing tap and an impulse-loading test rig.

Embodiment Design—Chapter 7

During this phase, designers start with the selected concept and work through the steps shown in Figure 7.1 to produce a definitive layout of the proposed technical product or system in accordance with technical and economic requirements.

About 40% of the book is devoted to this phase and the authors discuss the basic rules, principles and guidelines of embodiment design, followed by a comprehensive example of the embodiment design of the impulse-loading test rig introduced in Chapter 6.

The chapter on detail design has again been omitted, but a new Section 7.8 outlining the steps of this phase has been introduced (see Figure 7.164).

Mechanical Connections, Mechatronics and Adaptronics—Chapter 8

This chapter is new to the English series of Pahl/Beitz. Three classes of generic solutions are presented in a way that is consistent with the systematic approach presented in this book. Because of their overriding importance in mechanical design, mechanical connections are the first class to be discussed. Because of their growing importance, the other two classes are mechatronic and adaptronic systems.

The decision was taken to leave out drives, control systems and composite structures as these are covered extensively in the English literature.

Size Ranges and Modular Products—Chapter 9

This chapter presents methods for systematically developing size ranges and modular products to meet a wide range of requirements while at the same time reducing costs. In this edition the concepts of product architecture and platform construction are introduced.

Design for Quality—Chapter 10

The chapter on design for quality now includes a discussion of Quality Function Deployment (QFD).

Design for Minimum Cost—Chapter 11

This chapter now includes a section on Target Costing.

Summary—Chapter 12

The short final chapter provides a summary of the ideas covered in the book. Figures 12.1 and 12.2 provide a quick reference to the main steps in the design process and the appropriate working methods.

Every design must meet both task-specific and general requirements and constraints. To remind designers of these during all stages of the design process, a set of checklists is used throughout the book. An overview of these checklists is provided in Figure 12.3.

Translation Issues

The aim of the translation has been to render each section of the book comprehensible in its own right and to avoid specialist terminology. Terms are defined as they arise, rather than in a separate glossary, and their meanings should be clear from their usage. On occasions other authors have used slightly different terms, but it is hoped that no misunderstandings arise and that the translation is clear and consistent throughout.

Some terms, however, require special mention. The German methodology includes a standard concept introduced with the German prefix 'wirk'. Translators have used a number of different English terms to translate 'wirk', including 'active', 'working' and 'effective'. After careful consideration, we decided to continue to use 'working' as in the previous English edition, so, for example, 'wirkprinzip' becomes 'working principle', 'wirkort' become 'working location', 'wirkfläche' becomes 'working surface' and 'wirkbewegung' becomes 'working motion'. In

English 'working' does not immediately convey fully the correct German meaning. In German, the 'wirk' prefix is used to focus on the principles, locations and surfaces, etc. that ensure the desired physical effect takes place. So, for example, 'wirkort' (working location) is where the physical effect takes place using two or more 'wirkflächen' (working surfaces) and a 'wirkbewegung' (working motion). 'Wirkprinzip' brings these ideas together as the 'working principle'. For example 'clamping' is the working principle that can realise the friction effect by preventing certain working motions through an appropriate combination of suitable working surfaces (see Figure 2.12).

The term 'drawing' is used in this book to represent the output of either a traditional design approach, i.e. a physical drawing, or a modern computer-supported approach, i.e. a CAD model or drawing.

Of the four phases of the product design process, only the terminology used for the third, 'embodiment design', requires some explanation. Other translations, in a similar context, have used layout design, main design, scheme design or draft design. The input to this third phase is a design concept and the output is a technical description, often in the form of a scale drawing or CAD model. Depending on the particular company involved, this drawing is referred to as a general arrangement, a layout, a scheme, a draft, or a configuration, and it defines the arrangement and preliminary shapes of the components in a technical artefact. The term 'layout' is widely used and was selected for this book. The idea to introduce the term embodiment design came from French's book, *Engineering Design: The Conceptual Stage*, published in 1971. Embodiment design incorporates both layout design (the arrangement of components and their relative motions) and form design (the shapes and materials of individual components). The term 'form design' is widely used in the literature, and its meaning ranges from the overall form of a product in an industrial design context, to the more restricted form of individual components in an engineering context. This book tends towards the latter usage.

There are numerous references to DIN (Deutsche Industrie Normen) standards and VDI (Verein Deutscher Ingenieure) guidelines, a few of which have been translated into English. Examples are the DIN ISO standards and the translation of VDI 2221. In important cases, references to DIN standards and VDI guidelines have been retained in the English text, but elsewhere they have simply been listed along with the other references. In technical examples, DIN standards have been referred to without any attempt to find English equivalents.

The original text includes many references. Most of these are in German and therefore not of immediate interest to the majority of English readers. However, to have omitted them would have detracted from the authority of the book and its value as an important source of reference. The references have therefore been retained in full but grouped together at the end of the book, rather than at the end of each chapter as in the German text. An English bibliography has been added by the Editors, as well as an overview of the main engineering design conference series and journals.

It must be stressed that nothing was deleted that detracted from the main aim of the original German book, that is, to present a comprehensive, consistent and clear approach to systematic engineering design.

Acknowledgements

Donald Welbourn was responsible for encouraging the translation of the first English edition in the late 1970s, and he helped and supported the task in numerous ways. Many of the challenges that arose with the translation and terminology at the time were resolved with the help of Arnold Pomerans.

We first worked together on the translation of the second English edition, and Frank Bauert assisted us with the new figures. Nicholas Pinfield from Springer provided encouragement and support throughout.

For the third English edition, we worked jointly on the overall task of translation and editing.

John Clarkson helped with the compilation of the English bibliography. Anthony Doyle and Nicolas Wilson from Springer contributed enormously to the overall production of the book and their help and patience are gratefully acknowledged. Sorina Moosdorf from LE-TEX in Germany was responsible for the detailed task of typesetting the book. She and her colleagues did an excellent job.

Finally, and most sincerely, we must thank Professor Pahl, Professor Feldhusen and Professor Grote for trusting us with the translation of the book.

As with the previous two editions, it is hoped that this translation faithfully conveys the ideas of *Pahl/Beitz – Konstruktionslehre* while adopting an English style.

Ken Wallace and Luciënne Blessing
Cambridge and Berlin, November 2006

Contents

1 Introduction

1.1 The Engineering Designer

1.1.1 Tasks and Activities

The main task of engineers is to apply their scientific and engineering knowledge to the solution of technical problems, and then to optimise those solutions within the requirements and constraints set by material, technological, economic, legal, environmental and human-related considerations. Problems become concrete tasks after the problems that engineers have to solve to create new technical products (artefacts) are clarified and defined. This happens in individual work as well as in teams in order to realise interdisciplinary product development. The mental creation of a new product is the task of design and development engineers, whereas its physical realisation is the responsibility of production engineers.

In this book, *designer* is used synonymously to mean design and development engineers. Designers contribute to finding solutions and developing products in a very specific way. They carry a heavy burden of responsibility, since their ideas, knowledge and skills determine the technical, economic and ecological properties of the product in a decisive way.

Design is an interesting engineering activity that:

- affects almost all areas of human life
- uses the laws and insights of science
- builds upon special experience
- provides the prerequisites for the physical realisation of solution ideas
- requires professional integrity and responsibility.

Dixon [1.39] and later Penny [1.144] placed the work of engineering designers at the centre of two intersecting cultural and technical streams (see Figure 1.1).

However, other models are also available. In *psychological* respects, designing is a creative activity that calls for a sound grounding in mathematics, physics, chemistry, mechanics, thermodynamics, hydrodynamics, electrical engineering, production engineering, materials technology, machine elements and design theory, as well as knowledge and experience of the domain of interest. Initiative,

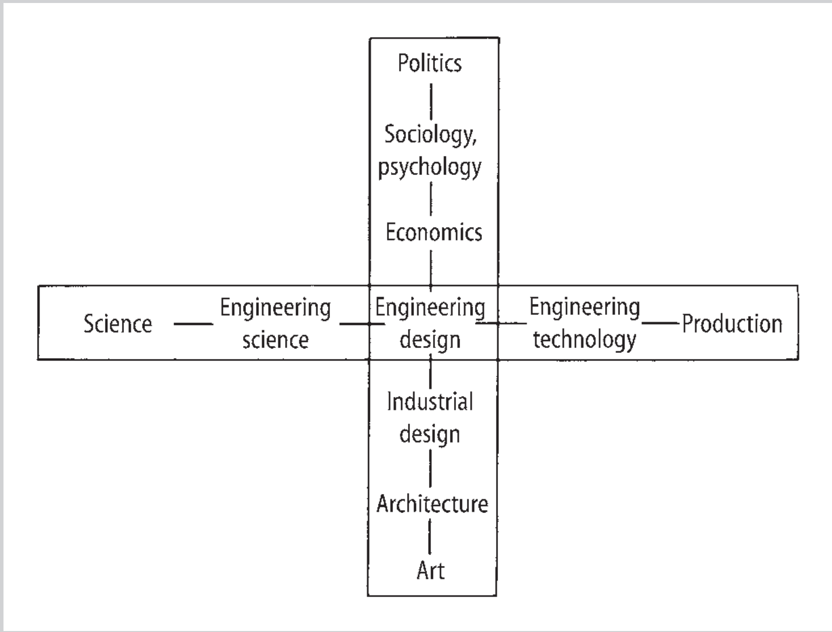


Figure 1.1. The central activity of engineering design. After [1.39, 1.144]

resolution, economic insight, tenacity, optimism and teamwork are qualities that stand all designers in good stead and are indispensable to those in responsible positions [1.130] (see Section 2.2.2).

In *systematic* respects, designing is the optimisation of given objectives within partly conflicting constraints. Requirements change with time, so that a particular solution can only be optimised for a particular set of circumstances.

In *organisational* respects, design is an essential part of the product life cycle. This cycle is triggered by a market need or a new idea. It starts with product planning and ends—when the product's useful life is over—with recycling or environmentally safe disposal (see Figure 1.2). This cycle represents a process of converting raw materials into economic products of high added value. Designers must undertake their tasks in close cooperation with specialists in a wide range of disciplines and with different skills (see Section 1.1.2).

The tasks and activities of designers are influenced by several characteristics.

Origin of the task: Projects related to mass production and batch production are usually started by a product planning group after carrying out a thorough analysis of the market (see Section 3.1). The requirements established by the product planning group usually leave a large solution space for designers.

In the case of a customer order for a specific one-off or small batch product, however, there are usually tighter requirements to fulfil. In these cases it is wise for designers to base their solutions on the existing company know-how that has been built up from previous developments and orders. Such developments usually take place in small incremental steps in order to limit the risks involved.

If the development involves only part of a product (assembly or module), the requirements and the design space are even tighter and the need to interact with other design groups is very high. When it comes to the production of a product,

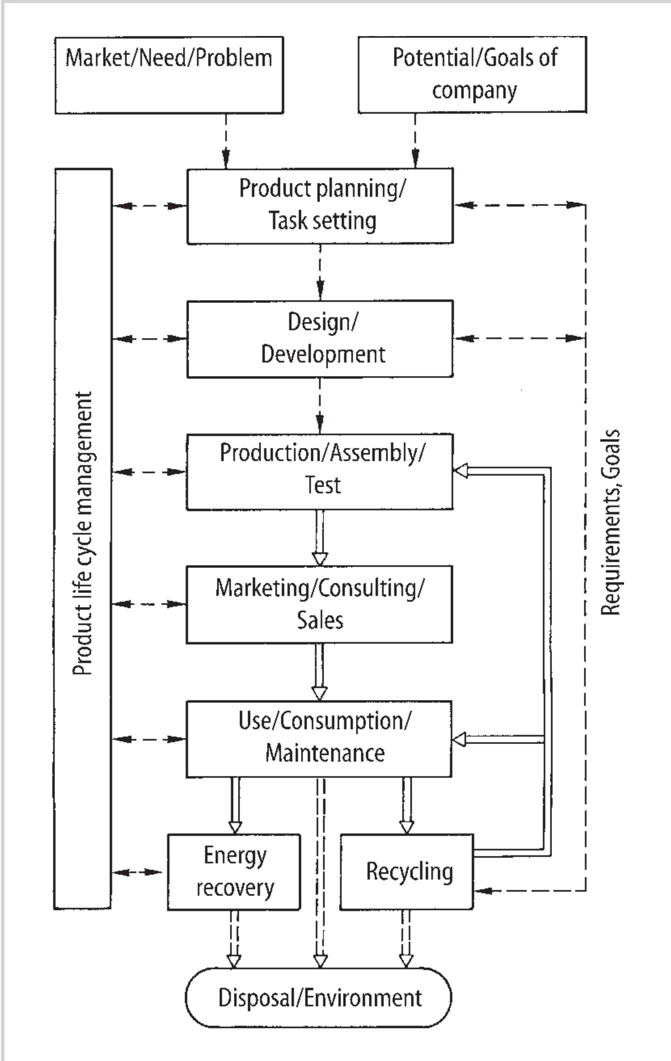


Figure 1.2. Life cycle of a product

there are design tasks related to production machines, jigs and fixtures, and inspection equipment. For these tasks, fulfilling the functional requirements and technological constraints is especially important.

Organisation: The organisation of the design and development process depends in the first instance on the overall organisation of the company. In product-oriented companies, responsibility for product development and subsequent production is split between separate divisions of the company based on specific product types (e.g. rotary compressor division, piston compressor division, accessory equipment division).

Problem-oriented companies split the responsibility according to the way the overall task is broken down into partial tasks (e.g. mechanical engineering, control systems, materials selection, stress analysis). In this arrangement the project manager must pay particular attention to the coordination of the work as it passes from group to group. In some cases the project manager leads independent temporary project teams recruited from the various groups. These teams report directly to the head of development or senior management (see Section 4.3).

Other organisational structures are possible, for example based on the particular phase of the design process (conceptual design, embodiment design, detail design), the domain (mechanical engineering, electrical engineering, software development), or the stage of the product development process (research, design, development, pre-production) (see Section 4.2). In large projects with clearly delineated domains, it is often necessary to develop individual modules for the product in parallel.

Novelty: New tasks and problems that are realised by *original designs* incorporate new solution principles. These can be realised either by selecting and combining known principles and technology, or by inventing completely new technology. The term original design is also used when existing or slightly changed tasks are solved using new solution principles. Original designs usually proceed through all design phases, depend on physical and process fundamentals and require a careful technical and economic analysis of the task. Original designs can involve the whole product or just assemblies or components.

In *adaptive design*, one keeps to known and established solution principles and adapts the embodiment to changed requirements. It may be necessary to undertake original designs of individual assemblies or components. In this type of design the emphasis is on geometrical (strength, stiffness, etc.), production and material issues.

In *variant design*, the sizes and arrangements of parts and assemblies are varied within the limits set by previously designed product structures (e.g. size ranges and modular products, see Chapter 9). Variant design requires original design effort only once and does not present significant design problems for a particular order. It includes designs in which only the dimensions of individual parts are changed to meet a specific task. In [1.124, 1.167] this type of design is referred to as *principle design* or *design with fixed principle*.

In practice it is often not possible to define precisely the boundaries between the three types of design, and this must be considered to be only a broad classification.

Batch size: The design of one-off and small batch products requires particularly careful design of all physical processes and embodiment details to minimise risk. In these cases it is usually not economic to produce development prototypes. Often functionality and reliability have a higher priority than economic optimisation.

Products to be made in large quantities (large batch or mass production) must have their technical and economic characteristics fully checked prior to full-scale production. This is achieved using models and prototypes and often requires several development steps (see Figure 1.3).

Branch: Mechanical engineering covers a wide range of tasks. As a consequence the requirements and the type of solutions are exceptionally diverse and always require the application of the methods and tools used to be adapted to the specific task in hand. Domain-specific embodiments are also common. For example, food processing machines have to fulfil specific requirements regarding hygiene; machine tools have to fulfil specific requirements regarding precision and operating speed; prime movers have to fulfil specific requirements regarding power-to-weight ratio and efficiency; agricultural machines have to fulfil specific requirements re-

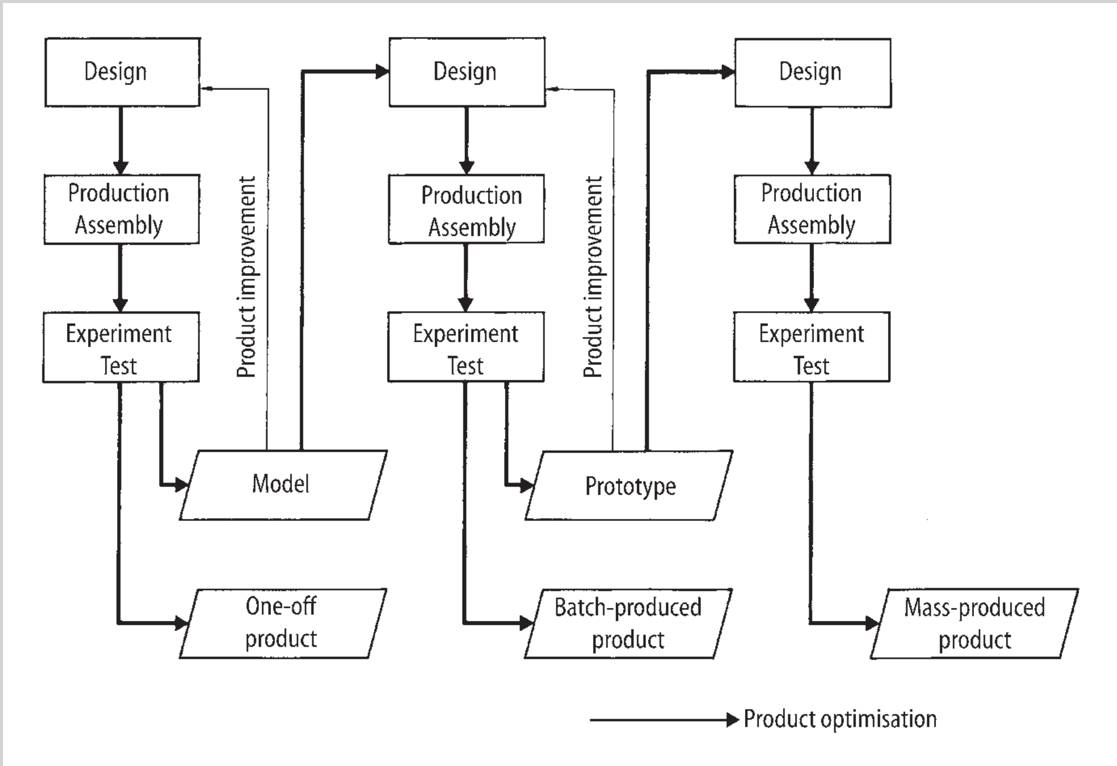


Figure 1.3. Stepwise development of a mass-produced product. After [1.191]

garding functionality and robustness; and office machines have to fulfil specific requirements regarding ergonomics and noise levels.

Goals: Design tasks must be directed towards meeting the goals to be optimised, taking into account the given restrictions. New functions, longer life, lower costs, production problems, and changed ergonomic requirements are all examples of possible reasons for establishing new design goals.

Moreover, an increased awareness of environmental issues frequently requires completely new products and processes for which the task and the solution principle have to be revisited. This requires a holistic view on the part of designers and collaboration with specialists from other disciplines.

To cope with this wide variety of tasks, designers have to adopt different approaches, use a wide range of skills and tools, have broad design knowledge and consult specialists on specific problems. This becomes easier if designers master a general working procedure (see Section 2.2.4), understand generation and evaluation methods (see Chapter 3) and are familiar with well-known solutions to existing problems (see Chapters 7 and 8).

The activities of designers can be roughly classified into:

- Conceptualising, i.e. searching for solution principles (see Chapter 6). Generally applicable methods can be used along with the special methods described in Chapter 3.
- Embodying, i.e. engineering a solution principle by determining the general arrangement and preliminary shapes and materials of all components. The methods described in Chapters 7 and 9 are useful.

- Detailing, i.e. finalising production and operating details.
- Computing, representing and information collecting. These occur during all phases of the design process.

Another common classification is the distinction between *direct* design activities (e.g. conceptualising, embodying, detailing, computing), and *indirect* design activities (e.g. collecting and processing information, attending meetings, coordinating staff). One should aim to keep the proportion of the indirect activities as low as possible.

In the design process, the required design activities have to be structured in a purposeful way that forms a clear sequence of main phases and individual working steps, so that the flow of work can be planned and controlled (see Chapter 4).

1.1.2 Position of the Design Process within a Company

The design and development department is of central importance in any company. Designers determine the properties of every product in terms of function, safety, ergonomics, production, transport, operation, maintenance, recycling and disposal. In addition, designers have a large influence on production and operating costs, on quality and on production lead times. Because of this weight of responsibility, designers must continuously reappraise the general goals of the task in hand (see Section 2.1.7).

A further reason for the central role of designers in the company is the position of design and development in the overall product development process. The links and information flows between departments are shown in Figure 1.4, from which it can be seen that production and assembly depend fundamentally on information from product planning, design and development. However, design and development are strongly influenced by knowledge and experience from production and assembly.

Because of current market pressures to increase product performance, lower prices and reduce the time-to-market, product planning, sales and marketing must draw increasingly upon specialised engineering knowledge. Because of their key position in the product development process, it is therefore particularly important to make full use of the theoretical knowledge and product experience of designers (see Section 3.1 and Chapter 5).

Current product liability legislation [1.12] demands not only professional and responsible product development using the best technology but also the highest possible production quality.

1.1.3 Trends

The most important impact in recent years on the design process, and on the activities of designers, has come from computer-based data processing. Computer-aided design (CAD) is influencing design methods, organisational structures, the division of work, e.g. between conceptual designers and detail designers,

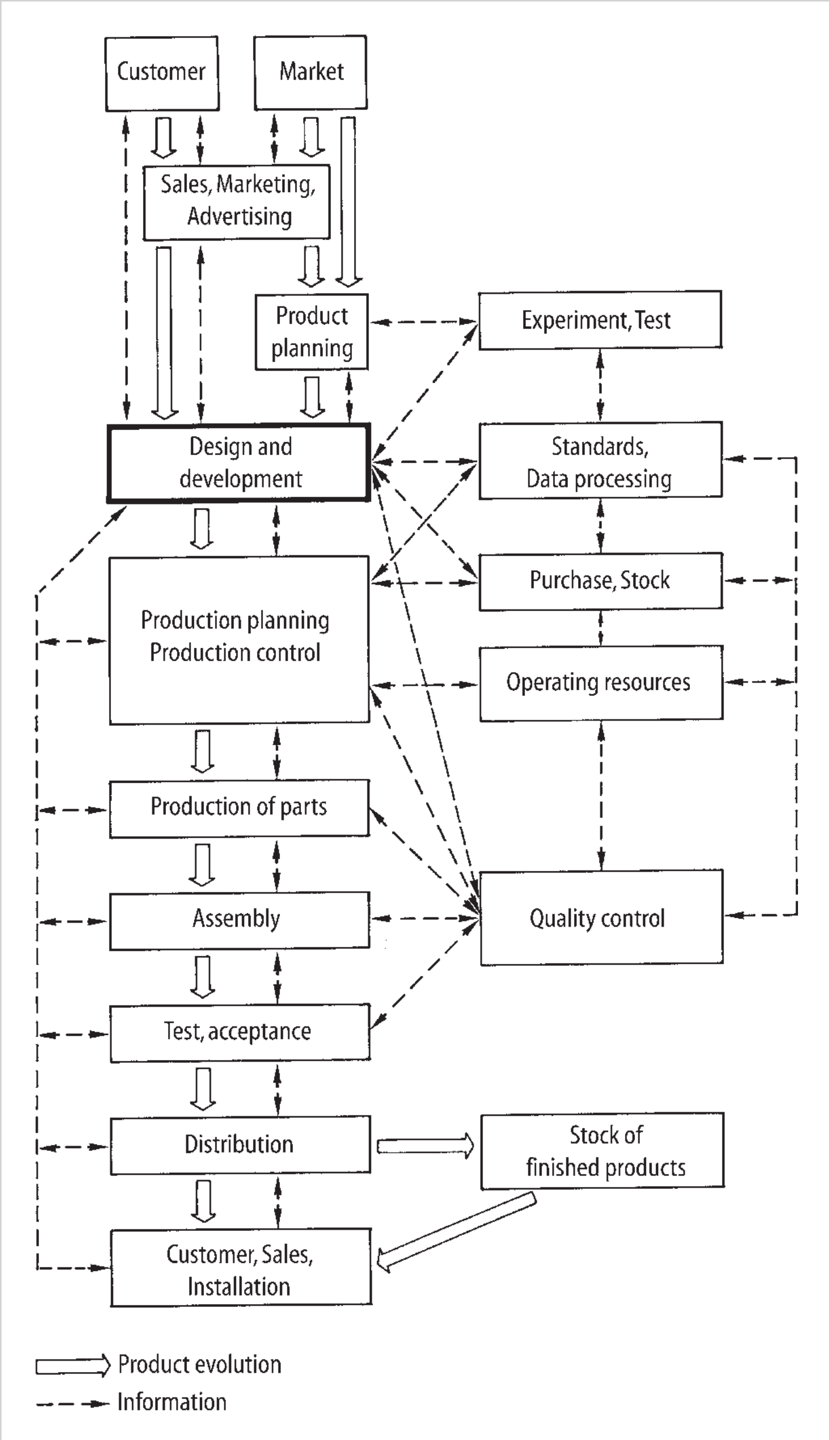


Figure 1.4. Information flows between departments

as well as the creativity and thought processes of individual designers (see Section 2.2). New staff, e.g. system managers, CAD specialists, etc., are being introduced into the design process. In the future, routine tasks such as variant designs will be largely undertaken by the computer, leaving designers free to concentrate on new designs and customer-specific one-off products. These tasks will be supported by computer tools that enhance the creativity, engineering knowledge and experience of designers. The development of knowledge-based systems (expert systems) [1.72, 1.108, 1.178, 1.183] and electronic component catalogues [1.19, 1.20, 1.53, 1.151, 1.183] will increase the ease with which information can be retrieved, including specific design data, details of standard components, information about existing products as well as their design processes and other design knowledge. These systems will also aid the analysis, optimisation and combination of solutions, but they will not replace designers. On the contrary, the decision-making abilities of designers will be even more crucial because of the very large number of solutions it will be possible to generate, and also because of the need to coordinate the inputs from the many specialists now required in modern multidisciplinary projects.

A further strong trend is for companies to concentrate their design and development activities on so-called core competences, and thus acting as system integrators, buying in assemblies and components as required from other companies (outsourcing). Designers therefore need the ability to assess and evaluate these outsourced items, even though they have not created these themselves. This critical assessment process is enhanced through broad technical knowledge, accumulated experience and a systematic use of evaluation procedures (see Section 3.3).

Computer-integrated manufacturing (CIM) has consequences for designers in terms of company organisation and information exchange. The system within a CIM structure makes better planning and control of the design process necessary and possible. The same holds true for *simultaneous engineering* (see Section 4.3 [1.13, 1.40, 1.188]), where development times are reduced by focusing on the flexible and partially parallel activities of product optimisation, production optimisation and quality optimisation. The trend is to bring production planning forward into the design process through the application of computers.

Apart from these developments that influence the working methods of designers, designers must increasingly take into account rapid technological developments (e.g. new production and assembly procedures, microelectronics and software) and new materials (e.g. composites, ceramics and recyclable materials). The integration of mechanical, electronic and software engineering (mechatronics) has led to many exciting product developments. Designers now have to give equal weight to these three aspects of modern products.

In summary, it can be concluded that there is already much pressure on designers and this pressure will increase further. This requires continuous further education for existing designers. However, the initial education of designers must take into account the many changes taking place [1.127, 1.187]. It is essential that future designers not only understand traditional science and engineering fundamentals (physics, chemistry, mathematics, mechanics, thermodynamics, fluid mechanics, electronics, electrical engineering, materials science, machine elements) but also specific domain knowledge (instrumentation, control, transmission technology, production technology, electrical drives, electronic controls). The education of future designers should include courses where they actually apply their design knowledge in order to solve design tasks. They also need specialist courses in design methodology, including CAD and CAE.

1.2 Necessity for Systematic Design

1.2.1 Requirements and the Need for Systematic Design

In view of the central responsibility of designers for the technical and economic properties of a product, and the commercial importance of timely and efficient product development, it is important to have a defined design procedure that finds good solutions. This procedure must be flexible and at the same time be capable of being planned, optimised and verified. Such a procedure, however, cannot be realised if the designers do not have the necessary domain knowledge and cannot work in a systematic way. Furthermore, the use of such a procedure should be encouraged and supported by the organisation.

Nowadays one distinguishes between design science and design methodology [1.90]. *Design science* uses scientific methods to analyse the structures of technical systems and their relationships with the environment. The aim is to derive rules for the development of these systems from the system elements and their relationships.

Design methodology, however, is a concrete course of action for the design of technical systems that derives its knowledge from design science and cognitive psychology, and from practical experience in different domains. It includes plans of action that link working steps and design phases according to content and organisation. These plans must be adapted in a flexible manner to the specific task at hand (see Chapter 4). It also includes strategies, rules and principles to achieve general and specific goals (see Chapter 7 and Chapters 9–11) as well as methods to solve individual design problems or partial tasks (see Chapters 3 and 6).

This is not meant to detract from the importance of *intuition* or experience; quite the contrary—the additional use of systematic procedures can only serve to increase the output and inventiveness of talented designers. Any logical and systematic approach, however exacting, involves a measure of intuition; that is, an inkling of the overall solution. No real success is likely without intuition.

Design methodology should therefore foster and guide the abilities of designers, encourage creativity, and at the same time drive home the need for objective evaluation of the results. Only in this way is it possible to raise the general standing of designers and the regard in which their work is held. Systematic procedures help to render designing comprehensible and also enable the subject to be taught. However, what is learned and recognised about design methodology should not be taken as dogma. Such procedures merely try to steer the efforts of designers from unconscious into conscious and more purposeful paths. As a result, when they collaborate with other engineers, designers will not merely be holding their own, but will be able to take the lead [1.130].

Systematic design provides an effective way to *rationalise* the design and production processes. In original design, an ordered and stepwise approach—even if this is on a partially abstract level—will provide solutions that can be used again. Structuring the problem and task makes it easier to recognise application possibilities for established solutions from previous projects and to use design catalogues. The stepwise concretisation of established solution principles makes it possible to select and optimise them at an early stage with a smaller amount of effort. The approach of developing size ranges and modular products is an important start to rationalisation in the design area, but is especially important for the production process (see Chapter 9).

A design methodology is also a prerequisite for flexible and continuous *computer support* of the design process using product models stored in the computer. Without this methodology it is not possible to: develop knowledge-based systems; use stored data and methods; link separate programs, especially geometric modellers with analysis programs; ensure the continuity of data flow; and link data from different company divisions (CIM, PDM). Systematic procedures also make it easier to divide the work between designers and computers in a meaningful way.

A rational approach must also cover the cost of computation and quality considerations. More accurate and speedy preliminary calculations with the help of better data are a necessity in the design field, as is the early recognition of weak points in a solution. All this calls for systematic processing of the design documentation.

A *design methodology*, therefore, must:

- allow a problem-directed approach; i.e. it must be applicable to every type of design activity, no matter which specialist field it involves
- foster inventiveness and understanding; i.e. facilitate the search for optimum solutions
- be compatible with the concepts, methods and findings of other disciplines
- not rely on finding solutions by chance
- facilitate the application of known solutions to related tasks
- be compatible with electronic data processing
- be easily taught and learned
- reflect the findings of cognitive psychology and modern management science; i.e. reduce workload, save time, prevent human error, and help to maintain active interest
- ease the planning and management of teamwork in an integrated and interdisciplinary product development process
- provide guidance for leaders of product development teams.

1.2.2 Historical Background

It is difficult to determine the origins of systematic design. Can we trace it back to Leonardo da Vinci? Anyone looking at the sketches of this early master must be surprised to see—and the modern systematist delights in discovering—the great extent to which Leonardo used systematic variation of possible solutions [1.118]. Right up to the industrial era, designing was closely associated with arts and crafts.

With the rise of mechanisation in the nineteenth century, as Redtenbacher [1.150] pointed out early on in his *Prinzipien der Mechanik und des Maschinenbaus* (Principles of Mechanics and of Machine Construction), attention became increasingly focused on a number of characteristics and principles that continue to be of great importance, namely: sufficient strength, sufficient stiffness, low wear, low friction, minimum use of materials, easy handling, easy assembly and maximum rationalisation.

Redtenbacher's pupil Reuleaux [1.152] developed these ideas but, in view of their often conflicting requirements, suggested that the assessment of their relative importance must be left to the intelligence and discretion of individual designers. They cannot be treated in a general way or be taught.

Important contributions to the development of engineering design were also made by Bach [1.11] and Riedler [1.153], who realised that the selection of materials, the choice of production methods and the provision of adequate strength are of equal importance and that they influence one another.

Rotscher [1.164] mentions the following essential characteristics of design: specified purpose, effective load paths, and efficient production and assembly. Loads should be conducted along the shortest paths, and if possible by axial forces rather than by bending moments. Longer load paths not only waste materials and increase costs but also require considerable changes in shape. Calculation and laying out must go hand-in-hand. Designers start with what they are given and with ready-made assemblies. As soon as possible, they should make scale drawings to ensure the correct spatial layout. Calculation can be used to obtain either rough estimates for the preliminary layout or precise values that are used to check the detail design.

Laudien [1.107], upon examining the load paths in machine parts, gave the following advice: for a rigid connection, join the parts in the direction of the load; if flexibility is required, join the parts along indirect load paths; do not make unnecessary provisions; do not over-specify; do not fulfil more demands than are required; save by simplification and economical construction.

Modern systematic ideas were pioneered by Erkens [1.46] in the 1920s. He insisted on a *step-by-step approach* based on *constant testing and evaluation*, and also on the *balancing of conflicting demands*, a process that must be continued until a network of ideas—the design—emerges.

A more comprehensive account of the "technique of design" has been presented by Wögerbauer [1.206], whose contribution we consider to be the origin of systematic design. He divides the *overall task* into *subsidiary tasks*, and these into operational and implementational tasks. He also examines (but fails to present in systematic form) the numerous interrelationships between the identifiable constraints designers must take into account. Wögerbauer himself does not proceed to a systematic elaboration of solutions. His systematic search starts with a solution discovered more or less intuitively and varied as comprehensively as possible in respect to the basic form, materials and method of production. The resulting profusion of possible solutions is then reduced by tests and evaluations, with cost being a crucial criterion. Wögerbauer's very comprehensive list of *characteristics* helps in the search for an optimum solution and also when testing and evaluating the results.

Franke [1.54] discovered a comprehensive structure for transmission systems using a logical–functional analogy based on elements with different physical effects (electrical, mechanical, hydraulic effects for identical logical functions guiding, coupling and separating). For this reason he is regarded as a representative of those working on the functional comparison of physically different solution elements. Rodenacker in particular used this analogical approach [1.155].

Though some need to improve and rationalise the design process was felt even before World War II, progress was impeded by the absence of a reliable means of representing abstract ideas and the widespread view that designing is a form of art, not a technical activity like any other. A period of staff shortages in the 1960s [1.190] created a strong impetus to adopt systematic thinking more widely. Important pioneers were Kesselring, Tschochner, Niemann, Matousek and Leyer. Their work continues to provide most useful suggestions for handling the individual phases and steps of systematic design.

Kesselring [1.98] first explained the basis of his method of successive approximations in 1942 (for a summary see [1.96, 1.97] and VDI Guideline 2225 [1.195]). Its salient feature is the evaluation of form variants according to *technical* and *economic criteria*. In his theory, he mentions five overlying principles:

- the principle of minimum production costs
- the principle of minimum space requirement
- the principle of minimum weight
- the principle of minimum losses
- the principle of optimum handling.

The design and optimisation of individual parts and simple technical artefacts is the aim of the theory of form design. It is characterised by the simultaneous application of physical and economic laws, and leads to a determination of the shape and dimensions of components and an appropriate choice of materials, production methods, etc. If selected optimisation characteristics are taken into account, the best solution can be found with the help of mathematical methods.

Tschochner [1.179] mentions four fundamental design factors, namely the *working principle*, the *material*, the *form* and the *size*. They are interconnected and dependent on the requirements, the number of units, costs, etc. Designers start from the solution principle, determine the other fundamental factors—material and form—and match them with the help of the chosen dimensions.

Niemann [1.121] starts out with a scale layout of the overall design, showing the main dimensions and the general arrangement. Next he divides the overall design into parts that can be developed in parallel. He proceeds from a *definition of the task* to a systematic *variation of possible solutions* and finally to a *critical and formal selection of the optimum solution*. These steps are in general agreement with those used in more recent methods. Niemann also draws attention to the then lack of methods for arriving at new solutions. He must be considered a pioneer of systematic design inasmuch as he consistently demanded and encouraged its development.

Matousek [1.112] lists four essential factors: *working principle*, *material*, *production* and *form* design, and then, following Wögerbauer [1.206], elaborates an overall working plan based on these four factors considered in the order given. He adds that, if the cost aspect is unsatisfactory, these factors have to be reexamined in an iterative manner.

Leyer [1.109] is mainly concerned with form design, for which he develops fundamental *guidelines* and *principles*. He distinguishes three main design phases. In the first, the working principle is laid down with the help of an idea, an invention, or established facts; the second phase is that of actual design; the third phase is that of implementation. His second phase is essentially that of embodiment; that is, layout and form design supported by calculations. During this phase, principles or rules have to be taken into account—for instance, the principle of constant wall thickness, the principle of lightweight construction, the principle of shortest load paths, and the principle of homogeneity. Leyer's rules of form design are so valuable because, in practice, failure is still far less frequently the result of bad working principles than of poor detail design.

These preliminary attempts made way for the intensive development of methods, mainly by university professors who had learnt the fundamentals of design by designing technical products of increasing complexity in industry before becoming professors. They realised that a greater reliance on physics, mathematics and information theory, and the use of systematic methods, were not only possible but, with the growing division of labour, quite indispensable. Needless to say, these developments were strongly affected by the requirements of the particular industries in which they originated. Most came from precision, power transmission and electromechanical engineering, in which systematic relationships are more obvious than in heavy engineering.

Hansen and other members of the *Ilmenau School* (Bischoff, Bock) first put forward their systematic design proposals in the early 1950s [1.21, 1.25, 1.78]. Hansen presented a more comprehensive design system in the second edition of his standard work published in 1965 [1.77].

Hansen's approach is defined in a so-called *basic system*. The four working steps in this approach are applied in the same way in conceptual, embodiment and detail design. Hansen begins with the analysis, critique, and specification of the task, which leads to the *basic principle* of the development (the crux of the task). The basic principle encompasses the overall function that has been derived from the task, the prevailing conditions, as well as the required measures. The overall function (the goal and the constraints) and the context (elements and properties) constitute the crux of the task together with the given constraints.

The second working step is a systematic search for solution elements and their combination into *working means* and *working principles*.

Hansen attaches great importance to the third step, in which any shortcomings of the developed working means are analysed with respect to their properties and quality characteristics, and then, if necessary, improved.

In the fourth and last step, these improved working means are evaluated to determine the optimum working means for the task.

In 1974 Hansen published another work, entitled *Konstruktionswissenschaft* (Science of Design) [1.76]. The book is more concerned with theoretical fundamentals than with rules of practical design.

Similarly, Müller [1.116] in his *Grundlagen der systematischen Heuristik* (Fundamentals of Systematic Heuristics) presents a theoretical and abstract picture of the design process. This book offers essential foundations of design science. Further important publications are [1.114, 1.115, 1.117].

After Hansen, it is Rodenacker [1.155–1.157] who became preeminent by developing an original design method. His approach is characterised by developing the required overall *working interrelationship* by defining in sequence the *logical*, *physical* and *embodiment* relationships. He emphasises the recognition and suppression of disturbing influences and failures as early as possible during formulation of the physical process; the adoption of a general selection strategy from simple to complex; and the evaluation of all parameters of the technical system against the criteria *quantity*, *quality* and *cost*. Other characteristics of his method are the emphasis on logical function structures based on *binary logic* (connecting and separating), and on a conceptual design stage based on the recognition that product optimisation can only take place once a suitable solution principle has been found. The most important aspect of Rodenacker's systematic design approach is undoubtedly his emphasis on establishing the physical process. Based on this, he not only deals with the systematic processing of concrete design tasks, but also with a methodology for inventing new technical systems. For the latter he starts with the question: For what new application can a known physical effect be used? He then searches systematically to discover completely new solutions.

In addition to the methods we have been describing, there is a view that a one-sided emphasis on discursive methods does not present the complete picture. Thus Wächtler [1.199, 1.200] argues, by analogy with cybernetic concepts such as control and learning, that creative design is the most complex form of the "learning process". Learning represents a higher form of control, one that involves not only quantitative changes at constant quality (rules), but also changes in the quality itself.

What matters is that, for the purpose of optimisation, the design process should be treated, not statically, but dynamically as a control process in which the information feedback must be repeated until the information content has reached the level at which the optimum solution can be found. The learning process thus keeps increasing the level of information and hence facilitates the search for a solution.

The systematic design methods of Leyer, Hansen, Rodenacker and Wächtler are still being applied today, having been integrated into the more recent developments in design methodology.

1.2.3 Current Methods

1. Systems Theory

In socio–economic–technical processes, procedures and methods of *systems theory* are becoming increasingly important. The interdisciplinary science of systems theory uses special methods, procedures and aids for the analysis, planning, selection and optimum design of complex systems [1.14–1.16, 1.23, 1.29, 1.30, 1.143, 1.208].

Technical artefacts, including the products of light and heavy engineering industry, are artificial, concrete and mostly dynamic systems consisting of sets of ordered elements, interrelated by virtue of their properties. A system is also characterised by the fact that it has a boundary which cuts across its links with the environment (see Figure 1.5). These links determine the external behaviour of the system, so that it is possible to define a function expressing the relationship between inputs and outputs, and hence changes in the magnitudes of the system variables (see Section 2.1.3).

From the idea that technical artefacts can be represented as systems, it was a short step to the application of systems theory to the design process, the more so as the objectives of systems theory correspond very largely to the expectations we have of a good design method, as specified at the beginning of this chapter [1.16]. The systems approach reflects the general appreciation that complex problems are best tackled in fixed steps, each involving analysis and synthesis (see Section 2.2.5).

Figure 1.6 shows the steps of the systems approach. The first of these is the gathering of information about the system under consideration by means of market analyses, trend studies or known requirements. In general this step can be called problem analysis. The aim here is the clear formulation of the problem (or subproblem) to be solved, which is the actual starting point for the development of the system. In the second step, or perhaps even during the first step, a programme is drawn up in order to give formal expression to the goals of the system (problem formulation). Such goals provide important criteria for the subsequent evaluation of solution variants and hence for the discovery of the optimum solution. Several solution variants are then synthesised on the basis of the information acquired during the first two steps.

Before these variants can be evaluated, the performance of each must be analysed for its properties and behaviour. In the evaluation that follows, the performance of each variant is compared with the original goals, and on the basis of this a decision is made and the optimum system selected. Finally, information is given out in the form of system implementation plans. As Figure 1.6 shows, the steps do not always lead straight to the final goal, so that iterative procedures may be needed. Built-in decision steps facilitate this optimisation process, which constitutes a transformation of information.

In a systems theory process model [1.23, 1.52], the steps repeat themselves in so-called life cycle phases of the system in which the chronological progression of a system goes from abstract to concrete (see Figure 1.7).

2. Value Analysis

The main aim of *Value Analysis*, as described in DIN 69910 [1.37, 1.66, 1.196–1.198], is to reduce cost (see Chapter 11). To that end a systematic overall approach is proposed which is applicable, in particular, to the further development of existing

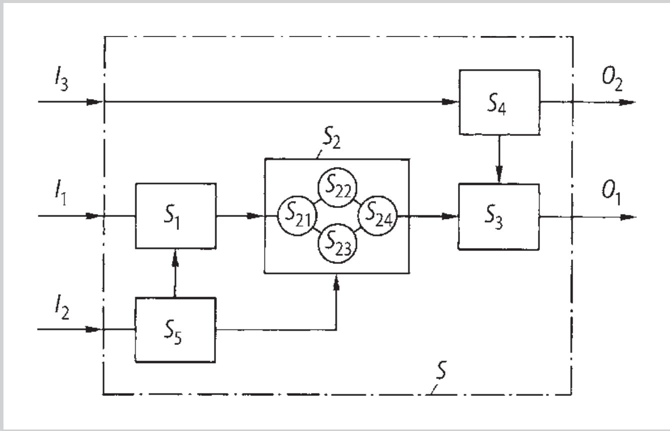


Figure 1.5. Structure of a system. S: system boundary; S_1–S_5: subsystems of S; S_{21}–S_{24}: subsystems or elements of S_2; I_1–I_3: inputs; O_1–O_2 outputs

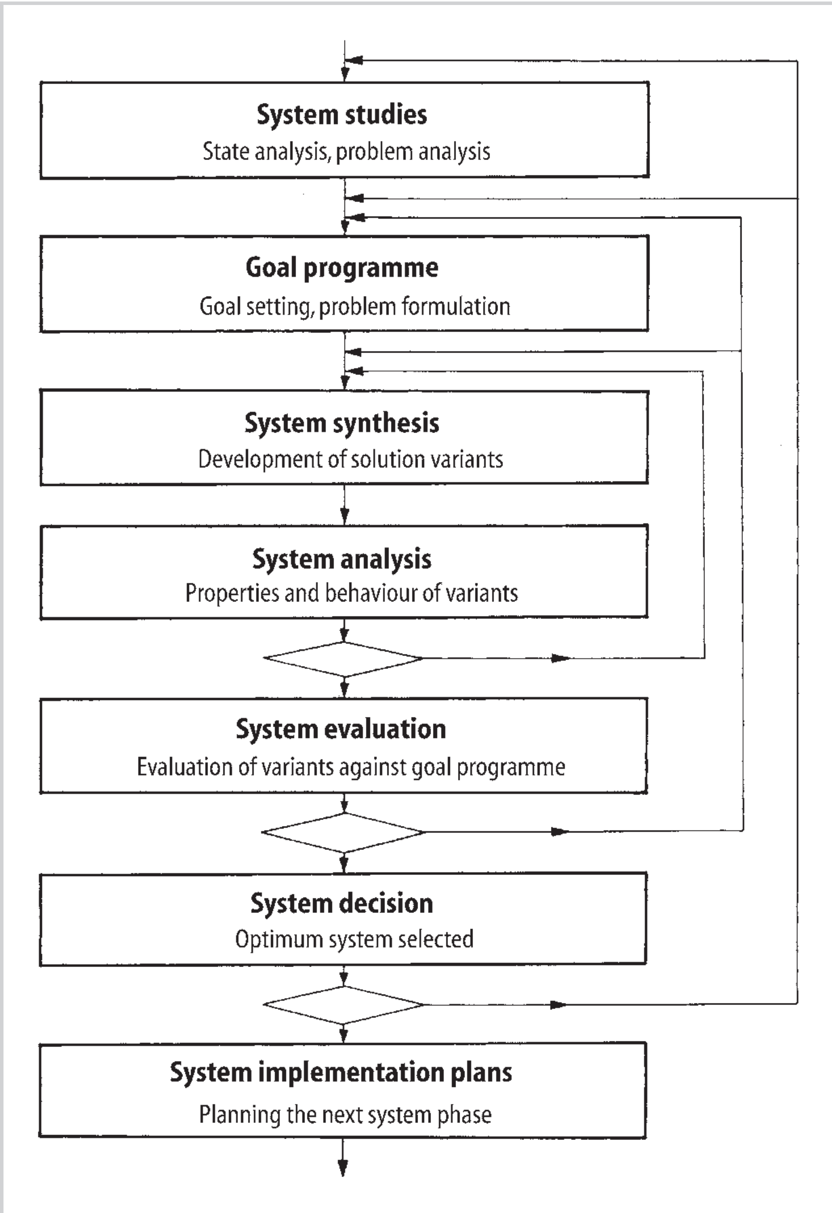


Figure 1.6. Steps of the systems approach

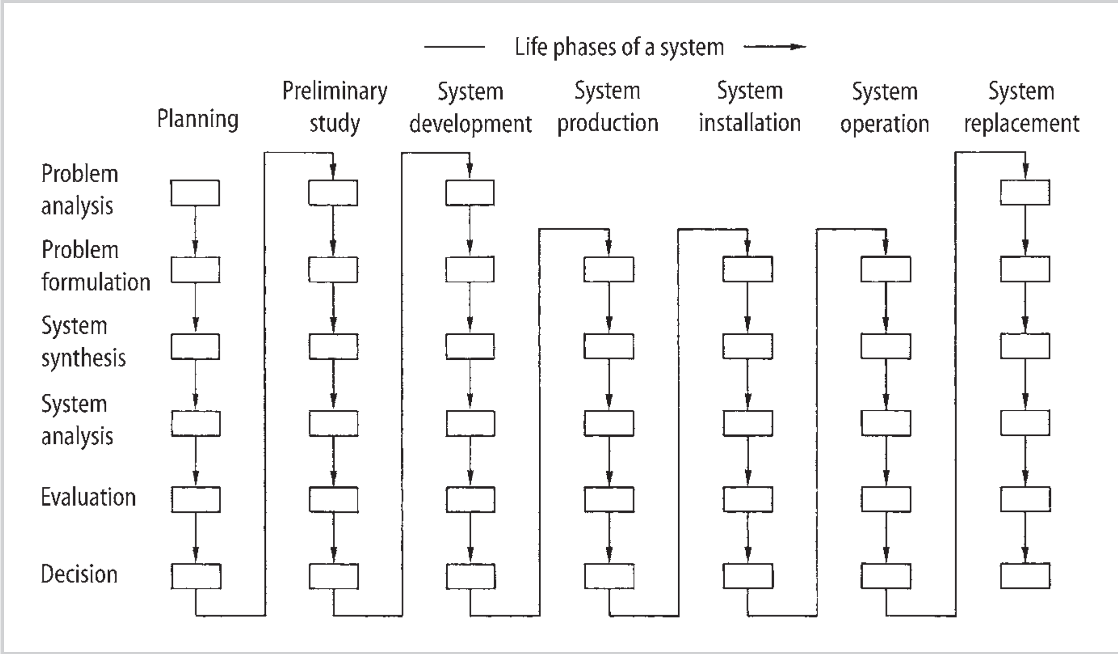


Figure 1.7. Model of the systems approach. After [1.23, 1.52]

Prepare project

- Assemble team
- Define scope of Value Analysis
- Define organisation and procedure

Analyse actual state

- Recognise functions
- Determine function costs

Determine target state

- Define target functions
- Identify additional requirements
- Match target costs with target functions

Develop solution ideas

- Collect existing ideas
- Search for new ideas

Determine solutions

- Evaluate ideas
- Develop ideas into solutions
- Evaluate and decide upon solutions

Realise solutions

- Detail chosen solutions
- Plan their implementation

Figure 1.8. Basic working steps of Value Analysis. After DIN 69910

products. Figure 1.8 shows the basic working steps of Value Analysis. In general, a start is made with an existing design, which is analysed with respect to the required functions and costs. Solution ideas are then proposed to meet the new targets. Because of its emphasis on functions and the stepwise search for better solutions, Value Analysis has much in common with systematic design.

Various methods are available to estimate costs and assess cost breakdowns (see Chapter 11). Teamwork is essential. Good communication between staff in sales, purchase, design, production and cost estimation (the Value Analysis team) ensures a holistic view of the requirements, embodiment design, materials selection, production processes, storage requirements, standards and marketing.

A further essential aspect is the division of the required overall function into subfunctions in the order of descending complexity along with their allocation to function carriers (assemblies, individual components). The costs of fulfilling all of the functions up to and including the overall function can be estimated from the costs calculated for the individual components. Such "function costs" can then provide the basis for evaluating the concepts or embodiment variants. The aim is to minimise these function costs and where possible eliminate functions that are not really necessary.

It has been suggested that the application of the Value Analysis method should not be left until after the layout and detail drawings have been finalised, but should be started during conceptual design in order to "design in" value [1.65]. In this way, Value Analysis approaches the goals of systematic design.

3. Design Methods

VDI Guideline 2222 [1.192, 1.193] defines an approach and individual methods for the conceptual design of technical products and is therefore particularly suitable for the development of new products. The more recent *VDI Guideline 2221* [1.191] (English translation: [1.186]) proposes a generic approach to the design of technical systems and products, emphasising the general applicability of the approach in the fields of mechanical, precision, control, software and process engineering. The approach (see Figure 1.9) includes seven basic working steps that accord with the fundamentals of technical systems (see Section 2.1) and company strategy (see Chapter 4). Both guidelines have been developed by a VDI Committee comprising senior designers from industry and many of the previously mentioned design methodologists from the former West Germany. Because the aim is for general applicability, the design process has been only roughly structured, thus permitting product-specific and company-specific variations. Figure 1.9 should therefore be regarded as a guideline to which detailed working procedures can be assigned. Special emphasis is placed on the iterative nature of the approach and the sequence of the steps must not be considered rigid. Some steps might be omitted, and others repeated frequently. Such flexibility is in accordance with practical design experience and is very important for the application of all design methods.

The design methodologists and senior designers from industry who collaborated to produce these VDI Guidelines often represented different schools of thought or had developed their own design methods. Several contributions to design methodology were made by colleagues in other countries. In this book, references are made to all of these many inputs when the individual methods and procedures are discussed in detail.

A comprehensive overview of the international design teaching and research activities since 1981 can be found in the proceedings of the ICED conference series (International Conference on Engineering Design) [1.148].

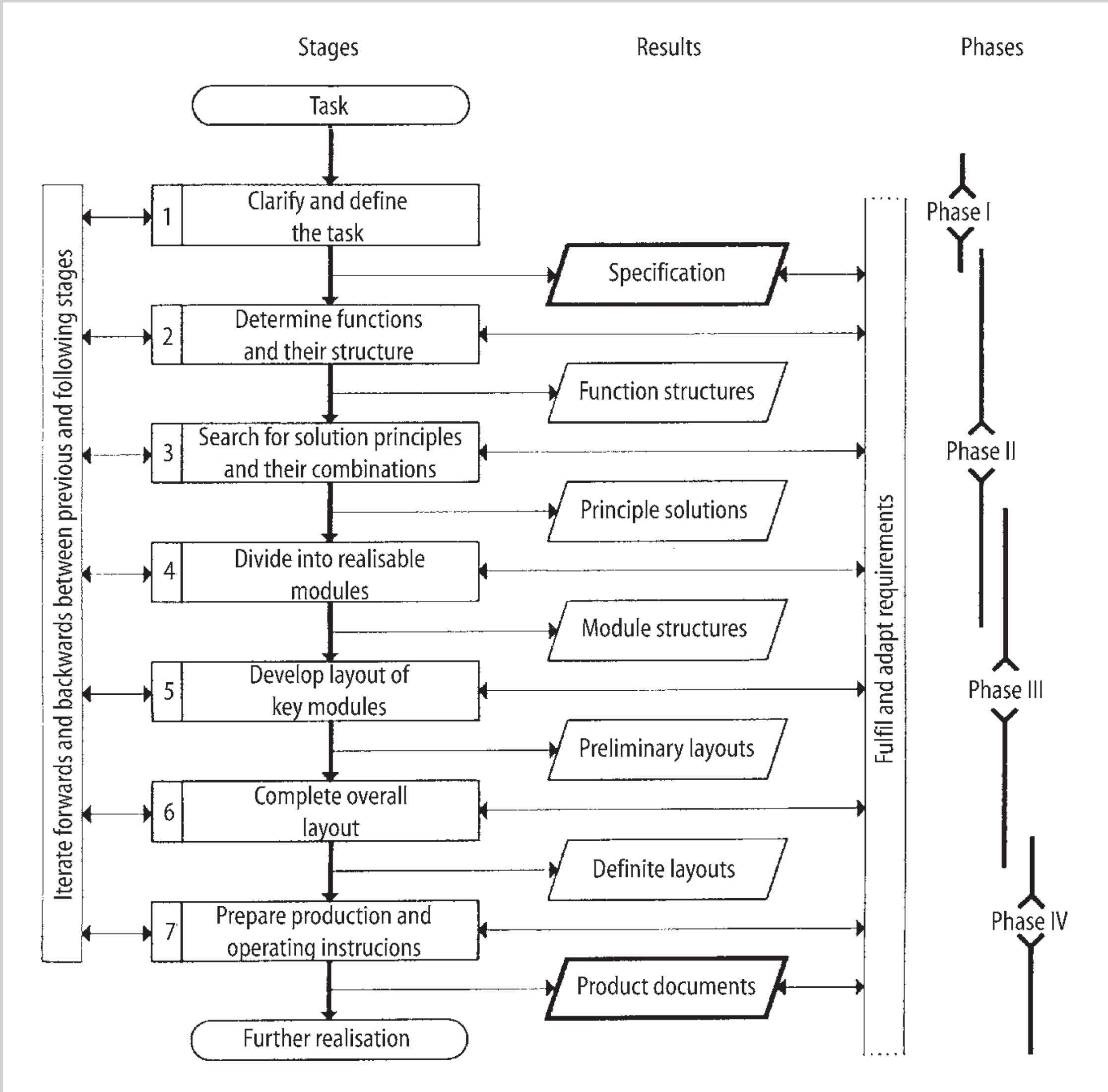


Figure 1.9. General approach to design. After [1.191]

In Table 1.1, the main publications on design methodology are given in chronological order. This table replaces and extends in a more compact form the individual efforts and achievements that were described in Chapter 1 of the second English edition of this book. Further contributions from the authors listed in the table can be seen from their entries in the list of references at the end of the book.

1.2.4 Aims and Objectives of this Book

On closer examination the methods we have been describing have been strongly influenced by their authors' specialist fields. They nevertheless resemble one another far more closely than the various concepts and terms might suggest. VDI Guidelines 2222 and 2221 confirm these resemblances as they were developed in collaboration with a wide range of experienced contributors.

Based on our experience in the heavy machinery industry and railway and automotive engineering and many years spent in engineering design education at

Table 1.1. Chronological overview of the development of design methodology

Year	Author	Theme/Title	Country	Literature
1953	Bischoff, Hansen	Rationelles Konstruieren	DDR	[1.21]
1955	Bock	Konstruktionssystematik—die Methode der ordnenden Gesichtspunkte	DDR	[1.25]
1956	Hansen	Konstruktionssystematik	DDR	[1.78]
1963	Pahl	Konstruktionstechnik im thermischen Maschinenbau	DE	[1.131]
1966	Dixon	Design Engineering: Inventiveness, Analysis and Decision-Making	USA	[1.39]
1967	Harrisberger	Engineermanship	USA	[1.79]
1968	Roth	Systematik der Maschinen und ihrer mechanischen elementaren Funktionen	DE	[1.163]
1969	Glegg	The Design of the Design, The Development of Design, The Science of Design	GB	[1.68–1.70]
	Tribus	Rational Descriptions, Decisions and Design	USA	[1.177]
1970	Beitz	Systemtechnik im Ingenieurbereich	DE	[1.16]
	Gregory	Creativity in Engineering	GB	[1.71]
	Pahl	Wege zur Lösungsfindung	DE	[1.129]
	Rodenacker	Methodisches Konstruieren (4th Edition 1991)	DE	[1.155]
1971	French	Conceptual Design for Engineers, 1st Edition (3rd Edition 1999)	GB	[1.58]
1972	Pahl, Beitz	Series of articles „Für die Konstruktionspraxis“ (1972–1974)	DE	[1.142]
1973	Altschuller	Erfinden: Anleitung für Neuerer und Erfinder	USSR	[1.5]
	VDI	VDI-Richtlinie 2222, Blatt 1 (Entwurf): Konzipieren technischer Produkte	DE	[1.192]
1974	Adams	Conceptual Blockbusting: A Guide to Better Ideas	USA	[1.1]
1976	Hennig	Methodik der Verarbeitungsmaschinen	DDR	[1.82]
1977	Flursheim	Engineering Design Interfaces	GB	[1.49, 1.50]
	Ostrofsky	Design, Planning and Development Methodology	USA	[1.126]
	Pahl, Beitz	Konstruktionslehre, 1st Edition (6th Edition 2005)	DE	[1.134]
	VDI	VDI-Richtlinie 2222 Blatt 1: Konzipieren technischer Produkte	DE	[1.192]
1978	Rugenstein	Arbeitsblätter Konstruktionstechnik	DDR	[1.165]
1979	Frick	Integration der industriellen Formgestaltung in den Erzeugnis-Entwicklungsprozess, Arbeiten zum Industrial Design	DDR	[1.60–1.62]
	Klose	Zur Entwicklung einer speicherunterstützten Konstruktion von Maschinen unter Wieder-verwendung von Baugruppen	DDR	[1.99, 1.100]
	Polovnikin	Untersuchung und Entwicklung von Konstruktionsmethoden	USSR	[1.146, 1.147]
1981	Gierse	Wertanalyse und Konstruktionsmethodik in der Produktentwicklung	DE	[1.67]
	Kozma, Straub (Pahl/Beitz)	Hungarian translation of Pahl/Beitz Engineering Design	H	[1.141]
	Nadler	The Planning and Design Approach	USA	[1.119]

Table 1.1. (continued)

Year	Author	Theme/Title	Country	Literature
	Proceedings of ICED by Hubka	WDK Series biannually from 1981 to 2001; Design Society Series from 2003	CH	[1.148]
	Schregenberger	Methodenbewusstes Problemlösen	CH	[1.170]
1982	Dietrych, Rugenstein	Einführung in die Konstruktions-wissenschaft	PL/D	[1.36]
	Roth	Konstruieren mit Konstruktions-katalogen, 1st Edition (3rd Edition 2001)	DE	[1.160, 1.161], [1.162]
	VDI	VDI-Richtlinie 2222 Blatt 2: Erstellung und Anwendung von Konstruktionskatalogen	DE	[1.193]
1983	Andreasen et al.	Design for Assembly	DK	[1.8]
	Höhne	Struktursynthese und Variationstechnik beim Konstruieren	DDR	[1.84]
1984	Hawkes, Abinett	The Engineering Design Process	GB	[1.80]
	Altschuller	Erfinden – Wege zur Lösung technischer Probleme	USSR	[1.4]
	Hubka	Theorie technischer Systeme	CH	[1.86, 1.87]
	Walczack (Pahl/Beitz)	Polish translation of Pahl/Beitz Engineering Design	PL	[1.139]
	Wallace (Pahl/Beitz)	English translation of Pahl/Beitz Engineering Design, 1st Edition (3rd Edition 2006)	GB	[1.140]
	Yoshikawa	Automation in Thinking in Design	J	[1.207]
1985	Archer	The Implications for the Study for Design Methods of Recent Development in Neighbouring Disciplines	GB	[1.10]
	Ehrlenspiel, Lindemann	Kostengünstig Entwickeln und Konstruieren	DE	[1.41, 1.43]
	Franke	Konstruktionsmethodik und Konstruktions-praxis—eine kritische Betrachtung	DE	[1.51]
	Koller	Konstruktionslehre für den Maschinenbau. Grundlagen, Arbeitsschritte, Prinziplösungen. (3rd Edition 1994)	DE	[1.101, 1.102], [1.103, 1.104]
	van den Kroonenberg	Design Methodology as a Condition for Computer-Aided Design	NL	[1.185]
1986	Odrin	Morphologische Synthese von Systemen	USSR	[1.122]
	Altschuller	Theory of Inventive Problem Solving	USSR	[1.2, 1.3]
	Taguchi	Introduction of Quality Engineering	J	[1.175]
1987	Andreasen, Hein	Integrated Product Development	DK	[1.7]
	Erlenspiel, Figel	Application of Expert Systems in Machine Design	DE	[1.42]
	Gasparski	On Design Differently	PL	[1.63]
	Hales	Analysis of the Engineering Design Process in an Industrial Context, Managing Engineering Design	GB	[1.73–1.75]
	Schlottmann	Konstruktionslehre	DDR	[1.169]
	VDI/Wallace	VDI Design Handbook 2221: Systematic Approach to the Design of Technical Systems and Products. English translation	DE/GB	[1.186]
	Wallace, Hales	Detailed Analysis of an Engineering Design Project	GB	[1.203]
1988	Dixon	On Research Methodology—Towards A Scientific Theory of Engineering Design	USA	[1.38]

Table 1.1. (continued)

Year	Author	Theme/Title	Country	Literature
	French	Form, Structure and Mechanism, Invention and Evolution	GB	[1.57, 1.58]
	Hubka, Eder	Theory of Technical Systems—A Total Concept Theory for Engineering Design	CH/CA	[1.88, 1.89]
	Jakobsen	Functional Requirements in the Design Process	N	[1.92]
	Suh	The Principles of Design, Axiomatic Design	USA	[1.173, 1.174]
	Ullmann, Stauffer, Dietterich	A Model of the Mechanical Design Process Based on Empirical Data	USA	[1.182]
	Winner, Pennell, et al.	The Role of Concurrent Engineering in Weapon Acquisition	USA	[1.205]
1989	Cross	Engineering Design Methods	GB	[1.33]
	De Boer	Decision Methods and Techniques	NL	[1.35]
	Elmaragh, Seering, Ullmann	Design Theory and Methodology	USA	[1.45]
	Jung	Funktionale Gestaltbildung—Gestaltende Konstruktionslehre für Vorrichtungen, Geräte, Instrumente und Maschinen	DE	[1.93, 1.94]
	Pahl/Beitz	Chinese translation of Pahl/Beitz Engineering Design	PRC	[1.138]
	Ulrich, Seering	Synthesis of Schematic Description in Mechanical Design	USA	[1.184]
1990	Birkhofer	Von der Produktidee zum Produkt—Eine kritische Betrachtung zur Auswahl und Bewertung in der Konstruktion	DE	[1.17, 1.18]
	Konttinnen (Pahl/Beitz)	Finnish translation of Pahl/Beitz Engineering Design	FIN	[1.137]
	Kostelic	Design for Quality	YU	[1.105]
	Müller	Arbeitsmethoden der Technikwissenschaften—Systematik, Heuristik, Kreativität	DDR	[1.114]
	Pighini	Methodological Design of Machine Elements	I	[1.145]
	Pugh	Total Design; Integrated Methods for Successful Product Engineering	GB	[1.149]
	Rinderle	Design Theory and Methodology	USA	[1.154]
	Roozenburg, Eekels	Evaluation and Decision in Design	NL	[1.158, 1.159]
1991	Andreasen	Methodical Design Frame by New Procedures	DK	[1.6]
	Bjärnemo	Evaluation and Decision Techniques in the Engineering Design Process	S	[1.22]
	Boothroyd, Dieter	Assembly Automation and Product Design	USA	[1.26]
	Clark, Fujimoto	Product Development Performance: Strategy, Organisation and Management	USA	[1.31]
	Flemming	Die Bedeutung der Bauweisen für die Konstruktion	CH	[1.47, 1.48]
	Hongo, Nakajima	Relevant Features of the Decade 1981–1991 of the Theories of Design in Japan	J	[1.85]

Table 1.1. (continued)

Year	Author	Theme/Title	Country	Literature
	Kannapan, Marshek	Design Synthetic Reasoning: A Methodology for Mechanical Design	USA	[1.95]
	Stauffer (ed)	Design Theory and Methodology	USA	[1.172]
	Walton	Engineering Design: From Art to Practice	USA	[1.204]
1992	O'Grady, Young	Constraint Nets for Life Cycle: Concurrent Engineering	USA	[1.123]
	Seeger	Integration von Industrial Design in das methodische Konstruieren	DE	[1.171]
	Ullmann	The Mechanical Design Process	USA	[1.180, 1.181]
1993	Breiing, Flemming	Theorie und Methoden des Konstruierens	CH	[1.28]
	Linde, Hill	Erfolgreich Erfinden. Widerspruchsorientierte Innovationsstrategie	DE	[1.110]
	Miller	Concurrent Engineering Design	USA	[1.113]
	VDI	VDI-Richtlinie 2221: Methodik zum Entwickeln und Konstruieren technischer Systeme und Produkte	DE	[1.191]
1994	Clausing	Total Quality Development	USA	[1.32]
	Blessing	A Process-Based Approach to Computer-Supported Engineering Design	GB	[1.24]
	Pahl (Editor)	Psychologische und pädagogische Fragen beim methodischen Konstruieren	DE	[1.127]
1995	Ehrlenspiel	Integrierte Produktentwicklung	DE	[1.40]
	Pahl/Beitz	Japanese translation of Pahl/Beitz Engineering Design	J	[1.136]
	Wallace, Blessing; Bauert (Pahl/Beitz)	English translation of Pahl/Beitz Engineering Design, 2nd Edition	GB	[1.135]
1996	Bralla	Design for Excellence	USA	[1.27]
	Cross, Christiaans, Dorst	Analysing Design Activity	NL	[1.34]
	Hazelrigg	Systems Engineering: An Approach to Information-Based Design	USA	[1.81]
	Waldron, Waldron	Mechanical Design: Theory and Methodology	USA	[1.202]
1997	Frey, Rivin, Hatamura	Introduction of TRIZ in Japan	J	[1.59]
	Magrab	Integrated Product and Process Design and Development	USA	[1.111]
1998	Frankenberger, Badke-Schaub, Birkhofer	Konstrukteure als wichtigster Faktor einer erfolgreichen Produktentwicklung	DE	[1.55]
	Hyman	Fundamentals of Engineering Design	USA	[1.91]
	Pahl/Beitz	Korean translation of Pahl/Beitz Engineering Design	KR	[1.133]
	Terninko, Zusman, Zlotin, Herb (ed)	Systematic Innovation: An Introduction to TRIZ	USA	[1.176]
1999	Pahl	Denk- und Handlungsweisen beim Konstruieren	DE	[1.128]
	Samuel, Weir	Introduction to Engineering Design	AU	[1.168]

Table 1.1. (continued)

Year	Author	Theme/Title	Country	Literature
	VDI	VDI-Richtlinie 2223 (Entwurf): Methodisches Entwerfen technischer Produkte	DE	[1.194]
2000	Pahl/Beitz	Portuguese translation of Pahl/Beitz Engineering Design	BR	[1.132]
2001	Antonsson, Cagan	Formal Engineering Design Synthesis	USA	[1.9]
	Gausemeyer, Ebbesmeyer, Kallmeyer	Produktinnovation mit strategischer Planung	DE	[1.64]
	Kroll, Condoor, Jansson	Innovative Conceptual Design: Parameter Analysis	USA	[1.106]
2002	Sachse	Entwurfsdenken und Darstellungshandeln, Verfestigung von Gedanken beim Konzipieren	DE	[1.166]
	Eigner, Stelzer	Produktdatenmanagement-Systeme	DE	[1.44]
	Neudörfer	Konstruieren sicherheitsgerechter Produkte	DE	[1.120]
	Orloff	Grundlagen der klassischen TRIZ	DE	[1.125]
	Wagner	Wegweiser für Erfinder	DE	[1.201]

the undergraduate and graduate levels, this book sets out a comprehensive design methodology for all phases of the product planning, design and development processes for technical systems. Most of the arguments are elaborations of a seminal series of papers published by the authors Pahl and Beitz [1.142] and previous editions of this book. It should be emphasised that between the publication of the first German edition of the book in 1977 and the latest edition, none of the statements had to be dropped because they were outdated.

As before, although our own approach to design does not claim to be the final word on the subject it tries to:

- be useful in design practice and design education
- provide a "toolbox" of design methods presented in a compatible way without expressing a particular school of thought or including short-lived trends
- emphasise the importance of design fundamentals, principles and guidelines at a time when more and more products are designed with the help of computers and many assemblies and components are outsourced
- serve as a guide to help designers and design leaders manage successful product development irrespective of the organisational structure (project management, however, is not the focus of this book).

We hope that this systematic approach to engineering design may serve as an introduction and springboard for the learner, as a help and illustration for the teacher, and as a source of information and further learning for the practitioner. It is important to realise that the methods and guidelines presented here underpin successful product development and product improvement.

Readers who are familiar with the application of generally applicable design methods and the fundamentals of systematic design can jump to Chapter 5 and start directly with the systematic approach to product development, returning to the fundamentals described in Chapters 2–4 when necessary. However, it is extremely important that students and novices build a solid foundation and do not ignore these early chapters.

2 Fundamentals

To develop an approach to design that can serve as a strategy for the development of solutions, we must first examine the fundamentals of technical systems and procedures along with the prerequisites for computer support. Only when that has been done is it possible to make detailed recommendations for design work.

2.1 Fundamentals of Technical Systems

2.1.1 Systems, Plant, Equipment, Machines, Assemblies and Components

Technical tasks are performed with the help of *technical artefacts* that include plant, equipment, machines, assemblies and components, listed here in approximate order of their complexity. These terms may not have identical uses in different fields. Thus, a piece of equipment (reactor, evaporator) is sometimes considered to be more complex than a plant, and artefacts described as "plant" in certain fields may be described as "machines" in others. A machine consists of assemblies and components. Control equipment is used in plant and machines alike and may also be made up of assemblies and components, and perhaps even of small machines. The variation in use of these terms reflect historical developments and application areas. There are attempts to define standards in which energy-transforming technical artefacts are referred to as machines, material-transforming artefacts as apparatus and signal-transforming artefacts as devices. It is evident that a clear division on the basis of these characteristics is not always possible and that the current terminology is not ideal.

There is much to be said for Hubka's suggestion [2.22–2.24] that technical artefacts should be treated as *systems* connected to the environment by means of *inputs* and *outputs*. A system can be divided into subsystems. What belongs to a particular system is determined by the *system boundary*. The inputs and outputs cross the system boundary (see Section 1.2.3). With this approach, it is possible to define appropriate systems at every stage of abstraction, analysis or classification. As a rule, such systems are parts of larger, superior systems.

A concrete example is the combined coupling shown in Figure 2.1. It can be considered as a system "coupling" which, within a machine, or when joining two

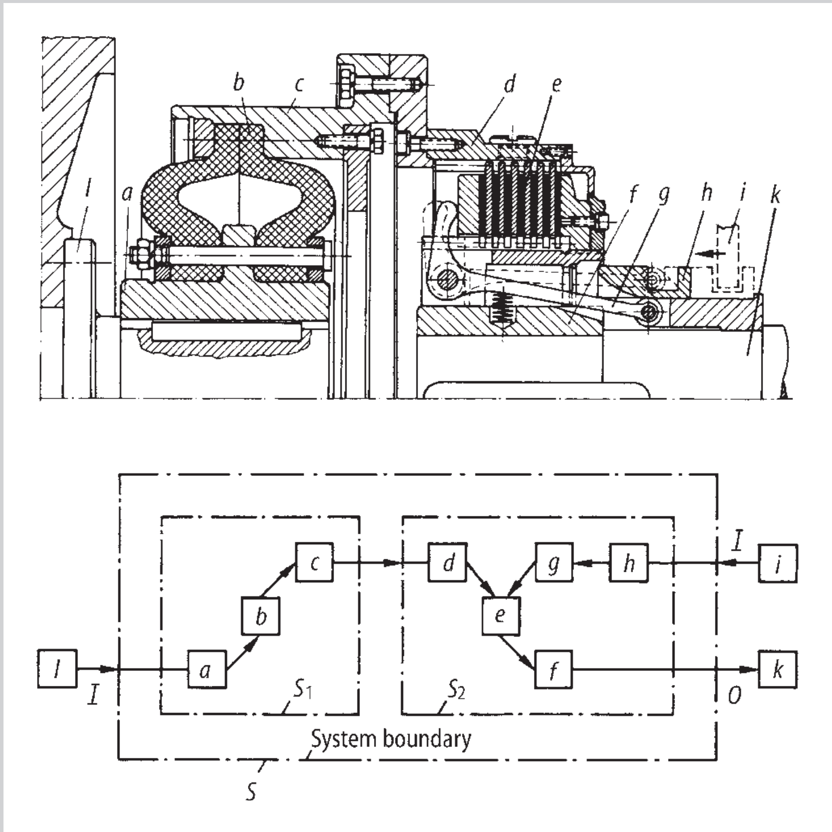


Figure 2.1. System "coupling": $a \ldots h$ system elements; $i \ldots l$ connecting elements, S overall system; S_1 subsystem "flexible coupling"; S_2 subsystem "clutch"; I inputs; O outputs

machines, can be considered to be an assembly. This coupling assembly can be treated as two *subsystems*—a "flexible coupling" and a "clutch". Each subsystem can, in turn, be subdivided into system elements, in this case components.

The system depicted in Figure 2.1 is based on its mechanical construction, referred to as the *construction structure*, see Figure 2.13. It is, however, equally possible to consider it in terms of its functions (see Section 2.1.3). In that case, the total system "coupling" can be split up into the subsystems "damping" and "clutching"; the second subsystem into the further subsystems "changing clutch operating force into normal force" and "transferring torque".

For example, the system element g could be treated as a subsystem whose *function* is to convert the actuating force into a larger normal force acting on the friction surface, and through its flexibility provide some equalisation of the wear.

Which viewpoint is used to divide the system depends on the intended purpose of the division. Common viewpoints are:

- Function: used to identify or describe the functional relationships
- Assembly: used to plan assembly operations
- Production: used to facilitate production and production planning.

Depending on their use, any number of such subdivisions may be made. Designers have to establish particular systems for particular purposes, and must specify their various inputs and outputs and fix their boundaries. In doing this, they may use what terminology they prefer or is customary in their particular field.

2.1.2 Conversion of Energy, Material and Signals

One encounters matter in many shapes and forms. Its natural form, or the form imposed upon it, provides information about its possible uses. Matter without form is inconceivable—form is a primary source of information about the state of matter. With the development of physics, the concept of force became essential. Force was conceived as being the means by which the motion of matter was changed. Ultimately this process was explained in terms of energy. The theory of relativity postulated the equivalence of energy and matter. Weizsacker [2.61] lists energy, matter and information as basic concepts. If change or flow is involved, time must be introduced as a fundamental quantity. Only by reference to time does the physical event in question become comprehensible, and can the interplay of energy, matter and information be adequately described.

In the technical sphere the previous terminology is usually linked to concrete physical or technical representations. *Energy* is often specified by its manifest form. We speak of, say, mechanical, electrical or optical energy. For matter, it is usual to substitute *material* with such properties as weight, colour, condition, etc. The general concept of information is generally given more concrete expression by means of the term *signal*—that is, the physical form in which the information is conveyed. Information exchanged between people is often called a message [2.20].

The analysis of technical systems—plant, equipment, machine, device, assembly or component—makes it clear that all of them involve technical processes in which energy, material and signals are channelled and converted. Such conversions of energy, material and signals have been analysed by Rodenacker [2.46].

Energy can be converted in a variety of ways. An electric motor converts electrical into mechanical and thermal energy, a combustion engine converts chemical into mechanical and thermal energy, a nuclear power station converts nuclear into thermal energy, and so on.

Materials too can be converted in a variety of ways. They can be mixed, separated, dyed, coated, packed, transported, reshaped and have their state changed. Raw materials are turned into part-finished and finished products. Mechanical parts are given particular shapes and surface finishes and some are destroyed for testing purposes.

Every plant must process information in the form of *signals*. Signals are received, prepared, compared and combined with others, transmitted, displayed, recorded, and so on.

In technical processes, one type of conversion (of energy, material or signals) may prevail over the others, depending on the problem or the type of solution. It is useful to consider these conversions as flows, and the prevailing one as the *main flow*. It is usually accompanied by a second type of flow, and quite frequently all three come into play. There can, for example, be no flow of material or signals without an accompanying flow of energy, however small. The provision and conversion of energy in such cases may not dominate, but it remains necessary to allow for them. Energy flow also involves the transfer of forces, torques, currents, etc., which are then referred to as force flow, torque flow and current flow.

The conversion of energy to produce electrical power, for example, is associated with a material conversion, even though no continuous material flow is visible in a nuclear power station compared to a coal-fired one. The associated flow of signals constitutes an important subsidiary flow for the control and regulation of the entire process.

However, numerous measuring instruments receive, transform and display signals without any flow of material. In many cases energy has to be specially provided for this purpose; in other cases latent energy can be drawn upon directly. Every flow of signals is associated with a flow of energy, though not necessarily with a flow of material.

In what follows, we shall be dealing with:

- Energy: mechanical, thermal, electrical, chemical, optical, nuclear …, also force, current, heat …
- Material: gas, liquid, solid, dust …, also raw material, test sample, workpiece …, end-product, component …
- Signals: magnitude, display, control impulse, data, information …

In this book technical systems whose main flow is energy-based are referred to as machines, those whose main flow is material-based as apparatus, and those whose main flow is signal-based as devices, unless these terms are not in line with established terminology.

In every type of proposed conversion, *quantity* and *quality* must be taken into consideration if rigorous criteria for the definition of the task, for the choice of solutions and for evaluation are to be established. No statement is fully defined unless its quantitative as well as its qualitative aspects are taken into account. Thus, the statement "100 kg/s of steam at 80 bar and 500 °C" is not a sufficient definition of the input of a steam turbine unless there is the further specification that these figures refer to a nominal quantity of steam and not, for instance, to the maximum flow capacity of the turbine, and the admissible fluctuations in the state of the steam are fixed at, say, 80 bar ± 5 bar and 500 °C ± 10 °C, that is, extended by a qualitative aspect.

In many applications, it is also essential to stipulate the *cost* or value of the inputs and the maximum permissible cost of the outputs (see [2.46], Categories: Quantity–Quality–Cost).

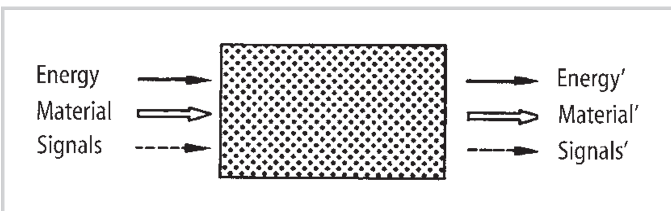


Figure 2.2. The conversion of energy, material and signals. Solution not yet known; task or function described on the basis of inputs and outputs

All technical systems, therefore, involve the conversion of energy, material and signals, which must be defined in quantitative, qualitative and economic terms (see Figure 2.2).

2.1.3 Functional Interrelationship

1. Task-Specific Description

In order to solve a technical problem, we need a system with a clear and easily reproduced relationship between inputs and outputs. In the case of material conversions, for instance, we require identical outputs for identical inputs. Also, between the beginning and the end of a process, for instance filling a tank, there must be a clear and reproducible relationship. Such relationships must always be planned—that is, designed to meet a specification. For the purpose of describing and solving design problems, it is useful to apply the term *function* to the intended input/output relationship of a system whose purpose is to perform a task.

For static processes it is enough to determine the inputs and outputs; for processes that change with time (dynamic processes), the task must be defined further by a description of the initial and final magnitudes. At this stage there is no need to stipulate what solution will satisfy this kind of function. The function thus becomes an abstract formulation of the task, independent of any particular solution. If the overall task has been adequately defined—that is, if the inputs and outputs of all the quantities involved and their actual or required properties are known—then it is possible to specify the *overall function*.

An overall function can often be divided directly into identifiable *subfunctions* corresponding to subtasks. The relationship between subfunctions and the overall function is very often governed by certain constraints, inasmuch as some subfunctions have to be satisfied before others.

On the other hand, it is usually possible to link subfunctions in various ways and hence to create variants. In all such cases, the links must be compatible.

The meaningful and compatible combination of subfunctions into an overall function produces a so-called *function structure*, which may be varied to satisfy the overall function. To that end it is useful to make a block diagram in which the processes and subsystems inside a given block (black box) are initially ignored, as shown in Figure 2.3 (see also Figure 2.2). The symbols used to represent subfunctions in a function structure are summarised in Figure 2.4.

Functions are usually defined by statements consisting of a verb and a noun, for example “increase pressure”, “transfer torque” and “reduce speed”. They are derived for each task from the conversions of energy, material and signals discussed in Section 2.1.2. So far as is possible, all of these data should be accompanied by specifications of the physical quantities. In most mechanical engineering applications, a combination of all three types of conversion is usually involved, with the conversion either of material or of energy influencing the function structure decisively. An analysis of all the functions involved is always useful (see also [2.59]).

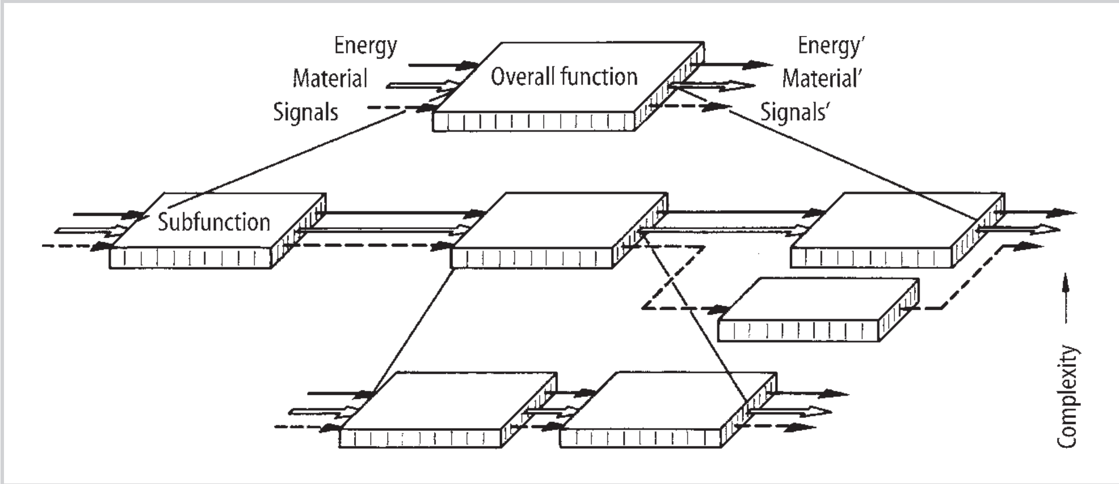


Figure 2.3. Establishing a function structure by breaking down an overall function into subfunctions

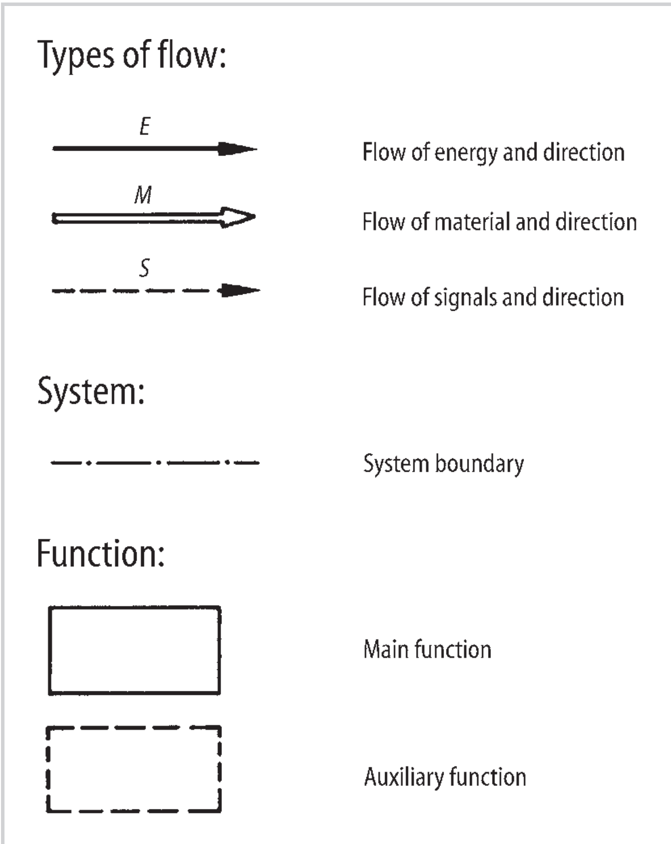


Figure 2.4. Symbols used to represent subfunctions in a function structure

It is useful to distinguish between main and auxiliary functions. While *main functions* are those subfunctions that serve the overall function directly, *auxiliary functions* are those that contribute to it indirectly. They have a supportive or complementary character and are often determined by the nature of the solutions for the main functions. These definitions are derived from Value Analysis [2.7,2.58, 2.60]. Although it may not always be possible to make a clear distinction between main and auxiliary functions, the terms are nevertheless useful. The division between them should be managed in a flexible manner. For example, a change in the system boundary resulting from a change of focus can transform an auxiliary function into a main function and vice versa.

It is also necessary to examine the relationship between the various subfunctions, and to pay particular attention to their logical sequence or required arrangement.

As an example, consider the packing of carpet tiles stamped out of a length of carpet. The first task is to introduce a method of control so that perfect tiles can be selected, counted and packed in specified lots. The main flow here is that of material, as shown in the form of a block diagram in Figure 2.5, which, in this case, is the only possible sequence. On closer examination we discover that this chain of subfunctions requires the introduction of auxiliary functions because:

- the stamping-out process creates offcuts that must be removed
- rejects must be removed separately and reprocessed
- packing material must be brought in.

The result is the function structure shown in Figure 2.6. It will be seen that the subfunction “count tiles” can also give the signal to pack the tiles into lots of a specified size, so it seems useful to introduce a signal flow with the subfunction “send signal to combine *n* tiles into one lot” into the function structure. The functions in this case are *task-specific functions*, whose definitions are derived from the terminology appropriate for the task being considered.

Outside the design domain, the term function is sometimes used in a broader sense, and sometimes in a narrower sense, depending on the context.

Brockhaus [2.40] has defined functions in general as activities, effects, goals and constraints. In mathematics, a function is the association of a magnitude y with a magnitude x such that a unique value (single-valued function) or more than one value (multi-valued function) of y is assigned for every value of x. According to the value analysis definition given in [2.7], functions define the behaviour of artefacts (tasks, activities, characteristics).

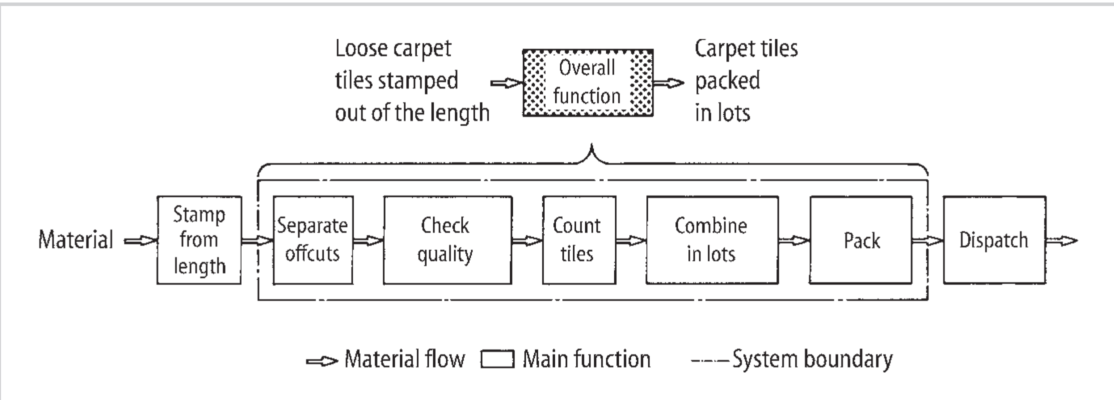


Figure 2.5. Function structure for the packing of carpet tiles

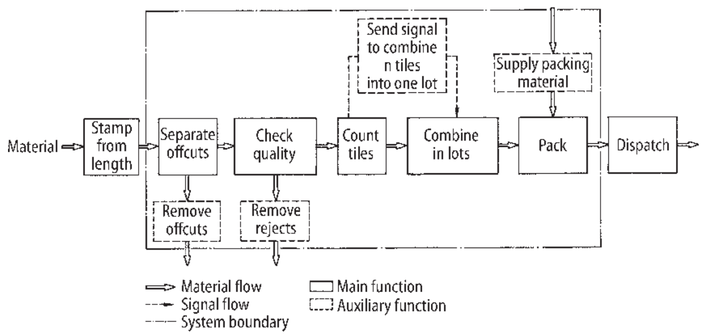


Figure 2.6. Function structure for the packing of carpet tiles as shown in Figure 2.5 but with auxiliary functions added

2. Generally Valid Description

Various design methodologists (see Section 1.2.3) have put forward wider or stricter definitions of *generally valid functions*. In theory, it is possible to classify functions so that the lowest level of the function structure consists exclusively of functions that cannot be subdivided further while remaining generally applicable. They therefore represent a high level of abstraction.

Rodenacker [2.46] has defined generally valid functions in terms of binary logic, Roth [2.47, 2.49] in terms of their general applicability, and Koller [2.28, 2.29] in terms of the required physical effects. Krumhauer [2.31] has examined general functions in the light of possible computer applications during the conceptual design phase, paying special attention to the relationship between inputs and outputs after changes in type, magnitude, number, place and time. By and large, he arrives at the same functions as Roth, except that by "change" he refers exclusively to changes in the type of input and output, while by "increase or decrease" he refers exclusively to changes in magnitude.

In the context of the design methodology presented here, the generally valid functions of Krumhauer will be used (see Figure 2.7).

The function chain shown in Figure 2.5 can be represented using generally valid functions, as shown in Figure 2.8.

A comparison between the functional representations in Figures 2.5 and 2.8 shows that the description that uses generally valid functions has a higher level of abstraction. For this reason, it leaves open all possible solutions and makes a systematic approach easier. However, using generally valid functions can represent a problem because such an abstract level can sometimes hinder the direct search for solutions. For more about the application of task-specific and generally valid functions, along with further examples, see Section 6.3.

3. Logical Description

The logical analysis of functional relationships starts with the search for the essential ones that must necessarily appear in a system if the overall problem is to

Characteristic Input (I)/Output (O)	Generally valid functions	Symbols	Explanations
Type	Change		Type and outward form of I and O differ
Magnitude	Vary		$I < O$ $I > O$
Number	Connect		Number of $I > O$ Number of $I < O$
Place	Channel		Place of $I \neq O$ Place of $I = O$
Time	Store		Time of $I \neq O$

Figure 2.7. Generally valid functions derived from the characteristics type, magnitude, number, place and time for the conversion of energy, materials and signals

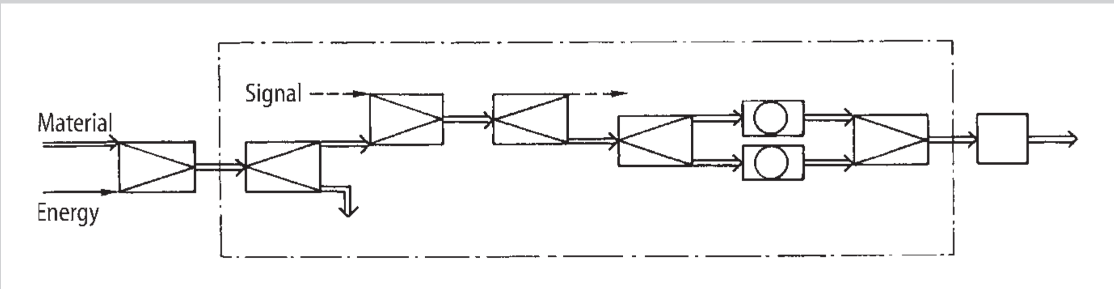


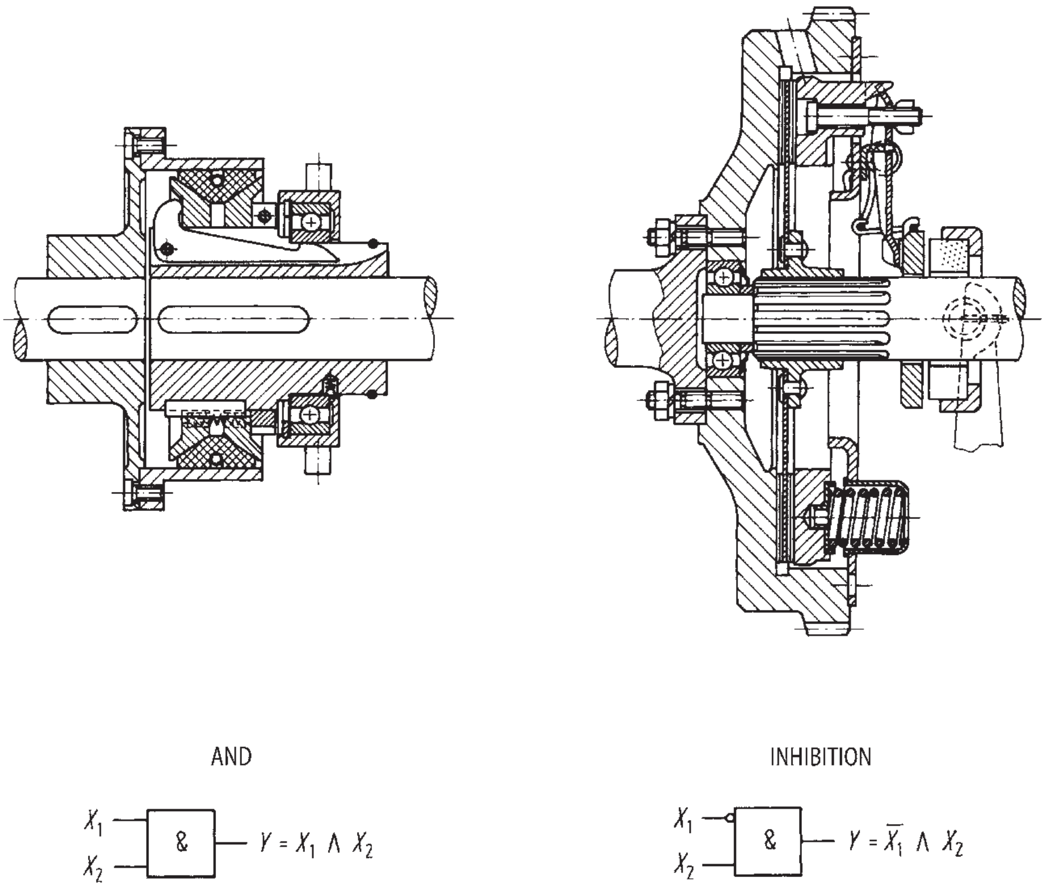
Figure 2.8. Same function structure as shown in Figure 2.5 but represented using generally valid functions, as defined in Figure 2.7

be solved. It may equally well be the relationships between subfunctions as those between inputs and outputs of particular subfunctions.

Let us first of all look at the relationships between subfunctions. As we have pointed out, certain subfunctions must be satisfied before another subfunction can be meaningfully introduced. The so-called "if–then" relationship helps to clarify this point: if subfunction *A* is present, then subfunction *B* can come into effect, and so on. Often several subfunctions must all be satisfied simultaneously before another subfunction can be put into effect. The arrangement of subfunctions thus determines the structure of the energy, material and signal conversions under consideration. Thus, during a test of tensile strength, the first subfunction—"load specimen"—must be satisfied before the other subfunctions—"measure force" and "measure deformation"—can be deployed. The last two subfunctions, moreover, must be satisfied simultaneously. Attention must be paid to consistency and order within the flow under consideration, and this is done by the unambiguous combination of the subfunctions.

Designation	AND-function (Conjunction)	OR-function (Disjunction)	NOT-function (Negation)
Symbol	X_1, X_2 — [&] — Y	X_1, X_2 — [≥1] — Y	X — [1]o — Y
Truth table	X_1: 0 1 0 1 X_2: 0 0 1 1 Y: 0 0 0 1	X_1: 0 1 0 1 X_2: 0 0 1 1 Y: 0 1 1 1	X: 0 1 Y: 1 0
Boolean algebra (Function)	$Y = X_2 \wedge X_1$	$Y = X_1 \vee X_2$	$Y = \overline{X}$

Figure 2.9. Logical functions. X independent statement (signal); Y dependent statement; "0", "1" value of statement, e.g. "off", "on"



X_1	(Signal supplied)	0	1	0	1
X_2	(Clutch engaged)	0	0	1	1
Y	(Torque transmitted)	0	0	0	1

X_1	0	1	0	1
X_2	1	0	0	1
Y	1	0	0	0

Figure 2.10. Logical function of two clutches

Logical relationships, moreover, must also be established between the inputs and outputs of a particular subfunction. In most cases there are several inputs and outputs whose relationships can be treated like propositions in binary logic. *Elementary logical links* of the input and output magnitudes exist for this purpose. In binary logic these are statements such as true/false, yes/no, in/out, fulfilled/unfulfilled, present/not present, which can be computed using Boolean algebra.

We distinguish between AND functions, OR functions and NOT functions, and also between their combination into more complex NOR functions (OR with NOT), NAND functions (AND with NOT) and storage functions with the help of flip-flops [2.4, 2.45, 2.46]. Grouped together, these are called *logical functions*.

In the case of AND functions, all signals on the input side must have the same validity if a valid signal is to appear on the output side.

In the case of OR functions, only one signal needs to be valid on the input side if a valid signal is to appear on the output side.

In the case of NOT functions, the signal on the input side is negated so that the negated signal appears on the output side.

All of these logical functions can be expressed by standard symbols, which can be found in [2.4]. The logical validity of any signal can be read from the truth table shown in Figure 2.9, in which all of the inputs are combined systematically to yield the relevant outputs. The Boolean equations have been added for the sake of completeness. Using logical functions it is possible to construct complex switchess and thus to increase the safety and reliability of control and communication systems.

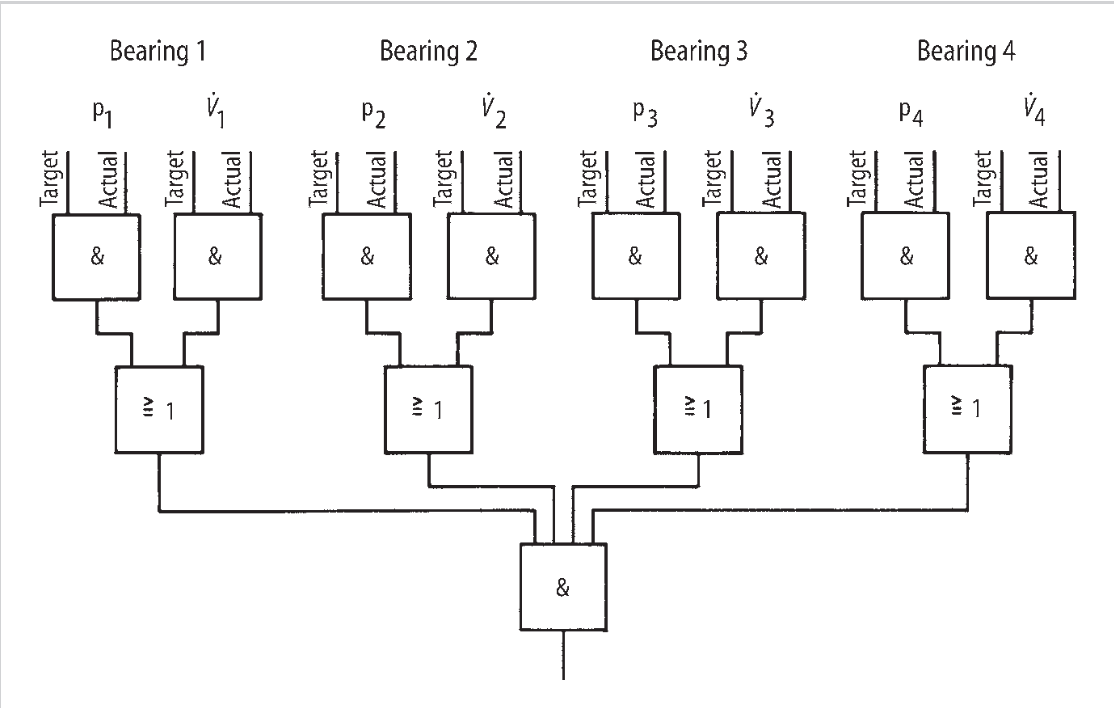


Figure 2.11. Logical functions for monitoring a bearing lubrication system. A positive signal for every bearing (oil present) permits operation. Monitor pressure p; monitor oil flow $\dot{V}$

Figure 2.10 shows two mechanical clutches with their characteristic logical functions. The workings of the clutch on the left can be represented by a simple AND function (the signal must be sent and the clutch engaged before the torque can be transmitted). The clutch on the right has been constructed such that, when the operating signal is given, the clutch is disengaged, meaning that X_1 must be negative if the torque is to be transmitted. In other words, only X_2 must be present or positive if the desired effect is to be produced.

Figure 2.11 shows a logical system for monitoring the bearing lubrication system of a multi-bearing machine shaft involving AND and OR functions. Every bearing position is monitored for oil pressure and oil flow by comparing a specified or target value with the actual value. However, only one positive value for each bearing position is needed to allow the system to operate.

2.1.4 Working Interrelationship

Establishing a function structure facilitates the discovery of solutions because it simplifies the general search for them, and also because solutions to subfunctions can be elaborated separately.

Individual subfunctions, originally represented by “black boxes”, must now be replaced with more concrete statements. Subfunctions are usually fulfilled by physical, chemical or biological processes—mechanical engineering solutions are based mainly on physical processes whereas process engineering solutions are based mainly on chemical and biological processes. If, in what follows, we refer to *physical processes*, we tacitly include the effects of possible chemical and biological processes.

A physical process realised by the selected *physical effects* and the determined *geometric* and *material characteristics* results in a *working interrelationship* that ensures the function is fulfilled in accordance with the task. Hence a working interrelationship comes into existence through physical effects in combination with the chosen geometric and material characteristics.

1. Physical Effects

Physical effects can be described quantitatively by means of the physical laws governing the physical quantities involved. Thus, the friction effect is described by Coulomb’s law, $F_F = \mu F_N$; the lever effect by the lever law $F_A \cdot a = F_B \cdot b$; and the expansion effect by the expansion law $\Delta l = \alpha \cdot l \cdot \Delta\vartheta$ (see Figure 2.12). Rodenacker [2.46] and Koller [2.28], in particular, have collated such effects.

Several physical effects may have to be combined in order to fulfil a subfunction. Thus the operation of a bimetallic strip is the result of a combination of two effects, namely thermal expansion and elasticity.

A subfunction can often be fulfilled by one of a number of physical effects. Thus a force can be amplified by the lever effect, the wedge effect, the electromagnetic effect, the hydraulic effect, etc. The physical effect chosen for a particular sub-function must, however, be compatible with the physical effects of other related

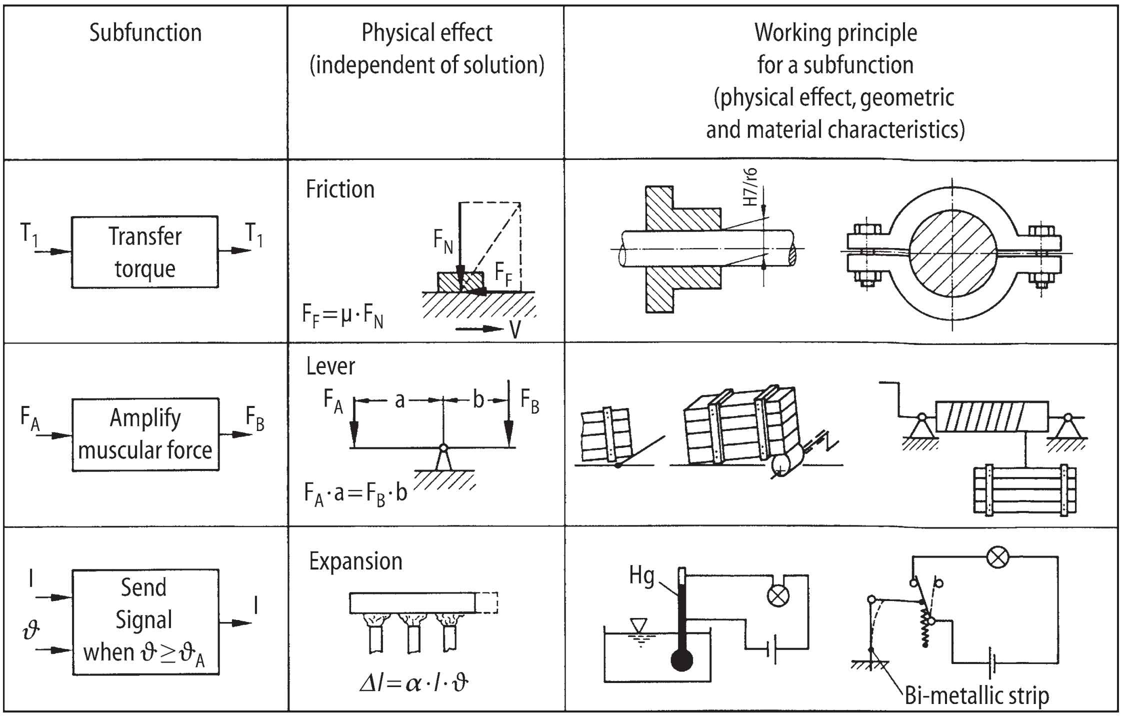


Figure 2.12. Fulfilling subfunctions by working principles built up from physical effects and geometric and material characteristics

subfunctions. A hydraulic amplifier, for instance, cannot be powered directly by an electric battery. Moreover, a particular physical effect can only fulfil a subfunction optimally under certain conditions. Thus a pneumatic control system will be superior to a mechanical or electrical control system only in particular circumstances.

As a rule, compatibility and optimal fulfilment can only be realistically assessed in relation to the overall function once the geometric and material characteristics have been established more concretely.

2. Geometric and Material Characteristics

The place where the physical process actually takes effect is the *working location*, i.e. the specific active location that is the focus of interest at the time. A function is fulfilled by the physical effect, which is realised by the *working geometry*, i.e. the arrangement of *working surfaces* (or working spaces), and by the choice of *working motions* [2.33].

The working surfaces are varied with respect to and determined by:

- Type
- Shape
- Position
- Size
- Number [2.46].

Similarly, the required working motions are determined by:

- Type: translation–rotation
- Nature: regular–irregular
- Direction: in x-, y-, z-directions and/or about x-, y-, z-axes
- Magnitude: velocity, etc.
- Number: one, several, etc.

In addition, we need a general idea of the *type of material* with which the working surfaces are to be produced, for example, whether it is solid, liquid or gaseous; rigid or flexible; elastic or plastic; stiff, hard or tough; or corrosion-resistant. A general idea of the final embodiment is often insufficient; the *main material properties* must be specified before a working interrelationship can be formulated adequately (see Figure 3.18).

Only the combination of the physical effect with the geometric and material characteristics (working surfaces, working motions and materials) allows the principle of the solution to emerge. This interrelationship is called the *working principle* (Hansen [2.19] refers to this as the working means), and it is the first concrete step in the implementation of the solution.

Figure 2.12 shows some examples:

- Transferring the torque through friction against a cylindrical *working surface* in accordance with Coulomb's law will, depending on the way in which the normal force is applied, lead to the selection of a shrink fit or a clamp connection as the working principle.
- Amplifying muscular force with the help of a lever in accordance with the lever law after determining the pivot and force application points (*working geometry*) and considering the necessary *working motion* will lead to a description of the working principle (lever solution, eccentric solution, etc.).
- Making electrical contact by bridging a gap using the expansion effect, applied in accordance with the linear expansion law, only leads to an overall working principle after determination of the sizes (e.g. the diameter and length) and the positions of the *working surfaces* needed for the *working motion* of the expanding medium: a *material*. For example, either mercury expanding by a fixed amount or a bimetallic strip serving as a switch.

To satisfy the overall function, the working principles of the various subfunctions have to be combined (see Section 3.2.4). There are obviously several ways in which this can be done. Guideline VDI 2222 [2.55] calls each combination a *combination of principles*.

The combination of several working principles results in the *working structure* of a solution. It is through this combination of working principles that the solution principle for fulfilling the overall task can be recognised. The working structure derived from the function structure thus represents how the solution will work at the fundamental principle level. Hubka refers to the working structure as the organ structure [2.22–2.24].

For known elements, a circuit diagram or a flow chart is sufficient as a means of representing a working structure. Mechanical artefacts can be effectively represented using engineering drawings, though new or uncommon elements may require additional explanatory sketches (see Figures 2.12 and 2.13).

Often the working structure alone will not be concrete enough to evaluate the solution principle. It may need to be quantified, for example by preliminary calculations and rough scale drawings, before the solution principle can be fixed. The result is called a *principle solution*.

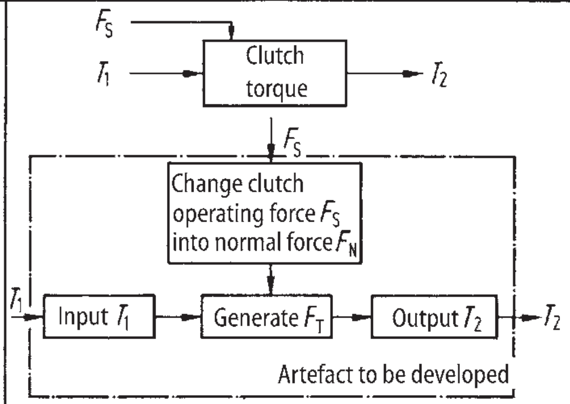
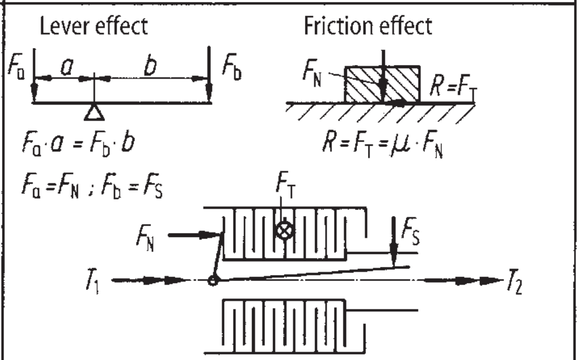
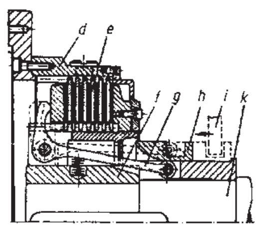
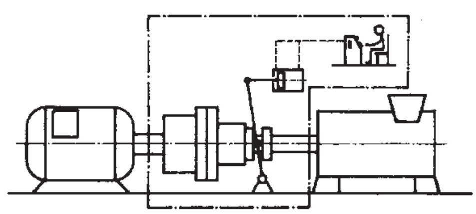
Interrelationships	Elements	Structures	Examples
Functional interrelationship	Functions	Functions structure	
Working interrelationship	Physical effects and geometric and material characteristics ↓ Working principles	Working structure	
Constructional interrelationship	Components Joints Assemblies	Construction structure	
System interrelationship	Artefacts Human beings Environment	System structure	

Figure 2.13. Interrelationships in technical systems

2.1.5 Constructional Interrelationship

The working interrelationship established in the working structure is the starting point for further concretisation leading to the *construction structure*. This interrelationship represents the concrete technical artefact or system by defining the components, assemblies and machines and their interconnections. The construction structure takes into account the needs of production, assembly, transport, etc. Figure 2.13 shows the fundamental interrelationships for the clutch shown in Figure 2.1. The increasing levels of concretisation can be seen clearly.

The concrete elements of a construction structure must satisfy the requirements of the selected working structure plus any other requirements necessary for the technical system to operate as intended. To identify these requirements fully, it is usually necessary to consider the system interrelationship.

2.1.6 System Interrelationship

Technical artefacts and systems do not operate in isolation and are, in general, part of a larger system. To fulfil its overall function, such a system often involves human beings who influence it through *input effects* (operating, controlling, etc.). The system returns *feedback effects* or signals that lead to further actions (see Figure 2.14). In this way, human beings support or enable the *intended effect* of the technical system.

Apart from desired inputs, undesired ones from the environment and from neighbouring systems can affect a technical system. Such *disturbing effects* (e.g. excess temperatures) can cause undesired *side-effects* (e.g. deviations from shape or shifts in position). Also, it is possible that in addition to the desired working interrelationship (intended effects), unwanted phenomena can occur (e.g. vibrations) as side-effects from individual components within the system or from the overall system itself. These side-effects can have an adverse effect on humans or the environment.

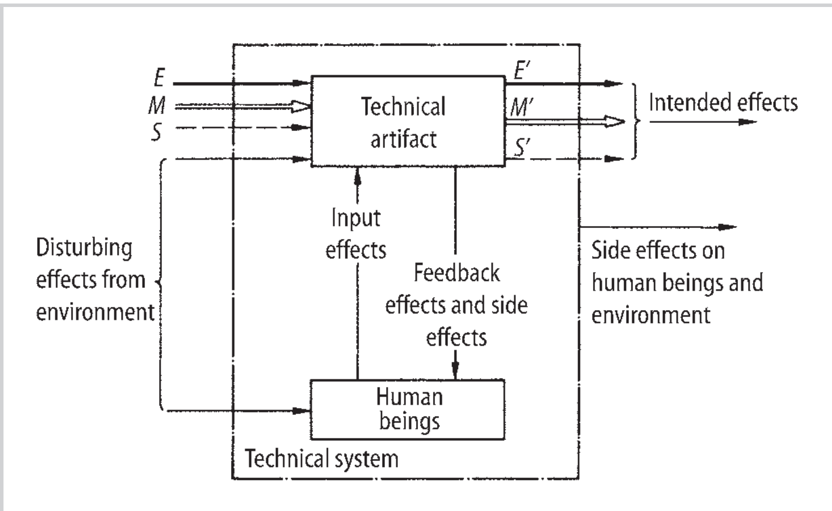


Figure 2.14. Interrelationships in technical systems including human beings

In accordance with Figure 2.14 it is useful to make the following distinctions (after [2.56]):

Intended effect:	Functionally desired effect in the sense of system operation.
Input effect:	Functional relationship due to human action on a technical system.
Feedback effect:	Functional relationship due to the action of a technical system on a human or another technical system.
Disturbing effect:	Functionally undesired influence from outside on a technical system or human that makes it difficult for a system to fulfil its function.
Side effect:	Functionally undesired and unintended effect of a technical system on a human or on the environment.

The overall interrelationship of all these effects must be carefully considered during the development of technical systems. To help recognise them in time, so that desired effects can be used and undesired ones avoided, it is helpful to follow a systematic guideline that adheres to the general objectives and constraints in Section 2.1.7.

2.1.7 Systematic Guideline

The solution of technical tasks is determined by the general objectives and constraints. The *fulfilment of the technical function*, the *attainment of economic feasibility* and the *observance of safety requirements* for humans and the environment can be considered as general objectives. The fulfilment of the technical function alone does not complete the task of designers; it would simply be an end unto itself. Economic feasibility is another essential requirement, and concern with human and environmental safety must impose itself for ethical reasons. Every one of these objectives has direct repercussions on the rest.

In addition, the solution of technical tasks imposes certain constraints or requirements resulting from ergonomics, production methods, transport facilities, the intended operation, etc., no matter whether these constraints are the result of the particular task or the general state of technology. In the first case we speak of task-specific constraints, in the second of general constraints that, although often not specified explicitly, must nevertheless be taken into account.

Hubka [2.22–2.24] separates the properties affected by the constraints into categories based variously on industrial, ergonomic, aesthetic, distribution, delivery, planning, design, production and economic factors.

Besides satisfying the functional and working interrelationships, a solution must also satisfy certain general or task-specific constraints. These can be classified under the following headings:

- Safety — also in the wider sense of reliability and availability
- Ergonomics — human–machine context, also aesthetics
- Production — production facilities and type of production

- Quality control throughout the design and production process
- Assembly during and after the production of components
- Transport inside and outside of the factory
- Operation intended use, handling
- Maintenance upkeep, inspection and repair
- Expenditure costs and schedules
- Recycling reuse, reconstitution, disposal, final storage.

The characteristics that can be derived from these constraints, which are generally formulated as requirements (see Section 5.2), affect the function, working and construction structures, and also influence one another. Hence they should be treated as *guidelines* throughout the design process, and adapted to each level of embodiment (see Figs. 2.15 and 12.3).

In addition there are influences from the designer, the development team and the suppliers as well as the customer, the specific context and the environment.

It is advisable to consider these guidelines even during the *conceptual phase*, at least in essence. During the *embodiment phase*, when the layout and form design of the more or less qualitatively elaborated working structure is first quantified, both the objectives of the task and also the general and task-specific constraints must be considered in concrete detail. This involves several steps—the collection of further information, layout and form design, and the elimination of weak spots, together with a fresh, if limited, search for solutions for a variety of subtasks, until finally,

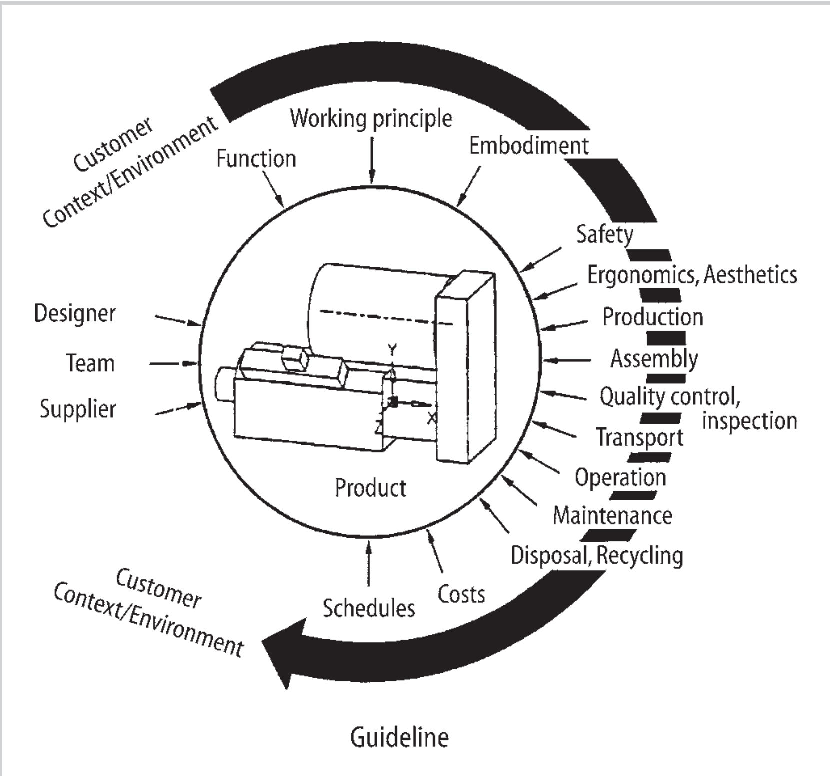


Figure 2.15. Influences and constraints during design and development. These can provide a guideline for quality control

in the *detail phase*, the elaboration of detailed production instructions brings the design process to a conclusion (see Chapters 5 to 7).

2.2 Fundamentals of the Systematic Approach

Before we deal with the specific steps and rules of systematic design, we must first discuss cognitive psychological relationships and general methodical principles. These help to structure the proposed procedures and individual methods so that they can be applied to the solution of design tasks in a purposeful way. The ideas come from a host of different disciplines, mainly non-technical ones, and are usually built on interdisciplinary fundamentals. Work science, psychology and philosophy are among the main inspirations, which is not surprising when we consider that methods designed to improve working procedures impinge on the qualities, capacities and limitations of human thought [2.41].

2.2.1 Problem Solving Process

Designers are often confronted with tasks containing problems they cannot solve immediately. Problem solving in different areas of application and at different levels of concretisation is a characteristic of their work. Researching the essence of human thinking is the focus of cognitive psychology. The results of this research must be taken into account in engineering design. The following sections are based largely on the work of Dörner [2.8, 2.10].

A *problem* has three components:

- an undesirable initial state, i.e. the existence of an unsatisfactory situation
- a desirable goal state, i.e. the realisation of a satisfactory situation
- obstacles that prevent a transformation from the undesirable initial state to the desirable goal state at a particular point in time.

An *obstacle* that prevents a transformation can arise from the following:

- The means to overcome the obstacle are unknown and have to be found (synthesis or operator problem).
- The means are known, but they are so numerous or involve so many combinations that a systematic investigation is impossible (interpolation problem, combination and selection problem).
- The goals are only known vaguely or are not formulated clearly. Finding a solution involves continuous deliberation and the removal of conflicts until a satisfactory situation is reached (dialectic problem, search and application problem).

A *problem* has the following typical characteristics:

- *Complexity:* many components are involved and these components, through links of different strength, influence each other.

- *Uncertainty:* not all requirements are known; not all criteria are established; the effect of a partial solution on the overall solution or on other partial solutions is not fully understood, or only emerges gradually. The difficulties become more pronounced if the characteristics of the problem area change with time.

A *task* is distinct from a problem because:

- A *task* imposes mental requirements for which various means and methods are available to assist. An example is the design of a shaft with given loads, connecting dimensions and production methods.

Tasks and problems occur in design in a number of ways, often combined and not clearly separable initially. A specific design task can, for example, turn out to be a problem when looked at more closely. Many large tasks can be divided into subtasks, some of which can reveal difficult subproblems. On the other hand, it is sometimes possible for a problem to be solved by fulfilling several subtasks in a previously unknown combination.

Thinking processes take place in the brain and involve changes in memory content. When thinking, the contents of the memory, and the way in which they are linked, play an important role.

In simple terms, one can say that in order to start solving a problem humans need a certain level of *factual knowledge* about the domain of the problem. In cognitive psychology, when this knowledge has been transferred into memory it represents the *epistemic structure*.

Humans also need certain *procedures* (methods) to find solutions and to find these effectively. This aspect involves the *heuristic structure* of human thought.

It is possible to distinguish between short-term and long-term memory. Short-term memory is a kind of working storage. It has limited capacity and can only retain about seven arguments or facts at the same time. Long-term memory probably has unlimited capacity and contains factual and heuristic knowledge that appears to be stored in a structured way.

In this way, humans are able to recognise specific relationships in many possible ways, to use these relationships and to create new ones. Such relationships are very important in the technical domain, for example:

- concrete—abstract relationship

  e.g. angular contact bearing—ball bearing—rolling element bearing—bearing—guide—transfer force and locate component.
- whole—part relationship (hierarchy)

  e.g. plant—machine—assembly—component.
- space and time relationships

  e.g. arrangement: front—back, below—above,

  e.g. sequence: this first—that next.

The memory can be thought of as a semantic network with nodes (knowledge) and connections (relationships) which can be modified and extended. Figure 2.16 shows a possible, though not necessarily complete, semantic network related to

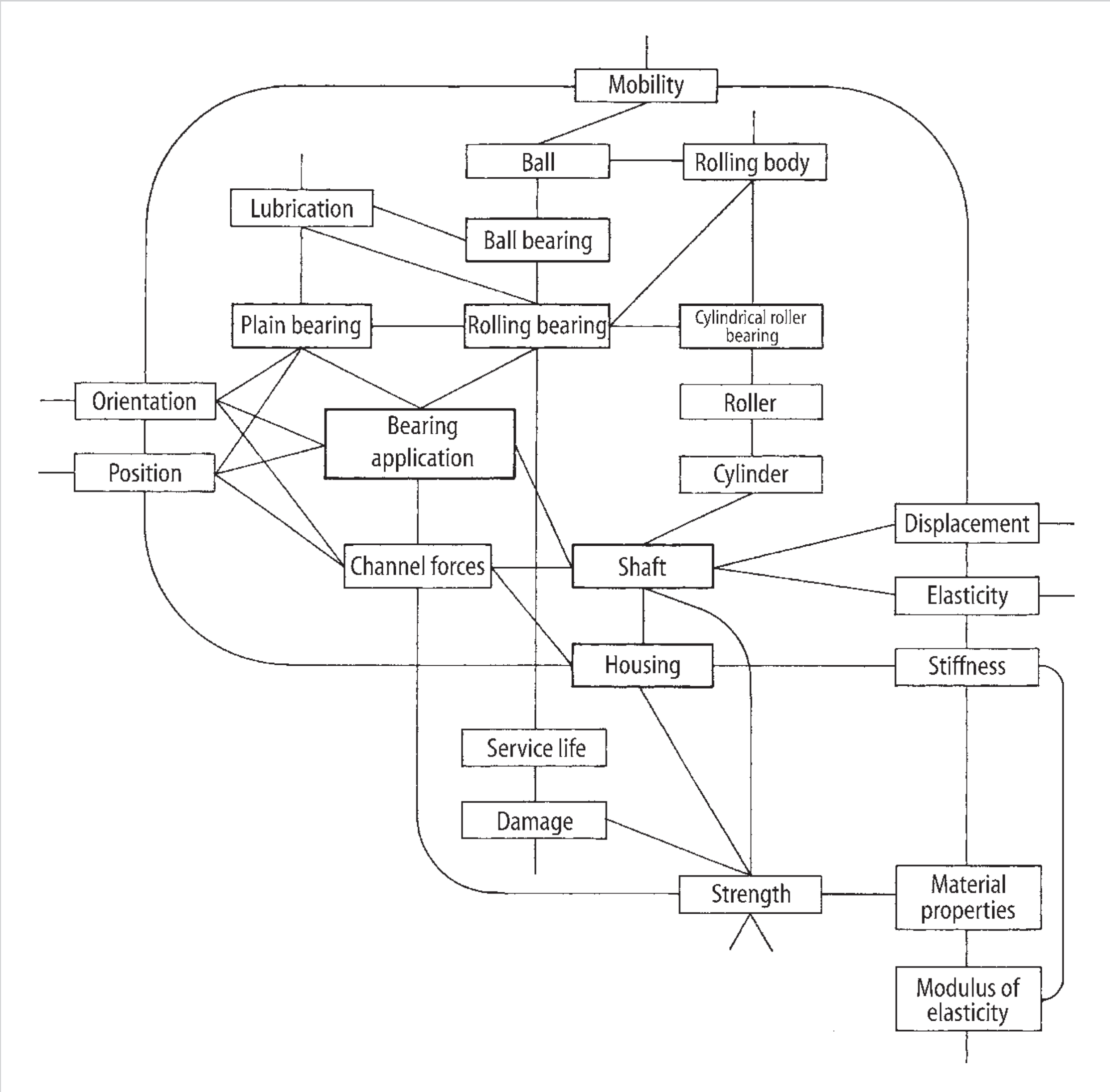


Figure 2.16. Extract of a semantic network related to bearings

the term "bearing". In this network it is possible to recognise the relationships mentioned above as well as others, such as property relationships and ones indicating opposites (polar relationships). Thinking involves building and restructuring such semantic networks, and the thinking process itself can proceed intuitively or discursively.

Intuitive thinking is strongly associated with flashes of inspiration. The actual thinking process takes place to a large extent unconsciously. Insights appear in the conscious mind suddenly, caused by some trigger or association. This is referred to as primary creativity [2.2, 2.30] and involves processing quite complex relations. In this context, Müller [2.36] refers to "silent knowledge", which includes common and background knowledge. This is also the knowledge that is available when one deals with episodic memories, vague concepts and imprecise definitions. It is activated by both conscious and unconscious thinking activities.

Generally time is needed for undisturbed and unconscious "thinking" before sudden insights appear. The length of this incubation period cannot be predetermined. Insights can be triggered, for example, by producing freehand sketches or engineering drawings of solution ideas. According to [2.14], these manual activities focus concentration on the subject, but still leave space in the mind that can by used by unconscious thinking processes, which can also be stimulated by such activities.

Discursive thinking is a conscious process that can be communicated and influenced. Facts and relationships are consciously analysed, varied, combined in new ways, checked, rejected, and considered further. In [2.2, 2.30] this is referred to as secondary creativity. This type of thinking involves checking exact and scientific knowledge and building this into a knowledge structure. In contrast to intuitive thinking, this process is slow and involves many small conscious steps.

In the memory structure, explicit and consciously acquired knowledge cannot be separated precisely from the vaguer common or background knowledge. Besides, the two types of knowledge influence each other. For knowledge to be easily retrieved and combined, it is thought that an ordered and logical structure of factual knowledge in the mind of the problem solver (epistemic structure) is decisive, and that this is true whether the thinking process is intuitive or discursive.

The *heuristic structure* includes explicit knowledge (i.e. knowledge that can be explained) as well as implicit knowledge. This is necessary in order to organise the sequence of thinking operations, including modifying operations (searching and finding) and testing operations (checking and assessing). It appears that problem solvers often start without a fixed plan in the hope of immediately finding a solution from their knowledge bases without much effort. Only when this approach fails, or when contradictions emerge, do they adopt a more clearly planned or systematic sequence of thinking operations.

The so-called TOTE model [2.33] represents an important fundamental sequence for thinking processes (see Figure 2.17). It consists of two processes: a modification process and a testing process. The TOTE model shows that before an operation of change takes place, an operation of testing (*T*est) is invoked to analyse the initial state. Only then is the chosen operation of change (*O*peration) executed. This is followed by another operation of testing (*T*est), during which the resulting state is checked. If the result is satisfactory, the process is exited (*E*xit); if not, the operation is adapted and repeated.

In more complex thinking processes, the TOTE sequences are linked in a chain or several modification processes are executed before a testing process takes place. Thus, when linking mental processes, many combinations and sequences are possible, but all of them can be mapped onto the basic TOTE model.

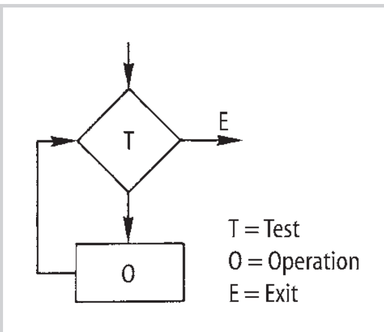


Figure 2.17. Basic TOTE model for thinking processes [2.8, 2.33]

2.2.2 Characteristics of Good Problem Solvers

The following statements are the result of the work of Dörner [2.9] and of research which has been undertaken with him by Ehrlenspiel and Pahl. The results of the research led by Ehrlenspiel and Pahl can be found in the publications of Rutz [2.50], Dylla [2.11, 2.12] and Fricke [2.15, 2.16]. This section provides a summary of their findings [2.42].

1. Intelligence and Creativity

In general, *intelligence* is thought to involve a certain cleverness, combined with the ability to understand and judge. Analytical approaches are often emphasised.

Creativity is an inspirational force that generates new ideas or produces novel combinations of existing ideas, leading to further solutions or deeper understanding. Creativity is often associated with an intuitive, synthesising approach.

Intelligence and creativity are personal characteristics. Up until now it has not been possible to come up with precise scientific definitions of or a clear distinction between intelligence and creativity. Attempts have been made to measure the level of intelligence of individuals using intelligence tests. The resulting Intelligence Quotients provide measures compared to the average of a large sample. Because of the different forms in which intelligence appears, various tests are needed to get a complete picture and draw tentative conclusions. The same is true for creativity tests.

For problem solving, a minimum level of intelligence is required and it appears that people with high Intelligence Quotients are more likely to be good problem solvers. However, according to [2.8, 2.9], intelligence tests on their own do not give much insight into which combination of factors makes a particular individual a good problem solver. The reason, according to Dörner [2.8], is that intelligence tests use tasks or problems that only require a few thinking steps to find a solution, so the sequence of steps seldom becomes conscious. Few intelligence tests require a large number of steps to be organised into a specific problem solving procedure. Such organisation requires switching between the different levels and possibilities of a general problem solving procedure, and is essential for the execution of long-term thinking activities.

Creativity tests too are often at such a low level that they do not address complex problem solving which involves planning and guiding one's own approach. Furthermore, in engineering design, creativity is always focused on a specific goal. Purely unfocused generation of ideas and variants can in fact hinder the problem solving process [2.2] or at best support a specific phase of the process.

2. Decision Making Behaviour

Apart from having well-structured factual knowledge, applying a systematic approach, and using focused creativity, designers have to master decision making processes. For decision making, the following mental activities and skills are essential:

- *Recognising Dependencies*

  In complex systems the dependencies between the individual elements can vary in strength. Recognising the types and strengths of such dependencies is an essential prerequisite for dividing the problem into more manageable, less complex subproblems or subgoals so that these can be addressed separately. However, those working on each separate subproblem must check to see how the short- and long-term effects of their own decisions influence the overall design.

- *Estimating Importance and Urgency*

  Good problem solvers know how to recognise *importance* (factual significance) and *urgency* (temporal significance), and how to use this information to modify their approach to problems.

  They try to resolve the most important things first and then tackle the dependent subproblems. They have the courage to be satisfied with suboptimal solutions for less significant problems if they have good or acceptable solutions for the most significant ones. By doing this they avoid immersing themselves in less relevant issues and thereby losing valuable time.

  The same is true when estimating the urgency. Good problem solvers estimate the time they need accurately. They prepare a demanding—but not impossible—time plan. Janis and Mann [2.25] have concluded that mild (i.e. bearable) stress is important for creativity. Therefore, realistic time planning has a positive effect on thinking processes, and new developments should take place under reasonable time pressure. But, of course, individuals react differently to time pressure.

- *Continuity and Flexibility*

  Continuity means an appropriate and continuous focus on achieving the goals, but there is a danger that excessive focus leads to a rigid approach. Flexibility means a ready ability to adapt to changing requirements. However, this should not lead to purposeless jumping from one approach to another.

  Good problem solvers find a suitable balance between continuity and flexibility. They demonstrate continuous and consistent, but at the same time flexible, behaviour. They stick to the given goals despite any hold-ups and difficulties they encounter. On the other hand, they adapt their approach immediately when the situation changes and when new problems occur.

  They consider heuristics, procedures and instructions first of all as guidelines and not as rigid prescriptions. Dörner states [2.8]: "Heuristics or heuristic plans should not degenerate into automatic procedures. Individuals should learn to develop what they have learnt. Heuristics should not be misinterpreted as prescriptions, but should be treated as guidelines that can, and often should, be developed."

- *Failures Cannot be Avoided*

  In complex systems with strong internal dependencies, at least partial failures are difficult to avoid because it is not possible to recognise all the potential effects simultaneously. When recognising such failures, the most important thing is the way one reacts. Being flexible is crucial, supported by the ability to analyse one's approach and the ability to make decisions that lead to corrective actions.

The results of cognitive psychology research are summarised below.

Good problem solvers:

- have a sound and structured technical knowledge, i.e. they have a well-structured model in their minds
- find an appropriate balance between concreteness and abstraction, depending on the situation
- can deal with uncertainty and fuzzy data
- continuously focus on the goals while adopting a flexible decision making behaviour.

Such heuristic competence depends largely on personal characteristics, but can be developed considerably through training on different types of problem.

The research mentioned earlier reveals that good designers demonstrate the following behaviour [2.42]:

- They thoroughly analyse the goals at the beginning of a task and continue to do so throughout the design process when formulating partial goals, in particular when the original problem formulation is vague.
- They first generate or identify the most suitable solution principles in a conceptual phase before developing concrete embodiments.
- They initially adopt a diverging search without generating too many variants and then quickly converge onto a small number of solutions; they choose the appropriate level of concretisation and switch easily between perspectives, e.g. abstract/concrete, overall problem/subproblem, working interrelationship/constructional interrelationship.
- They regularly assess their solutions using a comprehensive set of criteria, avoiding emphasising personal preferences.
- They continuously reflect on their approach and adapt it to the situation at hand.

These characteristics are in line with the aims and proposals for the design approach in this book.

2.2.3 Problem Solving as Information Processing

When we discussed the basic ideas of the systems approach (see Section 1.2.3), we found that problem solving demands a large and constant flow of information. Dörner [2.8] also views problem solving as information processing. The most important terms used in the theory of information processing are described in [2.5, 2.6]. Information is *received*, *processed* and *transmitted* (see Figure 2.18).

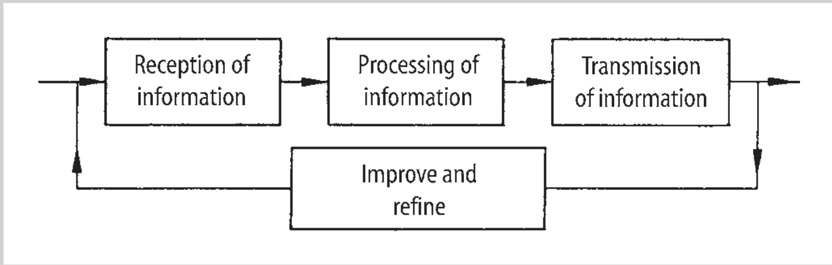


Figure 2.18. The conversion of information with iteration

Information is *received* from market analyses, trend studies, patents, technical journals, research results, licenses, inquiries from customers, concrete assignments, design catalogues, analyses of natural and artificial systems, calculations, experiments, analogies, general and in-house standards and regulations, stock sheets, delivery instructions, computer data, test reports, accident reports, and also by "asking questions". Data collection is an essential element of problem solving [2.3].

Information is *processed* by analysis and synthesis, the development of solution concepts, calculation, experiment, the elaboration of layout drawings and also the evaluation of solutions.

Information is *transmitted* by means of sketches, drawings, reports, tables, production documents, assembly manuals, user manuals, etc. These can be both in hard copy and electronic forms. Quite often provision must also be made for information to be *stored*.

In [2.32] some *criteria* for characterising information are given, and these can be used for formulating user information requirements. They include:

- Reliability: the probability of the information being available, trustworthy and correct.
- Sharpness: the precision and clarity of the information content.
- Volume and density: an indication of the number of words and pictures needed for the description of a system or process.
- Value: the importance of the information to the recipient.
- Actuality: an indication of the point in time when the information can be used.
- Form: the distinction between graphic and alphanumeric data.
- Originality: an indication of whether or not the original character of the information must be preserved.
- Complexity: the structure of, or connectivity between, information symbols and information elements, units or complexes.
- Degree of refinement: the quantity of detail in the information.

Information conversion is usually a very complicated process. Solving problems requires information of different types, content and range. In addition, to raise the level of information and improve it, it may be necessary to reiterate certain steps.

Iteration is the process by which a solution is approached step-by-step. In this process, one or more steps are repeated, each time at a higher level of information based on the results of the previous loop. Only in this way is it possible to obtain the information to refine a solution and ensure continuous improvement (see Figure 2.18). Such iterations occur frequently at all stages of the problem-solving process.

2.2.4 General Working Methodology

A general working methodology should be widely applicable, independent of discipline and should not require specific technical knowledge from the user. It should support a structured and effective thinking process. The following general ideas appear time and time again in specific approaches, either directly or slightly amended to adapt them to the special requirements of developing technical systems. The purpose of this section is to provide a general introduction to systematic procedures. The following procedures are based not only on our own professional experience and on the findings of cognitive psychology mentioned in Section 2.2.1, but also on the work of Holliger [2.20, 2.21], Nadler [2.38, 2.39], Müller [2.35, 2.36] and Schmidt [2.51]. They are also known as "heuristic principles" (a *heuristic* is a method for generating ideas and finding solutions) or "creativity techniques".

The following conditions must be satisfied by anyone using a systematic approach:

- *Define goals* by formulating the overall goal, the individual subgoals and their importance. This ensures the motivation to solve the task and supports insight into the problem.
- *Clarify conditions* by defining the initial and boundary constraints.
- *Dispel prejudice* to ensure the most wide-ranging search for solutions possible and to avoid logical errors.
- *Search for variants* to find a number of possible solutions or combinations of solutions from which the best can be selected.
- *Evaluate* based on the goals and conditions.
- *Make decisions*. This is facilitated by objective evaluations. Without decisions and experiencing their consequences there can be no progress.

To make these general methods work, the following *thinking* and *acting operations* must be considered.

1. Purposeful Thinking

As described in Section 2.2.1, intuitive and discursive thinking are possible. The former tends to be more unconscious, the latter more conscious.

Intuition has led to a large number of good and even excellent solutions. The prerequisite is, however, always a very conscious and intensive involvement with the given problem. Nevertheless, a purely intuitive approach has the following disadvantages:

- the right idea rarely comes at the right moment, since it cannot be elicited and elaborated at will
- the result depends strongly on individual talent and experience
- there is a danger that solutions will be circumscribed by preconceived ideas based on one's special training and experience.

It is therefore advisable to use more deliberate procedures that tackle problems step-by-step, and such procedures are denoted *discursive*. Here the steps are chosen intentionally; they can be influenced and communicated. Usually individual ideas or solution attempts are consciously analysed, varied and combined. It is an important aspect of this procedure that a problem is rarely tackled as a whole, but is first divided into manageable parts and then analysed.

It must, however, be stressed that intuitive and discursive methods are not opposites. Experience has shown that intuition is stimulated by discursive thought. Thus, while complex assignments must always be tackled one step at a time, the subsidiary problems involved may, and often should, be solved in intuitive ways.

In addition, it should be realised that creativity can be inhibited or encouraged by different influences [2.2]. It is, for example, often necessary to encourage intuitive thinking by interrupting the activity to provide some periods of incubation (see Section 2.2.1). On the other hand, too many interruptions can be disturbing and thereby inhibit creativity. A systematic approach including discursive elements and adopting different viewpoints encourages creativity. Examples include using different solution methods; moving between abstract and concrete ideas; collecting information using solution catalogues; and dividing work between team members. Furthermore, according to [2.25], realistic planning encourages rather than inhibits motivation and creativity.

2. Individual Working Styles

Designers should be given some freedom of action in their work to enable them to realise their own optimised working style. They should be free to select their preferred methods, the sequence in which they undertake individual working steps, and the sources of information they wish to consult. They should therefore be allowed to make their own plans for their area of responsibility and for them to have control over these plans. Obviously the individual working plans have to be compatible with the overall approach and make a useful contribution.

In general it is necessary to consider several subfunctions (subproblems) when developing new products. These functions, or combinations of them, lead to partial solutions. In such situations designers can proceed in different ways. One possibility is to search for working principles (solution principles) for every subfunction (or group of subfunctions), to roughly check their compatibility, and then to combine them into an overall working structure (solution concept). Finally the components are embodied, making sure their overall combination is compatible.

From a methodical point of view, this approach is systematic, stepwise and process-oriented; that is, the designer develops the different functional areas in parallel, from abstract (idea generation) to concrete (final embodiment) (see Figure 2.19a).

Another possibility is to proceed from idea generation to final embodiment for every problem or functional area, one after the other, and finally combine and modify these to make them all fit together. From a methodical point of view, this approach is problem-oriented; that is, the designer develops the different functional areas in sequence (see Figure 2.19b).

The investigations of Dylla [2.11, 2.12] and Fricke [2.15, 2.16] show that novices educated in systematic design tend to follow the process-oriented approach, whereas experienced designers tend to follow the problem-oriented approach. Experienced designers apply their wealth of experience, know a wide range of possible subsolutions, and are able to represent these solutions quickly. Hence they arrive relatively quickly at a concrete result. Then, using a corrective approach, they bring this together into an overall solution. This type of approach is successful in those cases where the individual components do not influence each other strongly and their properties are apparent. If these conditions are not met, this approach can lead to a relatively late recognition of a possible lack of compatibility between the functional areas. This approach can also result in different subsolutions being selected for identical, or similar, subfunctions, which is often not economic. In such cases further iterations are required to find other solutions.

The process-oriented approach largely avoids the potential disadvantages of the problem-oriented approach. However, more time is required because of the wider, more systematic perspective. This carries the danger of generating an unnecessarily large solution space. The process-oriented approach therefore requires designers to achieve an appropriate balance between abstract and concrete; that is, to know when a sufficiently large, but not too large, number of solution ideas has been generated (divergence), and the time has come to combine these into a concrete concept (convergence).

In practice, these two approaches (process-oriented and problem-oriented) are often not found in their pure form. They usually appear in various combinations depending on the problem situation. However, individual designers naturally tend to adopt one approach in preference to the other. Process-oriented approaches are recommended when subproblems are strongly interrelated and when breaking new ground. A problem-oriented approach is useful when the connectivity between functional areas is low and when subsolutions are known to exist in the area of application.

Similarly individual differences in approach can be observed during the search for solutions. If designers develop and investigate different solution principles or embodiment variants in parallel while searching for solutions for the individual subfunctions, and then compare these with one another to find the most suitable, this approach is called a *generative search for solutions* (see Figure 2.20a). If, on the other hand, a particular idea or example is used as a starting point and is then improved and adapted in a stepwise approach until a satisfactory solution emerges, this is called a *corrective search for solutions* (see Figure 2.20b). Adopting

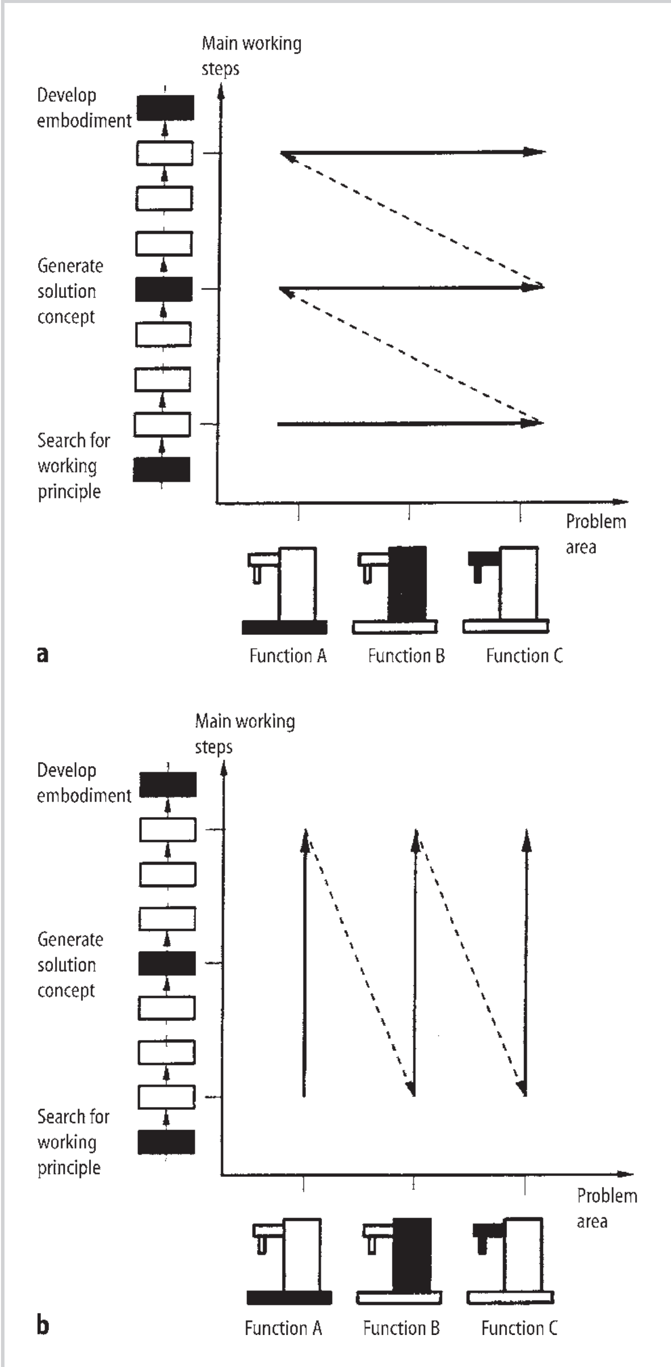


Figure 2.19. Different individual approaches during the development of solutions for a tea-making machine with several linked functional areas: baseplate/control (function A), water reservoir and heating element (function B), spout and closure (function C). **a** Systematic, stepwise, process-oriented, i.e. in every stage of development all functional areas are taken forward; **b** Problem-oriented, i.e. functional areas are developed in sequence before combining them (idealised process representation after Fricke [2.15, 2.16])

this latter approach will also result in a range of solution variants, if individual variants are not rejected.

A generative search for solutions increases the chances of finding new and unconventional ideas and considers many different principles, and thus may result in a larger solution space. The challenge, however, is a timely and goal-oriented

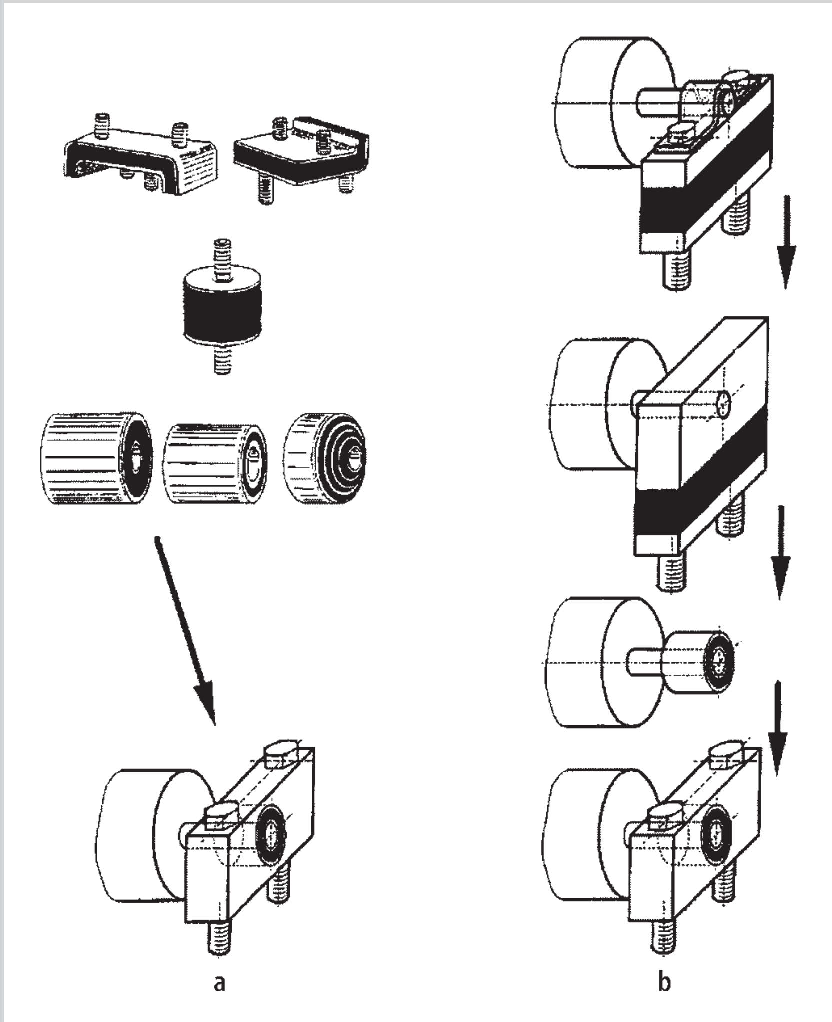


Figure 2.20. Different individual approaches during the search for solutions for an elastic support. **a** Generative, i.e. generation of various solutions and goal-oriented selection. **b** Corrective, i.e. search for solutions by improvement and adaptation of one idea

selection to avoid wasting time on unfeasible solutions. This type of search is typical for novices who have been taught systematic design and for designers who have adopted the systematic approach.

A corrective search for solutions is often used by inexperienced designers, in particular when they can think of a similar known solution in the application area. The advantage is that it is possible to concretise the solutions relatively quickly, even if these initial solutions are not really satisfactory. When adopting this type of search, designers tend to remain in their area of expertise and only expand this slowly. Possible dangers include fixating on solution ideas that are less suitable in principle and failing to recognise other better solution principles.

In practice, designers tend to adopt a mixture of search types with the main aim of minimising their work effort. However, designers clearly favour one or the other search type because of their individual talents and experience, usually without being aware of the advantages or dangers of their particular styles.

The consciously or unconsciously applied approaches depend on education and experience and can be influenced. Designers should not be forced into adopting a particular approach. On the contrary, it is better to make them aware of the advantages and dangers of the various approaches and leave the final decision up to them. It is, however, useful through training and further education, along with appropriate management during the project, to identify the most suitable overall approach and to agree on this.

2.2.5 Generally Applicable Methods

The following general methods provide further support for systematic work, and are widely used [2.21]. Often so-called "new" methods only involve repackaging one of the general methods described below.

1. Analysis

Analysis is the resolution of anything complex into its elements and the study of these elements and their interrelationships. It calls for identification, definition, structuring and arrangement. The acquired information is transformed into knowledge. If errors are to be minimised, then problems must be formulated clearly and unambiguously. To that end, they have to be analysed. *Problem analysis* means separating the essential from the nonessential and, in the case of complex problems, preparing a discursive solution by resolution into individual, more transparent, subproblems. If the search for the solution proves difficult, a new formulation of the problem may provide a better starting point. The reformulation of statements is often an effective means of finding new ideas and insights. Experience has shown that careful analysis and formulation of problems are among the most important steps of the systematic approach.

The solution of a problem can also be brought nearer by *structure analysis*, that is, the search for hierarchical structures or logical connections. In general, this type of analysis can be said to aim at the demonstration of similarities or repetitive features in different systems, for example by means of analogical reasoning (see Section 3.2.1).

Another helpful approach is *weak spot analysis*. It is based on the fact that every system has weaknesses caused by ignorance, mistaken ideas, external disturbances, physical limitations and production errors. During the development of a system it is therefore important to analyse the design concept or design embodiment for the express purpose of discovering possible weak spots and prescribing remedies. To that end, special selection and evaluation procedures (see Section 3.3) and weak spot identification methods (see Section 10.2) have been developed. Experience has shown that this type of analysis may not only lead to specific improvements of the chosen solution principle, but may also trigger off new solution principles.

2. Abstraction

Through abstraction it is possible to find a higher level interrelationship, that is, one which is more generic and comprehensive. Such a procedure reduces complexity and emphasises the essential characteristics of the problem and thereby provides an opportunity to search for and find other solutions containing the identified characteristics. At the same time new structures emerge in the minds of designers and these assist with the organisation and retrieval of the many ideas and representations. So abstraction supports both creativity and systematic thinking. It makes possible the definition of a problem in such a way that a coincidental solution path is avoided and a more generic solution is found (see example in Section 6.2).

3. Synthesis

Synthesis is the fitting together of parts or elements to produce new effects and to demonstrate that these effects create an overall order. It involves search and discovery, and also composition and combination. An essential feature of all design work is the combination of individual findings or subsolutions into an overall working system—in other words, the association of components to form a whole. During the process of synthesis the information discovered by analyses is processed as well. In general, it is advisable to base synthesis on a holistic or systems approach; in other words, to bear in mind the general task or course of events while working on subtasks or individual steps. Unless this is done, there is the grave risk that, despite the optimisation of individual assemblies or steps, no suitable overall solution will be reached. Appreciation of this fact is the basis of the interdisciplinary method known as Value Analysis, which proceeds from the analysis of the problem and structure to a holistic systems approach involving the early collaboration of all departments concerned with product development. Such an approach is also needed in large-scale projects, especially when preparing schedules by such techniques as critical path analysis (see Section 4.2.2). The entire systems approach and its methods are strongly based on holistic thinking, which is particularly important in the selection of evaluation criteria, because the value of a particular solution can only be gauged after overall assessment of all of the expectations, requirements and constraints (see Section 3.3.2).

4. Method of Persistent Questions

When using systematic procedures it is often a good idea to keep asking questions of both oneself and of others as a stimulus to fresh thought and intuition. A standard list of questions also fosters the discursive method. In short, asking questions is one of the most important methodological tools. This explains why many authors have drawn up special checklists for various working steps to support this method.

5. Method of Negation

The method of deliberate negation starts from a known solution, splits it into individual parts or describes it by individual statements, and negates these statements one-by-one or in groups. This deliberate inversion often creates new solution possibilities. Thus, when considering a "rotating" machine element, one might also examine the "static" case. Moreover, the mere omission of an element can be tantamount to a negation. This method is also known as "systematic doubting" [2.21].

6. Method of Forward Steps

Starting from a first solution attempt, one follows as many paths as possible to produce further solutions. This method is also called the method of divergent thought. It is not necessarily systematic, but frequently starts with an unsystematic divergence of ideas. The method is illustrated in Figure 2.21 for the development of a shaft–hub connection. The arrows indicate the direction of the thinking process.

Such a thinking process can be improved by using classifying criteria (see Figure 3.18) to support the systematic variation of the characteristics (see Figure 3.21). Where variation is done without conscious thought, even with well-structured representations, the identified characteristics are not used to their full potential.

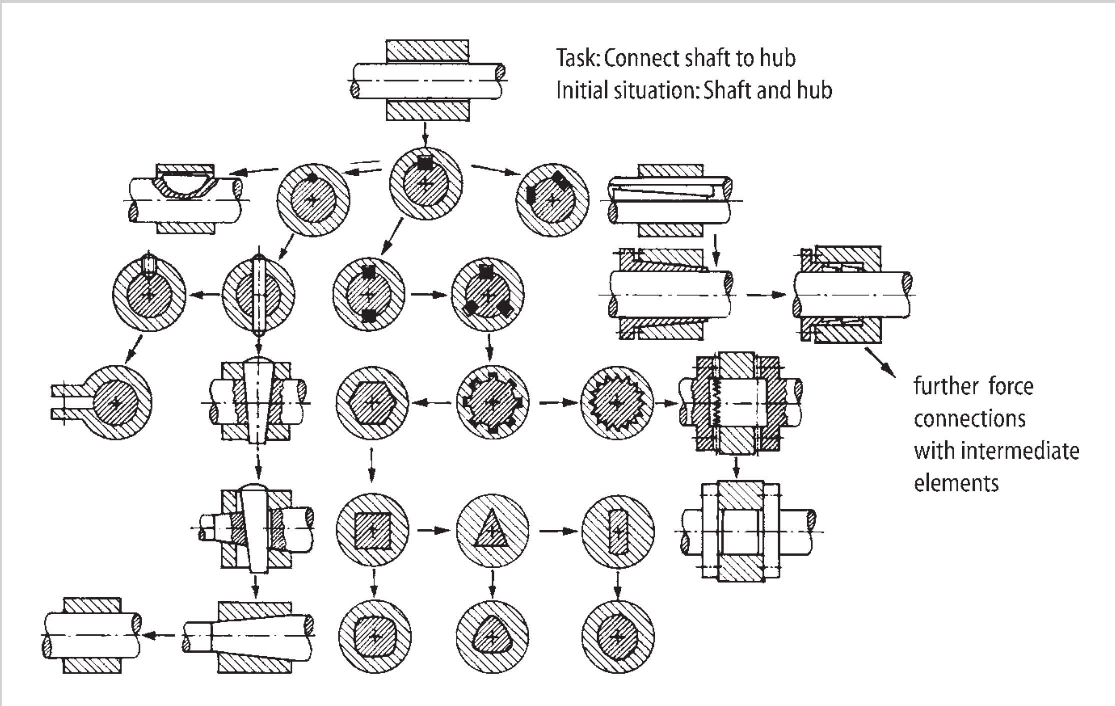


Figure 2.21. Development of shaft–hub connections in accordance with the method of forward steps

7. Method of Backward Steps

The starting point for this method is the goal rather than the initial problem. Beginning with the final objectives of the development, one retraces all of the possible paths that may have led up to it. This method is also called the method of convergent thought, because only ideas that converge on the ultimate goal are developed.

The method is particularly useful for drawing up production plans and developing systems for the production of components.

It is similar to the method of Nadler [2.38], who has proposed the construction of an ideal system that will satisfy all demands. This system is not developed in practice but formulated in the mind. It demands optimum conditions, such as an ideal environment which causes no external disturbances. Having formulated such a system, this is followed by a step-by-step investigation of what concessions must be made to turn this purely theoretical and ideal system into a technologically feasible one, and then finally into one that meets all the concrete requirements. Unfortunately, it is rarely possible to specify the ideal system in advance, because the ideal state of all functions, system elements and modules is difficult to specify, especially if they are linked together in a complex system.

8. Method of Factorisation

Factorisation involves breaking down a complex interrelationship or system into manageable, less complex and more easily definable individual elements (factors). The overall problem or task is divided into separate subproblems or subtasks that are, to a certain degree, independent (see Figure 2.3). Each of these subproblems or subtasks can initially be solved on its own, though the links between them in the overall structure must be kept in mind. Factorisation not only creates more manageable subtasks but it also clarifies their importance and influence in the overall structure, allowing priorities to be set. This approach is used in systematic design to divide an overall function into subfunctions and to develop function structures (see Sections 2.1.3 and 6.3), to search for working principles for subfunctions (see Section 6.4), and to plan the working steps during conceptual and embodiment design (see Section 4.2).

9. Method of Systematic Variation

Once the required characteristics of the solution are known, it is possible, by systematic variation, to develop a more or less complete solution field. This involves the construction of a generalised classification, that is, a schematic representation of the various characteristics and possible solutions (see Section 3.2.3). From the viewpoint of work science, too, it is obvious that the discovery of solutions is assisted by the construction and use of classification schemes. Nearly all authors consider systematic variation to be one of the most important methods.

10. Division of Labour and Collaboration

An essential finding of work science is that the implementation of large and complex tasks calls for the division of labour; more so as specialisation increases. This is also demanded by the increasingly tight schedules of modern industry. Now, division of labour implies interdisciplinary collaboration which, in turn, involves special organisational and staff arrangements along with appropriate staff attitudes, including receptiveness to the ideas of others. It must, however, be stressed that interdisciplinary collaboration and teamwork also demand a rigorous allocation of responsibility. Thus, the product manager should be in sole charge of the development of a particular product, regardless of departmental boundaries (see Section 4.3).

Systematic design, in combination with methods that make use of group dynamics, such as brainstorming, gallery method (see Section 3.2.3) and group evaluation (see Section 3.3), can overcome any lack of information exchange caused by the division of work, and can also help the search for solutions by stimulating ideas between team members.

2.2.6 Role of Computer Support

The systematic approach to design presented in this book can, in principle, be applied without the use of computers. However, the approach provides a sound basis for computer support of the design and development process that goes far beyond the use of complex analytical tools such as FEA and CFD, and the production of complex 3-D models. Computer support can be provided continuously throughout the process, for example, through the use of CAD, CAE, CAM, CIM, PDM and PLM software suites. The general use of IT also supports product improvement and reduces design and production effort.

It is not the purpose of this book, nor is there space, to describe the fundamental support that computers provide throughout the design process in detail. This topic is comprehensively covered in other texts, such as [2.1, 2.13, 2.17, 2.18, 2.27, 2.34, 2.37, 2.43, 2.44, 2.48, 2.52, 2.54, 2.57].

3 Product Planning, Solution Finding and Evaluation

A “methods toolbox” is presented in this chapter. Several of the methods, in particular the solution finding and evaluation ones, can be applied equally well in the different phases of the design process. Solution finding methods, such as brainstorming or the gallery method, can be useful, for example, in product planning and during conceptual design to find solution principles, as well as during embodiment design to find solutions for auxiliary functions. Evaluation methods can also be used in all of the phases. The only difference is the level of concretisation of the solutions under consideration.

Not every method is used in every product development process. Only those that seem appropriate for the problem situation and that contribute to a successful outcome are used. We provide recommendations for the practical application of each method to help the user assess its suitability in a given situation. Chapter 12 provides an overview of all recommendations.

3.1 Product Planning

One source of design and development tasks is a direct request (order) from a known client. This so-called business-to-business model [3.37, 3.47] is typical of made-to-order systems and process engineering equipment as well as for supply chain companies. For this type of order, there is a trend from client orientation to client integration [3.37], which has an influence on the work of the design and development department [3.2].

Assignments are set not only by clients, but increasingly—particularly in the case of original designs—they originate in the special planning departments of companies. In this case, designers are bound by the planning ideas of others (see Figure 1.2). Even then, however, the special skills of designers prove to be most useful in the medium- and long-term planning of products. The senior staff of the design department should therefore maintain close contacts not only with the production department, but also with the product planning department.

Planning can also be done by outside bodies, for instance by clients, by authorities, by consultancies, etc.

As will be discussed in Section 4.2 (see Figure 4.3), the design process for original designs starts with conceptualisation based on a requirements list (design specification). This preliminary list is usually based on requirements identified by product planning. It is therefore important for designers to know the essential points and steps of the product planning process. This will help them to understand the origin of the requirements and if necessary to add to the list. If there has not been a formal product planning phase, designers can organise the relevant steps using their own knowledge about product planning, or can undertake this phase themselves using simpler procedures.

In this chapter, and as shown in Figure 4.3, product planning and clarifying the task are consciously combined into one main phase. This is to emphasise the importance of integrating both activities. This remains important even when product planning and clarifying the task are undertaken separately within an organisation.

3.1.1 Degree of Novelty of a Product

As discussed in Section 1.1, the tasks of designers can have different degrees of novelty. The majority of tasks are adaptations to and variations on existing designs. This does not imply that these tasks are less challenging for designers. For product planning, the following differentiation of design tasks is of interest:

- *Original design:* New tasks and problems are solved using new or novel combinations of known solution principles. Two different cases can be distinguished:
  1. An invention is something truly new and is often based on the application of the latest scientific knowledge and insights [3.66].
  2. An innovation is a product that realises new functions and properties. This could be through novel or new combinations of existing solutions.
- *Adaptive design:* The solution principle remains unchanged; only the embodiment is adapted to new requirements and constraints.
- *Variant design:* The sizes and arrangements of parts and assemblies are varied within the limits set by previously designed product structures, which is typical of size ranges and modular products (see Chapter 9).

3.1.2 Product Life Cycle

Every product has a life cycle (see Figure 1.2), as illustrated in Figure 3.1. This is based on an economic viewpoint showing turnover, as well as profit and loss.

The *cycle time* depends strongly on the type of product and the branch of engineering, but in general cycles times are becoming shorter. This trend is likely to continue. This has a large effect on work in the design and development department because the time allocated for tasks that are identical, or very similar, to previous ones is reduced. As a consequence, it is necessary to adapt the product development process (see Chapter 4) as well as the methods discussed in this chapter.

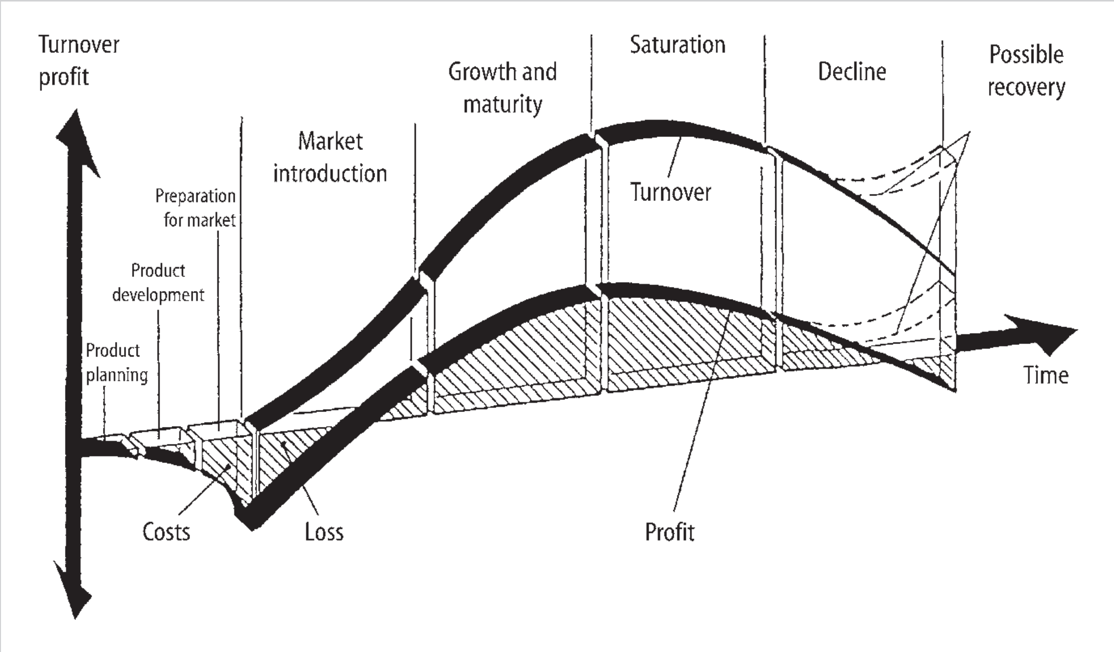


Figure 3.1. Life cycle of a product. After [3.45]

Measures to reactivate the market or generate new products have to be introduced when the saturation phase has been reached, at the latest. The introduction of these measures is an important task of product monitoring. A related activity in this context is the development of *market share*.

3.1.3 Company Goals and Their Effect

The main goal of every company is to make a profit. This goal has to be broken down into more concrete subgoals and related measures. To secure a lasting market presence, two generic strategies can be distinguished. The first strategy aims at achieving cost leadership. The corresponding company goals and implementation strategies are a broad sales base, large volumes, and rigorous product standardisation. The second strategy is that of performance differentiation. In this case, the goals and strategies focus on sales in special areas, highly effective flexible production, and specialisation in design and development. Both strategies have a time component, which is reflected in the company goal of being quicker to reach the market with a new product than its competitors.

One extreme strategy combines both strategies mentioned above, which, due to increasing competition, is becoming increasingly important.

Both of these goals—cost leadership and performance differentiation—affect the design and development department. At the next level down, many detailed goals are established, including those relating to the:

- *Product:* Such as functionality and properties
- *Market:* Such as time-to-market, which influences the time and budget made available (see Chapter 11) [3.12].

It is therefore very important for the design and development department to know the company's goals, their interrelationship and their relative importance. An important task for senior engineering managers is to convey the company goals relevant to engineering effectively to every member of staff.

3.1.4 Product Planning

1. Task and General Approach

Design and development start their work using a task description that, depending on the type of company, can come from different sources. In many cases, in particular in small- and medium-sized companies, it is left to the good sense of a director, or an individual member of staff, to develop and introduce the right product ideas at the right time and to formulate the necessary tasks. In larger companies, however, systematic procedures are increasingly used to find new products. An important aspect of this systematic approach is its potential to monitor the time and cost of product planning and product development more accurately. Those involved in product planning include marketing staff and product managers.

In many companies, therefore, the product planning department is expected to follow the development of a product idea through the design and production departments, and then to watch over its market behaviour. This includes monitoring the financial position and market success of the product and, if necessary, taking appropriate corrective measures (see Figure 1.2). In this book we shall only be dealing with product planning in the narrower sense, that is, as a preparation for product development.

The most important factor in finding new product ideas is client focus, which is increasingly directed towards client integration [3.2, 3.37]. One established method of identifying client wishes and translating these into product requirements is known as *Quality Function Deployment* (QFD) (see Section 10.5 [3.11, 3.38]).

Several systematic product planning approaches exist [3.5, 3.23, 3.33, 3.34, 3.42, 3.45, 3.69] and all of them have much in common (see Figure 3.2).

The stimuli for product plans come from outside (from the *market* or the *environment*) or from inside (from the *company* itself). These stimuli are usually identified by marketing.

Stimuli from the *market* include:

- the technical and economic position of the company's products in the market, in particular when changes occur, such as a reduction in turnover or a drop in market share
- changes in market requirements, for example new functions or fashions
- suggestions and complaints from customers
- the technical and economic superiority of competing products.

Stimuli from the *environment* include:

- economic and political changes, for example oil price increases, resource shortages, transport restrictions

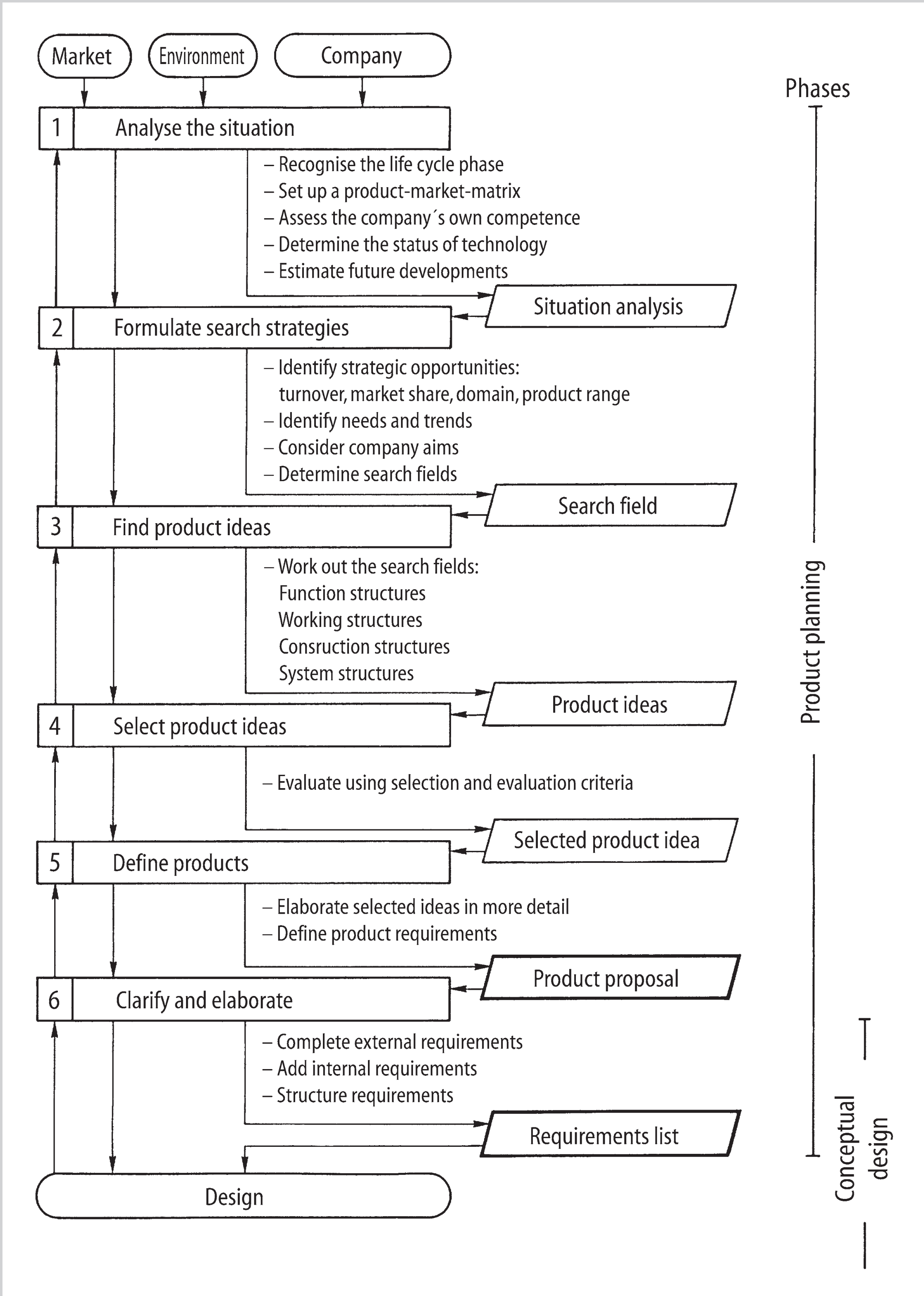


Figure 3.2. Product planning procedure. After [3.45, 3.69]

- new technologies and research results, for example microelectronics replacing mechanical solutions or laser cutting replacing flame cutting
- environmental and recycling issues.

Stimuli from within the *company* include:

- new ideas and results from company research applied during development and production
- new functions added to extend or satisfy the market
- the introduction of new production methods
- rationalisation of the product range and production
- increasing the degree of product diversification, that is, creating a range of products with life cycles that are intended to overlap.

These external and internal stimuli initiate five main working steps, which are illustrated, along with their outputs, in Figure 3.2.

These main working steps relate strongly to the general working methods described in Section 2.2 and more or less conform to systematic conceptual design (see Chapter 6 and Figure 4.3), and will be discussed in more detail in the following sections.

2. Analysing the Situation

The situation at the beginning of the product planning stage involves several aspects, and these must be clarified through a number of investigations, each with a different aim. The following steps have been found to be useful when analysing the situation, see also Figure 3.2.

Recognising the Life Cycle Phase

Consider the issues discussed in Sections 3.1.2 and 3.1.3. Life cycle analysis can also be used to recognise the need for diversification, in other words the phased development and sale of several different products. This will help to realise a balance of overlapping life cycles.

Setting Up a Product–Market Matrix

Recognising and clarifying the statuses of existing products from the company and from competitors in the various markets (field I in Figure 3.3) with respect to turnover, profit and market share should reveal the strengths and weaknesses of each of the products. A comparison with strong competitors is of particular interest.

Assessing the Company's Own Competence

This part of the analysis extends the previous one and provides the reasons for the current market position through an assessment of the company's technical weaknesses and through a comparison with competing companies (Figure 3.4). This analysis should not be based solely on orders, because these represent a selection that are already profitable for the company, but also on customer enquires and complaints, as well as installation and test reports.

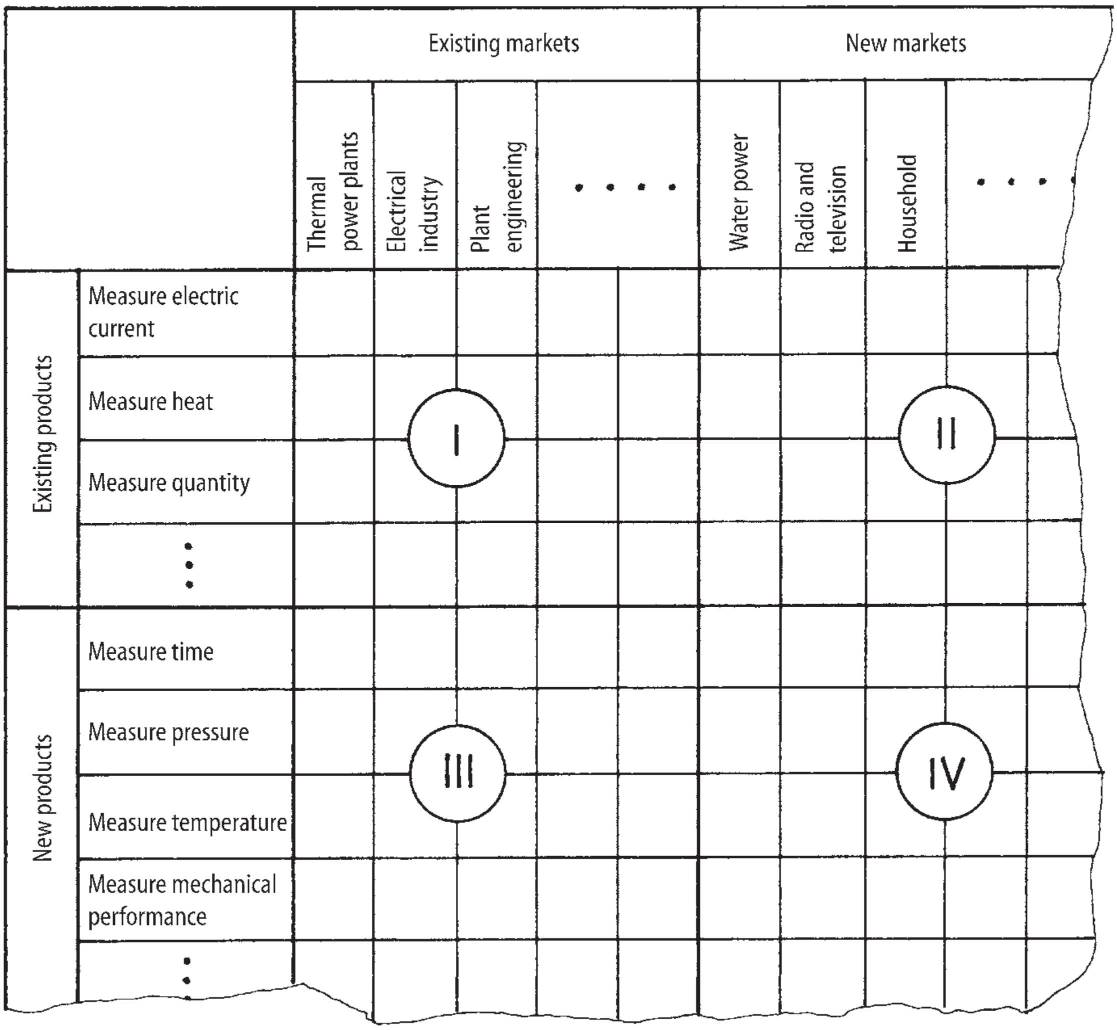


Figure 3.3. Product–market matrix, after [3.19] and [3.42], for a company producing measuring devices for industry

Determining the Status of Technology

This includes reviewing the products of the company, related technologies, concepts and products in the literature and patents, as well as competitors' products. In addition, the latest standards, guidelines and regulations are important.

Estimating Future Developments

Guidance can be obtained from knowledge of future projects, expected customer behaviour, technological trends, environmental requirements and the results of fundamental research.

A well-known method of visualising the technological situation, the international situation, the company situation and the competitive situation is portfolio analysis, which uses a multidimensional representation to present strategic business areas [3.38]. A distinction is made between the portfolios representing the present situation and the target situation. Figure 3.5 schematically shows a nine-cell portfolio matrix. It is also possible to use a simpler four-cell matrix. A distinction is made between business areas that are not profitable any more (cells 1, 2 and 3) and areas that should be targeted (cells 7, 8 and 9). If a business area is situated inbetween these (cells 4, 5 and 6), it is an indication that some action needs to be taken. Good examples of the factors labelled 1 and 2 in Fig-

Criteria	Competitors				
	A	B	C	D	.
Turnover					
Market share %					
Market situation					
– Conditions					
– Service					
– Delivery times					
–					
Product					
Management					
Product programme					

= same as us; + better; – worse

Figure 3.4. Analysis of competing companies. After [3.44]

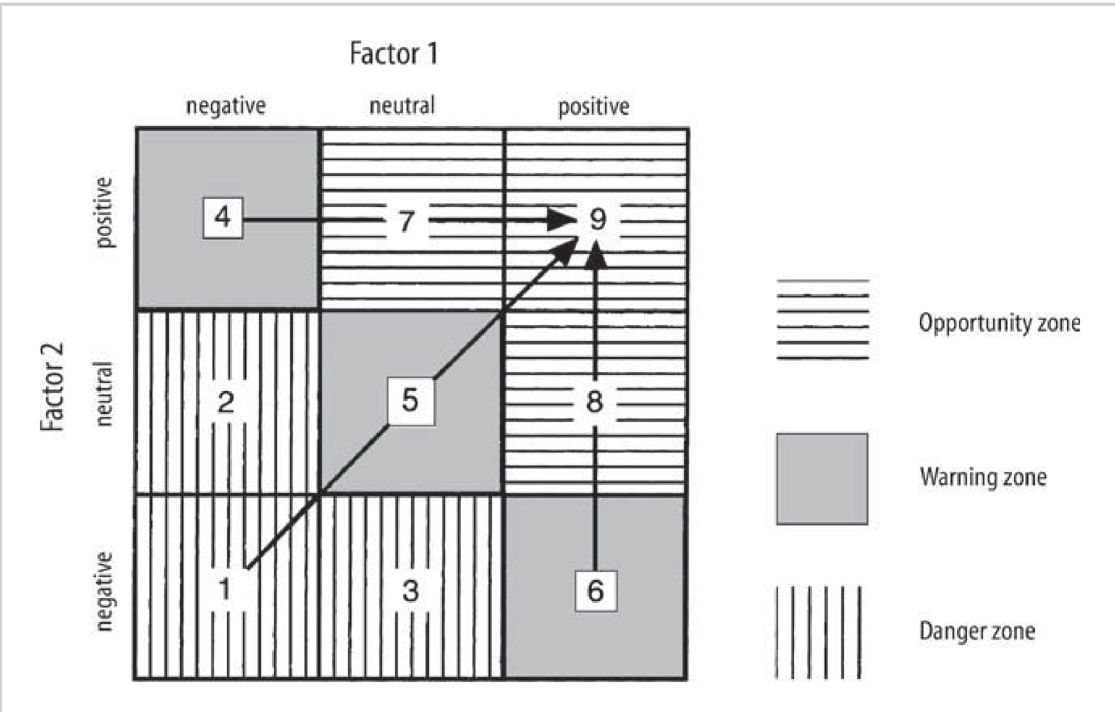


Figure 3.5. General structure of a portfolio matrix [3.21, 3.38, 3.45]

ure 3.5 would be: market growth—relative market share; market appeal—strength of competition; technological appeal—relative technological position; and market priority—technological priority [3.21].

Situation analysis determines the search strategies and the search fields that have to be addressed.

3. Formulating Search Strategies

Identifying Strategic Opportunities

It is possible that some gaps in the current product range or in the market are identified during the situation analysis. The task now is to determine which strategy to adopt: to introduce new products into the current market (field III in Figure 3.3); to open new markets with existing products (field II); or even to enter into new markets with new products (field IV). The latter involves the highest risk.

A promising gap that determines the search field [3.5, 3.33] must be found by taking into account the company's goals, strengths and market (see Table 3.1). Kramer [3.43] calls these strategic opportunities. They can relate to profit, market share, type of industry and product range. The weightings listed in Table 3.1 indicate that company goals are the most important criteria.

Table 3.1. Decision criteria for product planning

Criteria	Weighting
Company goals	$\geq$ 50%
Adequate financial cover	
High turnover	
High market growth	
Large market share (market leader)	
Short-term market opportunity	
Large functional advantages for users and excellent quality	
Differentiation from competitors	
Company strengths	$\geq$ 30%
Extensive know-how	
Favourable extension to range and/or product programme (diversification)	
Strong market position	
Limited need for investment	
Few sourcing problems	
Favourable rationalisation potential	
Market and other sources	$\geq$ 20%
Low danger of substitution	
Weak competition	
Favourable patent status	
Few general restrictions	

Identifying Needs and Trends

Most important for determining search fields is the identification of customer needs and market trends. Clues for these come from changes in customer behaviour caused, for example, by social developments such as environmental awareness, disposal problems, reduction in the working week, and transport problems. Another starting point could be changes in the length of the production supply chain, which can lead to new markets for suppliers. A commonly used tool is the *need–strength matrix* [3.42] (see Figure 3.6). In this matrix, one axis lists customer needs in decreasing order of importance, while the other lists the strengths and potentials of the company. The crossed fields in the top left corner of the matrix are the preferred search fields to be used in the preparation of the search field proposal. *Client–problem analysis* provides another tool [3.46].

Under subheading 1 of Section 3.1.4 we highlighted the importance of focussing on clients when planning new products and business areas. Here we describe an approach to achieve this objective. In the first step, the benefits currently required by the clients of a product or product group are extrapolated into the future. This is done to determine how the desired benefits are likely to change. If possible, all statements should be quantified, e.g. a noise reduction of 5 dB by the year 2006 and a reduction in energy consumption by 3 kW by 2007. In the second step, these requirements are allocated to suitable function carriers, i.e. assemblies or components. Next, the potential of the individual function carriers regarding the degree of fulfilment of future client requirements is estimated. In this step, requirements will

	Strengths and competences of the company (in decreasing order)					
Customer needs (in decreasing order)	Electronic measuring	Mechanical measuring	Electro-optics	Lithographic printing	Bonding	
Adaptive control, microprocessors	X			X	X	
Thermal control	X	X				
Remote control	X		X			
Traffic count	X		X			
Measure wind direction and velocity		X				

Figure 3.6. Need–strength matrix developed by a company, based on Figure 3.3 [3.42]

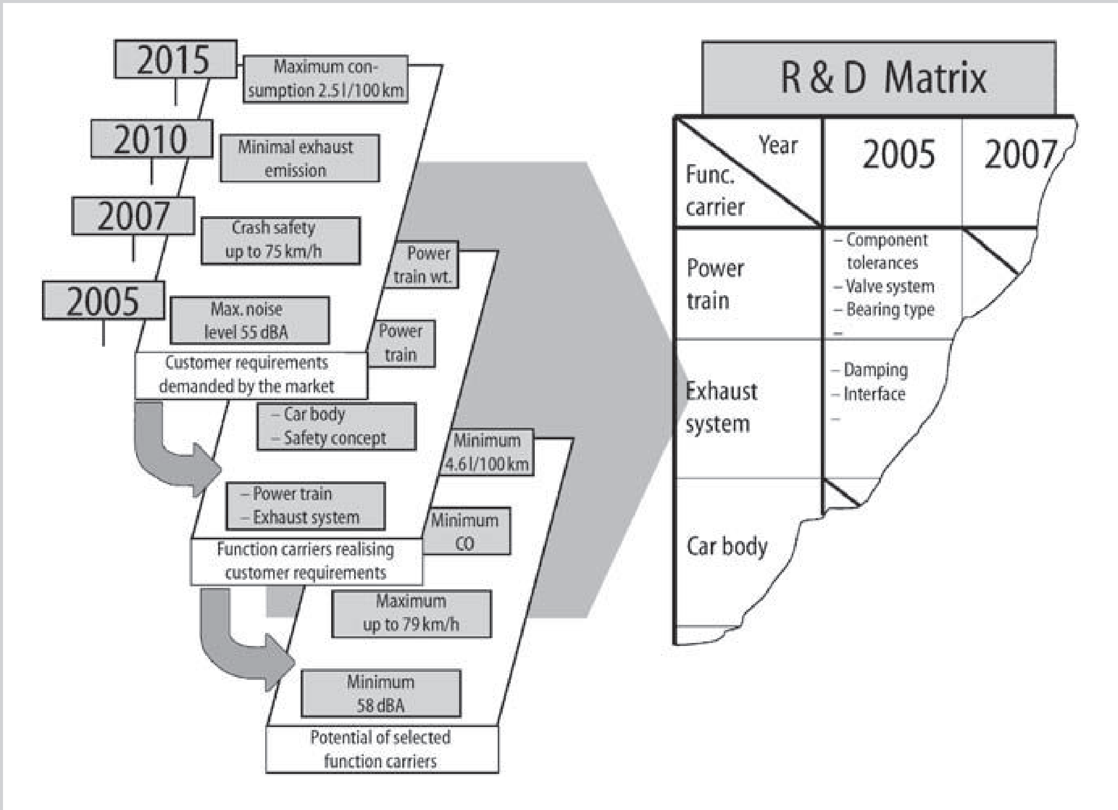


Figure 3.7. Product goals derived from customer requirements

arise for which no function carriers are yet available. As a result, this analysis can also reveal research and development needs regarding new and further developments of components, assemblies and products. It is useful to weight and prioritise the identified future requirements relating to client benefits. This will also rank the research and development themes. The results of these processes are the research and development goals that are used to plan the timing of research and development tasks. The product goals are also obtained. These goals provide the sequence in which new products, using the newly developed assemblies and components, are to be introduced into the market. Figure 3.7 illustrates this process.

For a medium- and long-term view of the future, scenario planning can be used to identify needs and trends, as well as profitable areas for the future [3.20, 3.21].

Considering Company Aims

Table 3.1 lists the goals and strengths of the company, which must be used to select a search field. The matrix in Figure 3.6 also emphasises the importance of the strengths and competences of the company in the selection of a worthwhile search field.

Determining Search Fields

The previously described steps of this product planning stage should lead, after a selection process, to a limited number of search fields (about 3–5 according to [3.22]) on which to concentrate the search for products.

4. Finding Product Ideas

The preferred search fields are now investigated in more detail using known search methods such as those that are used in product development (see Sections 3.2 and 6.4). These include: considering functions; intuitive methods such as brainstorming (see the so-called "idea-finding workshops" in [3.22]); and discursive methods such as ordering schemes, morphological charts and systematic combination.

When working out the search fields, a directed search for product ideas may be encouraged by the general relationships in technical products and their particular level of concretisation (see Section 2.1). Depending on the degree of novelty, the starting points for new products can be new product functions; other working principles; new embodiments; and rearrangements of an existing or new system structure. For a company producing measuring instruments, for example (see Figures 3.3 and 3.6), worthwhile product ideas can emerge from: new measuring functions; new physical effects (e.g. laser effect) used to fulfil known functions; or new embodiment goals (e.g. miniaturisation, better ergonomics and improved aesthetics).

The considerations follow the known interrelationship between function, working principle and embodiment:

Function:

- Which functions are required by the client?
- Which functions do we already fulfil?
- What complements existing functions?
- Which functions represent a generalisation of the existing ones?

For example, until now our company has only transported unit loads overland.

- What can we do in the future?
- Should we also use waterways?
- Should we start transporting very large, heavy items?
- Should we also transport bulk goods?
- Should we try to solve transport problems in general?

Working principle:

Existing products are based on a specific working principle. Would a change of working principle lead to better products?

Characteristics to look for are the types of energy and physical effects. For example, should a temperature-dependent flow-rate controller be based on the principle of fluid expansion, the bimetallic effect or the use of microprocessor-controlled temperature probes?

Embodiment:

- Is the space used still appropriate?
- Should we focus on miniaturisation?

- Is the shape still appealing?
- Could the ergonomics be better?

For example, is it still appropriate to use laces in shoes? Would Velcro or hooks be more appealing and more comfortable?

The answers to these questions determine the novelty of the product idea and therefore the developmental risks.

5. Selecting Product Ideas

The product ideas generated are first subjected to a selection procedure (see Section 3.3.1). For this initial selection, the criteria linked to the company's goals are sufficient in so far as they can be determined (see Table 3.1). At the very least, high turnover, large market share and functional advantages for the customer should be taken into account. A more detailed selection involves the other criteria. To identify promising product ideas, it is often sufficient, in the sense of an efficient application of selection procedures, to work only with binary values (yes/no).

6. Definiting Products

In this step, product ideas that seem promising are elaborated more concretely and in more detail. It is useful to consider the characteristics of requirements lists used in product development (see Section 5.2). During this step, at the latest, sales, marketing, research, development and design should work actively together. This can be encouraged by involving these groups in the evaluation and selection of product ideas.

Product ideas, after elaboration, are then subjected to an evaluation in which all of the criteria listed in Table 3.1, as far as they are known, are used.

Often some criteria, such as investment needs or sourcing problems, cannot be assessed because they are solution dependent. In these cases they will not be considered during this step. The best product definitions are given to the product development department as a product proposal together with a preliminary requirements list. The product development department then develops the actual product, using, for example, the systematic approach we propose.

The product proposal should:

- Describe the intended functions.
- Contain a preliminary requirements list that should have been compiled as far as possible using the characteristics used later to clarify the task and finalise the requirements list.
- Formulate all requirements in a solution-neutral way. The working principle should only be determined in so far as it is really necessary from the point of view of the overall functionality. For example, the same working principle

will be specified when an existing product range is being extended. Suggestions for working principles, however, should always be indicated, in particular when suitable solution principles have emerged during the idea-finding step. These should not prejudice product development (see also the solution-neutral formulation of requirements).

- Indicate a cost target or a budget linked to the company's goals which clarifies future intentions such as production volume, extensions to the product range, new suppliers, etc.

This concludes the product planning phase. By using the listed decision criteria, only those proposals that are likely to fit the company's goals and strengths, and that match the macro- and microeconomic situations, should enter the development stage. The development of the requirements list using the same method that will be applied in product development ensures an easy and seamless transition from product planning to product development.

For successful product planning and development, it is important that both groups work together using the same methods and similar evaluation and decision criteria. At the latest, product development should be actively involved when product ideas are selected and the product is defined. Together they should also develop the requirements list in a format suitable for product development (see Section 5.2).

7. Product Planning in Practice

Because of strong competition, new products have to meet market needs closely, be produced at a competitive cost and be economical to use. In addition, requirements relating to disposal and recycling, and to low environmental impact during production and use, are becoming increasingly important. Products with such complex requirements need to be planned systematically to meet these demands. Just relying on spontaneous ideas or incremental developments to existing products will not, in general, fulfil these demands. Systematic product planning often uses the same methods as concept development, and staff can usefully be exchanged between the two departments.

The following guidelines are important:

- The size of the company determines whether or not it is possible to set up interdisciplinary project groups or departments. In smaller companies it might be necessary to involve external consultants to supply expertise that is missing in the company.

- To use company expertise, however, can involve less risk and often increases client confidence.

- If product planning focuses on existing product lines, in other words further development or systematic variation, the development department responsible for the product line can monitor the new product, or this can be done by a special planning group that includes members from that department.

- When product planning takes place outside an existing product line, in other words the focus is on completely new products or diversification of the product programme, it is better to set up a new planning group. This group works on "innovative planning" and can either be set up as a permanent department or as a temporary working group.
- More elaborate analysis and conscious thought is required when planning for new markets than when dealing with known sales channels and existing client circles.
- When the starting situation is complex, it can be useful to undertake product planning and development using a stepwise and iterative approach. Acquisition of information and the decision making steps should be scheduled such that the anticipated effort and success can be reviewed and planned.
- Even when product ideas have been generated intuitively, a situation analysis and a feasibility study using the search strategies should still be performed.
- To identify customer problems, it is useful to have intensive collaboration with a few leading clients, referred to as "lead users" [3.22]. QFD methods can be used here too [3.11, 3.38].
- When new products are introduced, technical failures and weaknesses can have a far-reaching impact on the reputation of such products. Part of a careful product planning process, therefore, should include sufficient time for testing and the calculation of risks (see Section 7.5.12).
- Entry into the market later than announced can also have a negative effect on reputation because it suggests technical problems.
- During the planning and introduction of new products, it is useful to have a powerful product champion, e.g. a board member who identifies personally with the new product. This helps overcome a potential lack of interest and conventional resistance [3.22].
- Scenario planning (see [3.20, 3.22]) is particularly suitable for long-term forecasts. The effort required for scenario preparation, scenario field analysis, scenario forecasts and scenario building, however, is only worthwhile for business areas that are important to the company and its survival.

Finally, it should be stated that the procedure shown in Figure 3.2 does not represent a straight path with sequential steps, but a guideline for obtaining an essentially purposeful approach. The practical application of this approach will require an iterative procedure in which forward and backward steps at higher levels of information are necessary. This is quite normal in successful product finding.

3.2 Solution Finding Methods

The main advantage of the systematic approach is that designers do not have to rely on coming up with a good idea at the right moment. Solutions can be systematically elaborated using the relevant methods. These methods are the subject of this chapter.

An optimal solution:

- fulfils all demands in the requirements list as well as most of the wishes
- can be realised by the company within the constraints of budget (target costing), time-to-market, production facilities, etc.

Several steps are required to realise such a solution.

First, a range of possible solutions for the given task has to be generated. The basis for this is the function structure (see Section 2.1.3) that is used to divide the overall task into manageable subtasks. The function structure also provides the functional interrelationship between the subtasks, by describing the relationship between the inputs and outputs of each subfunction with respect to the flows of material, energy and signals.

In a second step, one or more possible physical effects are assigned to each of these solution-neutral subfunctions in order to realise them. This is done in accordance with the task-specific requirements. To realise a certain force, for example, a physical effect with the appropriate capability needs to be selected.

The approach described thus far typifies the traditional approach of an engineer. A solution space is created because variants are generated while developing the function structure and when selecting physical effects.

The use of a combination of solution-finding methods can be used to extend the solution space.

Often a subfunction can only be realised through a combination of several physical effects. This is another reason to use several solution finding methods. Those that are proposed or described in the following sections originate from, among others, the area of creativity techniques with its generally recurring methods that are described in Section 2.2.5. Others are based on analogical or logical reasoning.

The methods described here are mainly intended for the design and development of new products. However, they can be very helpful when existing patents of a competitor have to be circumvented or when existing products or components have to be optimised. The methods have to be selected for, adapted to and used in accordance with the context of the problem.

3.2.1 Conventional Methods

1. Information Gathering

For designers, access to state-of-the-art information is essential. As a first step, designers use a variety of collection techniques [3.45]. Information and data repositories, along with systems used to search and process the data, assist the active search for and the passive discovery of solutions. The internet enables a more effective and efficient application of the following conventional techniques:

- searching the literature
- analysing trade publications

- surveying the presentations from exhibitions and fairs
- assessing catalogues of competitors
- exploring patents, etc.

2. Analysis of Natural Systems

The study of natural forms, structures, organisms and processes can lead to very useful and novel technical solutions. The connections between biology and technology are investigated by bionics and biomechanics. Nature can stimulate the creative imagination of designers in a host of different ways [3.6, 3.29, 3.31, 3.35].

Technical applications of the design principles of natural forms include lightweight structures employing honeycombs, tubes and rods, the profiles of aircraft and ships, and the take-off and flying characteristics of aircraft. Lightweight structures in the form of thin stems are very important (see Figure 3.8). Another technical application is sandwich construction, and Figure 3.9 shows a few derivations of this natural principle that have proved useful in aircraft construction.

The hooks of a burr provided a solution that was incorporated into the Velcro fastener (see Figure 3.10). Further examples are given in Figure 3.11.

Fibre composites can be used to optimise the stiffness and deformation of structures that can equal or exceed those in found in nature. Carbon, glass and plastic fibres are aligned according to the principal stress directions and embedded in a predominantly polymer matrix of polyester, epoxy and other resins. This construction method requires an in-depth stress analysis along with a laying-up technique for the fibres adapted to that analysis, as well as extensive knowledge of plastics to select the fibre matrix composite. The basic relationships and ideas for the correct design of fibre composites and numerous literature references are provided by Flemming et al. [3.16].

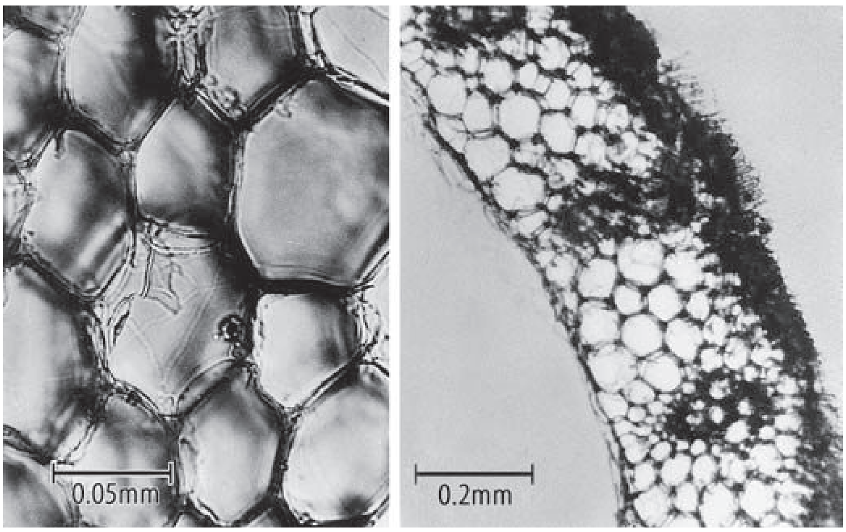


Figure 3.8. Wall of a wheat stem [3.29]

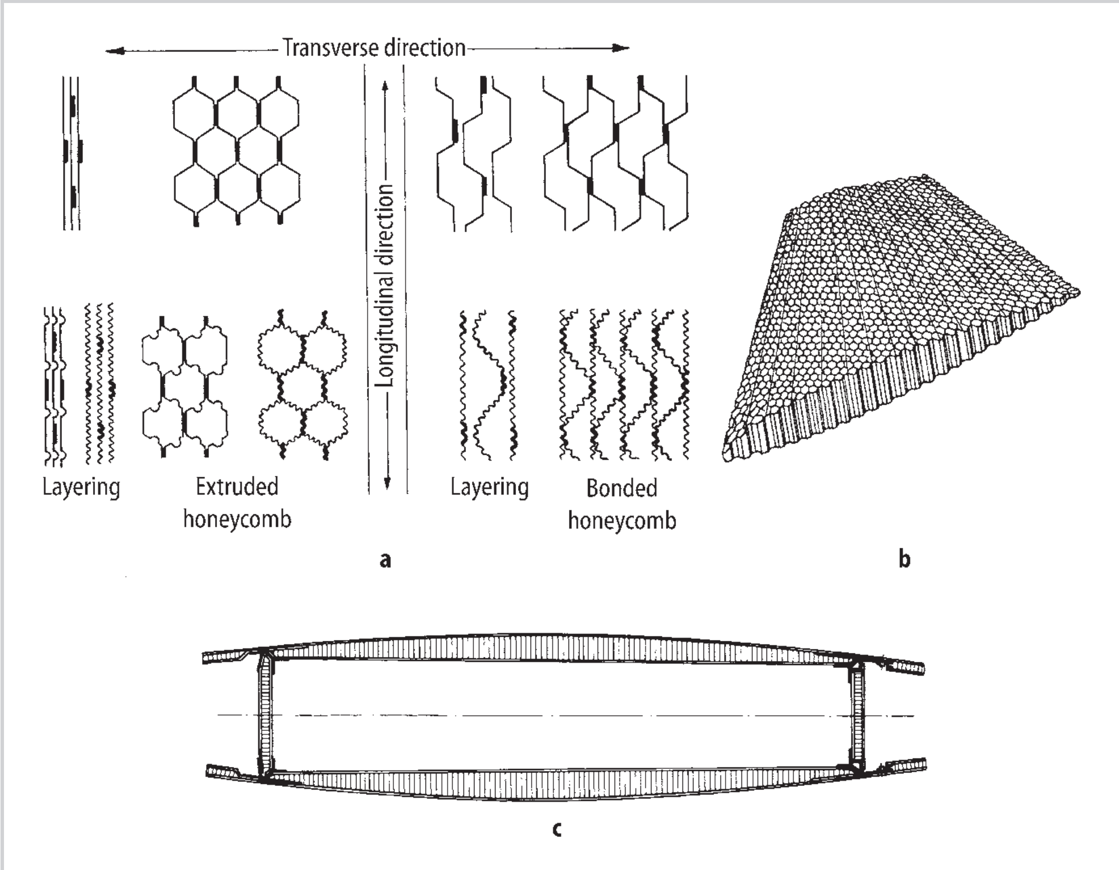


Figure 3.9. Sandwich construction for lightweight structures [3.30]. **a** A few honeycomb structures. **b** Completed honeycomb structure. **c** Sandwich box girder

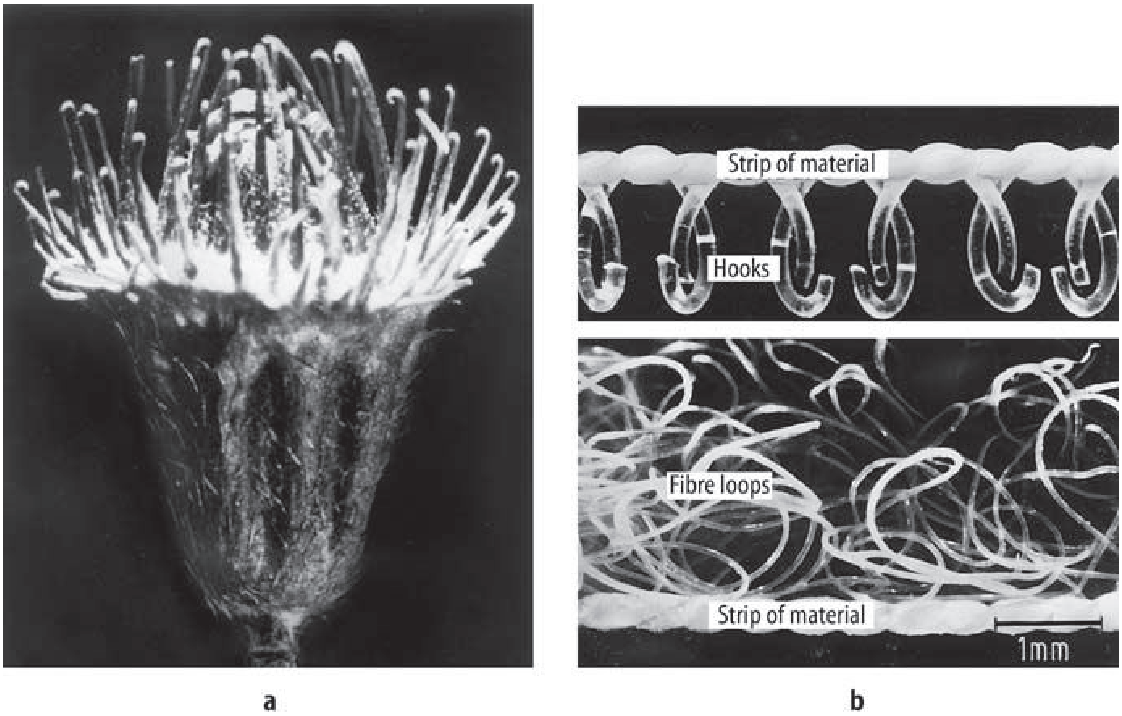


Figure 3.10. a Hooks of a burr. **b** Velcro fastener. After [3.29]

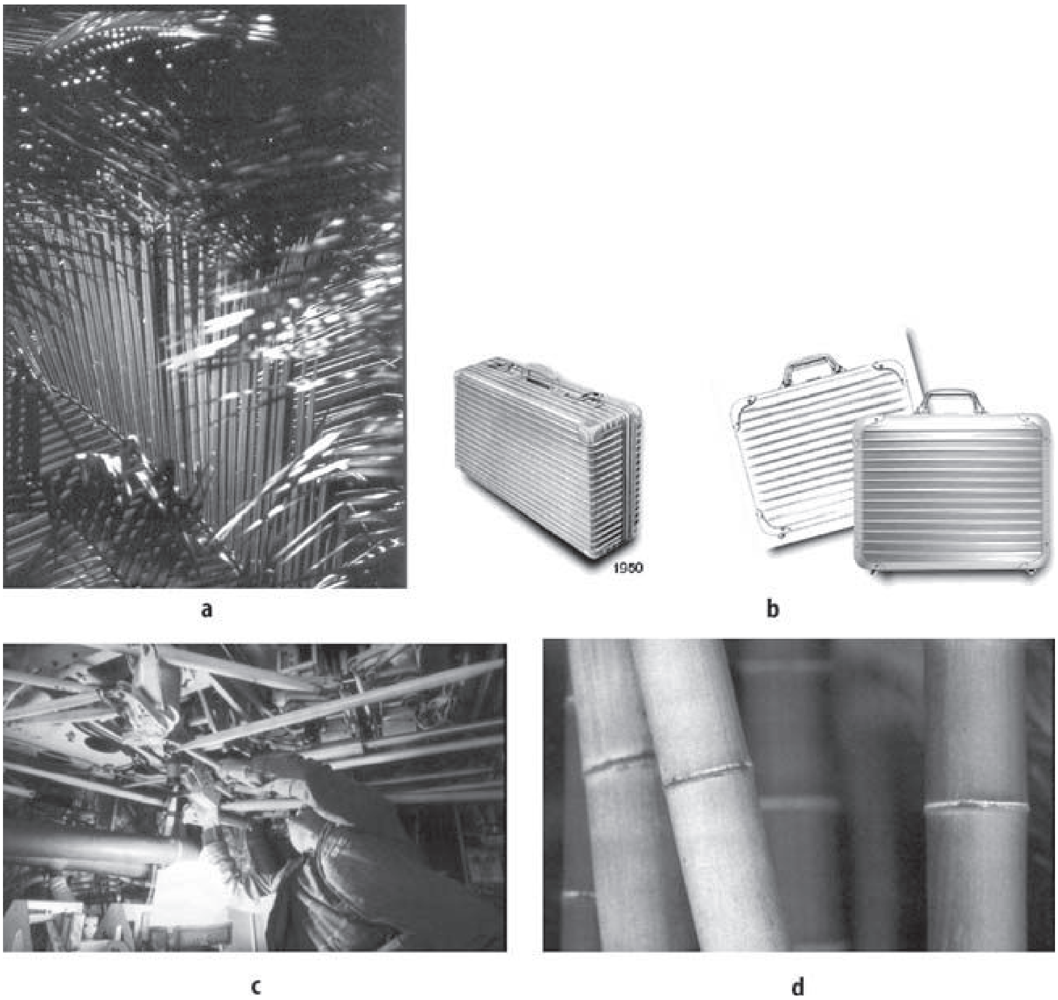


Figure 3.11. **a** Palm leaves (Lufthansa publication 2/96). **b** Aluminium suitcase (Rimowa Kofferfabrik 10/01). **c** Tubular structure in an aircraft. **d** Bamboo stems (Lufthansa publication 5/96)

3. Analysis of Existing Technical Systems

The analysis of existing technical systems is one of the most important means of generating new or improved solution variants in a step-by-step manner.

This analysis involves the mental or even physical dissection of finished products. It may be considered a form of structure analysis (see Subsection 1 in Section 2.2.5) aimed at the discovery of related logical, physical and embodiment design features. Figure 6.10 shows an example of this type of analysis. Here, subfunctions were derived from the existing configuration. From them, further analysis led to the identification of the physical effects involved which, in turn, might have suggested new solution principles for corresponding subfunctions. It is also possible to adopt solution principles discovered during the analysis.

Existing systems used for analysis might include:

- products or production methods from competing companies
- older products and production methods from one's own company
- similar products or assemblies in which some subfunctions or parts of the function structure correspond to those for which a solution is being sought.

Because the only systems to be analysed are those that have some bearing on the new problem as a whole or on parts of it, we could call this way of collecting information the systematic exploitation of proven ideas, or of experience. It proves particularly helpful for finding a first solution concept as a starting point for further variations. It must, however, be said that this approach carries the danger of causing designers to stick with known solutions instead of pursuing new paths.

4. Analogies

In the search for solutions and in the analysis of system properties, it is often useful to substitute an analogous problem (or system) for the one under consideration, and to treat it as a model. In technical systems, analogies may be obtained, for instance, by changing the type of energy used [3.3, 3.64]. Analogies chosen from the nontechnical sphere may prove very useful as well.

Besides helping in the search for solutions, analogies are also helpful in the study of the behaviour of a system during the early stages of its development by means of simulation and model techniques, and in the subsequent identification of essential new subsolutions and the introduction of early optimisations.

If the model is to be applied to systems of markedly different dimensions and conditions, a supportive similarity (dimensional) analysis should be undertaken (see Section 9.1.1).

5. Measurements and Model Tests

Measurements on existing systems, model tests supported by similarity analyses and other experimental studies are among the most important sources of information. Rodenacker [3.59] in particular stresses the importance of experimental studies, arguing that design can be interpreted as the reversal of physical experiment.

In the precision engineering and mass production industries, including those where micromechanisms and electronic products are developed, experimental investigations are an important and established means of arriving at solutions. This approach has organisational repercussions since, in the creation of such products, experimental development is often incorporated within the design activity (see Figure 1.3).

In a similar way, the testing and subsequent modification of software solutions belong to this empirically based group of methods.

3.2.2 Intuitive Methods

Designers often seek and discover solutions for difficult problems by intuition—that is, solutions come to them in a flash after a period of search and reflection. These solutions suddenly appear as conscious thoughts and often their origins cannot be traced. As Galtung of the International Peace Research Institute in Oslo has put it: “The good idea is not discovered or undiscovered; it comes, it happens”.

It is then developed, modified and amended, until such time as it leads to the solution of the problem.

Good ideas are always scrutinised by the subconscious or preconscious in the light of expert knowledge, experience and the task in hand, and often the simple impetus resulting from the association of ideas suffices to force them into consciousness. That impetus can also come from apparently unconnected external events or discussions. Frequently, a sudden idea will hit the bull's eye, so that all that needs to be done is to make changes or adaptations that lead straight to a final solution. If that is indeed the case and a successful product is created, then this represents the optimum procedure. Very many good solutions are born in that way and developed successfully. A good design method, far from trying to eliminate this process, should serve to back it up.

An industrial concern should nevertheless beware of exclusive reliance on the intuition of its designers, nor should designers themselves leave everything to chance or rare inspiration. Purely intuitive methods have the following disadvantages:

- The right idea does not always come at the right time, since it cannot be forced.
- Current conventions and personal prejudices may inhibit original developments.
- Because of inadequate information, new technologies or procedures may fail to reach the consciousness of the designer.

These dangers increase with specialisation, the division of tasks and with time pressure.

There are several methods of encouraging intuition and opening new paths by the association of ideas. The simplest and most common of these involves critical discussions with colleagues. Provided that such discussions are not allowed to stray too far and are based on the general methods of persistent questions, negation, forward steps, etc. (see Section 2.2.5), they can be very helpful and effective.

Methods with an intuitive bias such as Brainstorming, Synectics, Gallery Method, Method 635 and many others involve group dynamics that are used to generate the widest possible range of ideas. One of the effects of group dynamics is the uninhibited exchange of associated ideas between the members.

Most of these techniques were originally devised for the solution of nontechnical problems. They are, however, applicable to any field that demands new, unconventional ideas.

1. Brainstorming

Brainstorming can be described as a method of generating a flood of new ideas. It was originally suggested by Osborn [3.51] and provides conditions in which a group of open-minded people from as many different spheres of life as possible bring up, without prejudice, any thoughts that occur to them and thus trigger off new ideas in the minds of the other participants [3.74]. Brainstorming relies strongly on stimulation of the memory and on the association of ideas that have never been considered in the current context or have never been allowed to reach consciousness.

For maximum effect, brainstorming sessions should be run along the following lines:

Composition of the Group

- The group should have a leader and consist of a minimum of five and a maximum of 15 people. Fewer than five constitute a spectrum of opinion and experience that is too small, and hence produce too few stimuli. With more than 15, close collaboration may decline because of individual passivity and withdrawal.
- The group must not be confined to experts. It is important that as many fields and activities as possible are represented, the involvement of nontechnical members adding a rich new dimension.
- The group should not be hierarchically structured but, if possible, made up of equals in order to prevent the censoring of such thoughts as might give offence to superiors or subordinates.

Leadership of the Group

- The leader of the group should only take the initiative when dealing with organisational problems (invitation, composition, duration and evaluation). Before the actual brainstorming session, the leader must outline the problem and, during the session, must see to it that the rules are observed and, in particular, that the atmosphere remains free and easy. To that end the leader should start the session by expressing a few absurd ideas, or mentioning an example from another brainstorming session, but should never lead in the expression of ideas. On the other hand, the flow of new ideas should be encouraged whenever the productivity of the group slackens. The leader must ensure that no one criticises the ideas of other participants, and should appoint one or two members to take minutes.

Procedure

- All participants must try to shed their intellectual inhibitions; that is, they should avoid rejecting as absurd, false, embarrassing, stupid, well-known or redundant any ideas expressed spontaneously by themselves or by other members of the group.
- No participant should criticise any ideas that are brought up, and everyone must refrain from using such killer phrases as “we’ve heard it all before”, “it can’t be done”, “it will never work” and “this has nothing to do with the problem”.
- New ideas will be taken up by the other participants, who may change and develop them at will. It is also useful to combine several ideas into new proposals.
- All ideas should be written down, sketched out, or recorded.
- All suggestions should be concrete enough to allow the emergence of specific solution ideas.
- The practicability of the suggestions should be ignored at first.

- A session should not generally last for more than 30 to 45 minutes. Experience has shown that longer sessions produce nothing new and lead to unnecessary repetitions. It is better to make a fresh start with new ideas or with other participants later.

Evaluation

- The results should be reviewed by experts to find potential solution elements. If possible, these should be classified and graded in order of feasibility and then developed further.
- The final result should be reviewed with the entire group to avoid possible misunderstandings or one-sided interpretations on the part of the experts. New and more advanced ideas may well be expressed or developed during such a review session.

Brainstorming is indicated [3.56] whenever:

- No practical solution principle has been discovered.
- The physical process underlying a possible solution has not yet been identified.
- There is a general feeling that deadlock has been reached.
- A radical departure from the conventional approach is required.

Brainstorming is even useful in the solution of subproblems arising in known or existing systems. Moreover, it has a beneficial side-effect: all of the participants are supplied with new data, or at least with fresh ideas on possible procedures, applications, materials, combinations, etc., because the group represents a broad spectrum of opinion and expertise (for instance, designers, production engineers, sales persons, materials experts and buyers). It is astonishing what a profusion and range of ideas such a group can generate. The designers will remember the ideas brought up during brainstorming sessions on many future occasions. Brainstorming triggers off new lines of thought, stimulates interest and represents a break in the normal routine.

It should, however, be stressed that no miracles must be expected from brainstorming sessions. Most of the ideas expressed will not be technically or economically feasible, and those that are will often be familiar to the experts. Brainstorming is meant first of all to trigger off new ideas, but it cannot be expected to produce ready-made solutions because problems are generally too complex and too difficult to be solved by spontaneous ideas alone. However, if a session should produce one or two useful new ideas, or even some hints in what direction to go looking for the solution, it will have achieved a great deal.

An example of a solution obtained by Brainstorming can be found in Section 6.6, which also shows how the resulting ideas were evaluated and how classifying criteria for the subsequent search for solutions were derived from them.

2. Method 635

Brainstorming has been developed into Method 635 by Rohrbach [3.60]. After familiarising themselves with the task, and after careful analysis, each of six participants is asked to write down three rough solutions in the form of keywords. After some time, the solutions are handed to each participant's neighbour who, after reading the previous suggestions, enters three further solutions or developments. This process is continued until each original set of three solutions has been completed or developed through association by the five other participants, hence the name of the method.

Method 635 has the following advantages over Brainstorming:

- A good idea can be developed more systematically.
- It is possible to follow the development of an idea and to determine more or less reliably who originated the successful solution principle, which might prove advisable for legal reasons.
- The problem of group leadership rarely arises.

The method has the following disadvantage:

- Reduced creativity by the individual participants owing to isolation, and lack of stimulation in the absence of overt group activity.

3. Gallery Method

The Gallery Method developed by Hellfritz [3.27] combines individual work with group work, and is particularly suitable for any stage of the design process where solution proposals can be expressed in the form of sketches or drawings. The organisation and team building are similar to Brainstorming. The method consists of the following steps.

Introduction Step: The group leader presents the problem and explains the context.

Idea Generation Step 1: For 15 minutes the individual group members create solutions intuitively and without prejudice using sketches supported, where necessary, by text.

Association Step: The results from idea generation step 1 are hung on a wall as in an art gallery so that all group members can see and discuss them. The purpose of this 15-minute association step is to find new ideas or to identify complementary or improved proposals through negation and reappraisal.

Idea Generation Step 2: The ideas and insights from the association step are further developed individually by each of the group members.

Selection Step: All ideas generated are reviewed, classified and, if necessary, finalised. Promising solutions are then selected (see Section 3.3.1). It is also possible to identify potential solution characteristics that can be developed later using a discursive method (see Section 3.2.3).

The Gallery Method has the following advantages:

- Intuitive group working takes place without unduly lengthy discussions.
- An effective exchange of ideas using sketches is possible.
- Individual contributions can be identified.
- Documentary records are easily assessed and stored.

4. Delphi Method

In this method, experts in a particular field are asked for written opinions [3.7].

The requests take the following form:

First Round: What starting points for solving the given problem do you suggest? Please make spontaneous suggestions.

Second Round: Here is a list of various starting points for solving the given problem. Please go through this list and make what further suggestions occur to you.

Third Round: Here is the final evaluation of the first two rounds. Please go through the list and write down what suggestions you consider most practicable.

This elaborate procedure must be planned very carefully and is usually confined to general problems bearing on fundamental questions or on company policy. In the field of engineering design, the Delphi Method should be reserved for fundamental studies of long-term developments.

5. Synectics

Synectics is a word derived from Greek and it refers to the activity of combining various and apparently independent concepts. Synectics is comparable to Brainstorming, with the difference that its aim is to trigger off fruitful ideas with the help of analogies from nontechnical or semi-technical fields.

The method was first proposed by Gordon [3.25]. It is more systematic than Brainstorming, with its arbitrary flow of ideas. However, both methods call for complete frankness and lack of inhibition or criticism.

A synectics group should consist of no more than seven members, otherwise the ideas expressed will run away with themselves. The leader of the group has the additional task of helping the group to develop the proposed analogies by guiding them through the following steps:

- Presentation of the problem.
- Familiarisation with the problem (analysis).
- Grasping the problem.
- Rejection of familiar assumptions with the help of analogies drawn from other spheres.
- Analysis of one of the analogies.
- Comparison of the analogy with the existing problem.
- Development of a new idea from that comparison.
- Development of a possible solution.

If the result is unsatisfactory, the process may have to be repeated with a different analogy.

An example may help to illustrate this method. In a seminar set up for the purpose of discovering the best method of removing urinary calculi from the human body, several mechanical devices for gripping, holding and extracting these stones were mentioned. The device would have to stretch and open up inside the urethra. The keywords “stretch” and “open up” suggested the idea of an umbrella to one of the participants (see Figure 3.12).

Question: How can the umbrella analogy—(a) in Figure 3.12—be applied?

Possible answer 1: By (b) drilling through the stone, pushing the umbrella through the hole and opening it up. Not very feasible.

Possible answer 2: By (c) pushing a tube through the hole and blowing it up (balloon) behind the stone. Drilling of hole not feasible.

Possible answer 3: By (d) pushing the tube past the stone. When the tube is withdrawn the resistance may seriously damage the urethra.

Possible answer 4: By (e) adding a second balloon as a guide and by (f) embedding the stone in a gel between the two balloons and then pulling it out? This was found to be the best solution.

This example shows the association with a semi-technical analogy (umbrella) from which a solution was developed that took into account the special constraints that existed in this case. The solution shown here is not the final solution resulting from the seminar but represents an example of how the method was used.

Characteristic of this approach is the unrestricted use of analogies which, in the case of technical problems, are selected from nontechnical or semi-technical spheres. Such analogies will generally suggest themselves quite spontaneously at the first attempt but, during subsequent development and analysis, they will generally be derived more systematically.

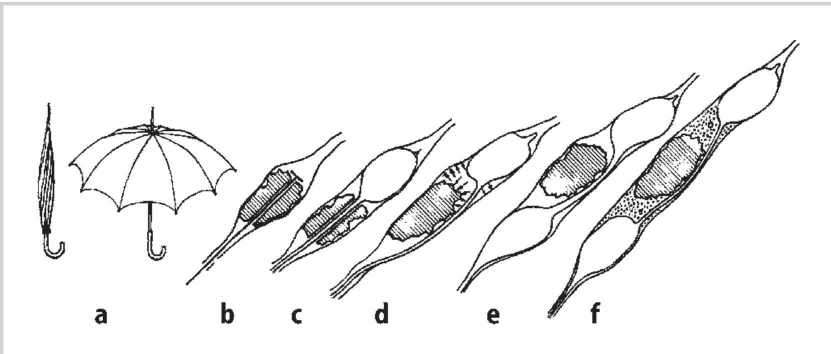


Figure 3.12. Step-by-step development of a solution principle for the removal of urinary calculi based on analogy and stepwise improvement

6. Combination of Methods

Any one of these methods taken by itself may not lead to the required goal. Experience has shown that:

- The group leader of, or another participant in, a brainstorming session may, when the flow of ideas dries up, introduce synectic procedures—deriving analogies, rejection of familiar assumptions, etc.—to release a new flood of ideas.

- A new idea or an analogy may radically change the approach and ideas of the group.
- A summary of what has been agreed so far may lead to new ideas.
- The explicit use of the methods of negation and reappraisal and of forward steps (see Section 2.2.5) can enrich and extend the variety of ideas.

In the seminar we mentioned, the presentation of the idea "destroy stone" produced a host of new suggestions, such as drilling, smashing, hammering, ultrasonic disintegration and so on. When the flow of ideas eventually dried up, the group leader asked, "How does nature destroy?", which immediately evoked a number of new suggestions, including weathering, heating and cooling, decay, putrefaction, bacterial action, ice expansion and chemical decomposition. A combination of the two principles "clasp stone" and "destroy stone" provoked the question, "What else?" This produced the answer "contact stone rather than clasp", which in turn threw up such new ideas as sucking, gluing, and applying various contact forces.

The different methods should be combined so as to best address particular cases. A pragmatic approach ensures the best results.

3.2.3 Discursive Methods

Methods with a discursive bias provide solutions in a deliberate step-by-step approach that can be influenced and communicated. Discursive methods do not exclude intuition, which can make its influence felt during individual steps and in the solution of individual problems, but not in the direct implementation of the overall task.

1. Systematic Study of Physical Processes

If the solution of a problem involves a known physical (chemical, biological) effect represented by an equation, and especially when several physical variables are involved, various solutions can be derived from the analysis of their interrelationships, that is, of the *relationship* between a dependent and an independent variable, all other quantities being kept constant. Thus, if we have an equation in the form $y = f(u, v, w)$, then, according to this method, we investigate solution variants for the relationships $y_1 = f(u, \underline{v}, \underline{w})$, $y_2 = f(\underline{u}, v, \underline{w})$ and $y_3 = f(\underline{u}, \underline{v}, w)$, the underlined quantities being kept constant.

Rodenacker has given several examples of this procedure, one of which concerns the development of a capillary viscometer [3.59]. Four solution variants can be derived from the well-known law of capillary action $\eta \sim \Delta p \cdot r^4/(\dot{V} \cdot l)$. They are shown schematically in Figure 3.13.

1. A solution in which the differential pressure Δp serves as a measure of the viscosity: $\Delta p \sim \eta$ ($\dot{V}$, r and l = constant).
2. A solution based on changes in radius of the capillary tube: $\Delta r \sim \eta$ ($\dot{V}$, Δp and l = constant).

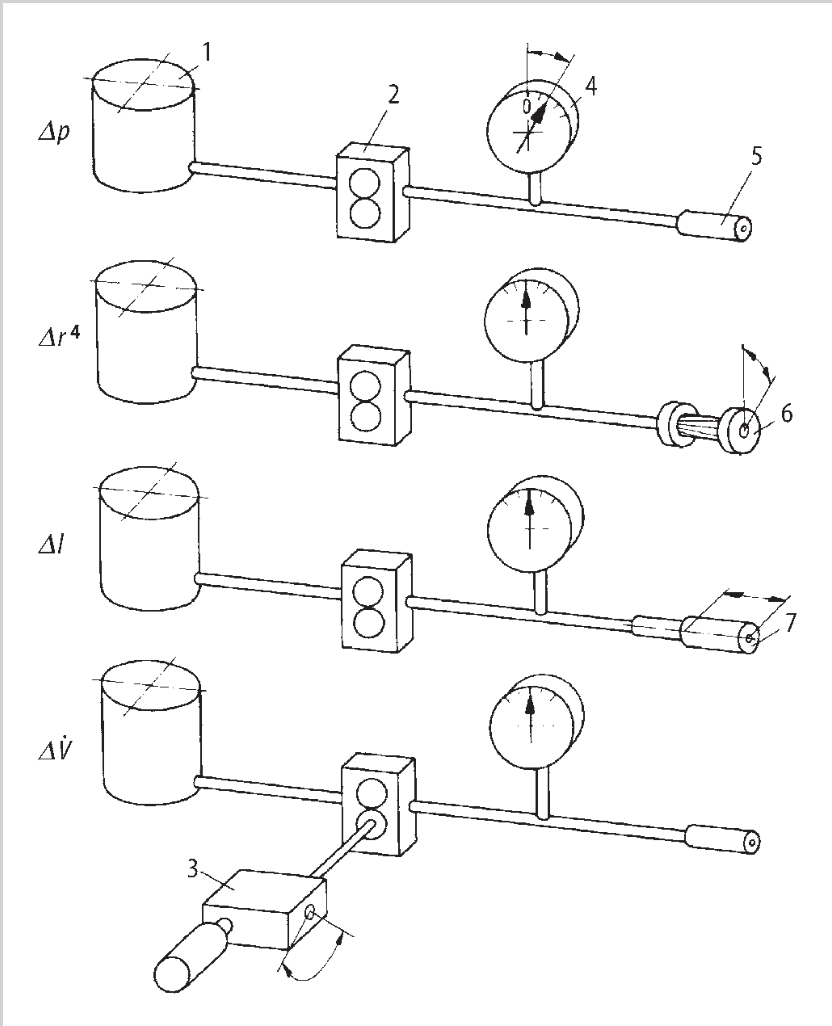


Figure 3.13. Schematic representation of four viscometers, after [3.59]. *1*, container; *2*, gear pump; *3*, variable drive; *4*, pressure gauge; *5*, fixed capillary tube; *6*, capillary tube with variable diameter; *7*, capillary tube with variable length

3. A solution based on changes in the length of the capillary tube: $\Delta l \sim \eta$ (Δp, $\dot{V}$ and r = constant).
4. A solution based on changes in the volume flow rate: $\Delta\dot{V} \sim \eta$ (Δp, r and l = constant).

Another way of obtaining new or improved solutions by the analysis of physical equations is the resolution of known physical effects into their individual components. Rodenacker [3.59], in particular, has used this approach in the design of novel devices and the development of new applications for existing ones.

By way of example, let us look at the development of a frictional thread locking device, based on the analysis of the equation governing the torque needed to release a threaded fastener:

$$T = P[(d/2)\tan(\phi_v - \beta) + (D/2)\mu_f] \tag{3.1}$$

The torque given by Equation (3.1) is made up of the following components:

Frictional torque in the thread:

$$T_t \sim P(d/2)\tan\phi_v = P(d/2)\mu_v \tag{3.2}$$

where

$$\tan\phi_v = \mu_t/\cos(\alpha/2) = \mu_v$$

Frictional torque on the bolt head or nut face:

$$T_f = P(D/2)\tan\phi_f = P(D/2)\mu_f \tag{3.3}$$

Release torque of the thread due to pre-load and thread pitch:

$$T_r \sim P(d/2)\tan(-\beta) = -P\cdot\frac{P}{2\pi} \tag{3.4}$$

(where p = thread pitch, β = helix angle, d = mean thread (t) diameter, P = pre-load, D = mean face (f) diameter, μ_v = virtual (v) coefficient of friction in the thread, μ_t = actual coefficient of friction in the thread, μ_f = coefficient of friction on the head or nut face, α = flank angle, ϕ = angle of friction).

To discover solution principles for the improvement of the locking properties of a threaded fastener, we must analyse the physical relationships further so as to identify the physical effects involved. The individual effects involved in Equations (3.2) and (3.3) are:

- the friction effect (Coulomb friction)

$$F_t = \mu_v P \quad\text{and}\quad F_f = \mu_f P$$

- the lever effect

$$T_t = F_t\, d/2 \quad\text{and}\quad T_f = F_f\, D/2$$

- the wedge effect

$$\mu_v = \mu_t/\cos(\alpha/2)$$

The individual effects in Equation (3.4) are:

- the wedge effect

$$F_r \sim P\tan(-\beta)$$

- the lever effect

$$T_r = F_r\, d/2$$

An examination of the individual physical effects will yield the following solution principles for the improvement of the locking properties of the fastener:

- Use of the wedge effect to reduce the tendency to loosen by decreasing the helix angle β.

- Use of the lever effect to increase the frictional moment on the head or nut face by increasing the mean face diameter D.
- Use of the friction effect to increase the frictional force by increasing the coefficient of friction μ.
- Use of the wedge effect to increase the frictional force on the face by means of conical surfaces ($P\mu_f/\sin\gamma$ with cone angle = 2γ). This method is used with automobile wheel attachment nuts.
- Increase of the flank angle α to increase the virtual coefficient of friction in the thread.

2. Systematic Search with the Help of Classification Schemes

In Section 2.2.5 we showed that the systematic presentation of information and data is helpful in two respects. On the one hand it stimulates the search for further solutions in various directions; on the other hand it facilitates the identification and combination of essential solution characteristics. Because of these advantages, a number of classification schemes have been drawn up, all with a similar basic structure. Dreibholz [3.10] has published a comprehensive survey of the possible applications of such classification schemes.

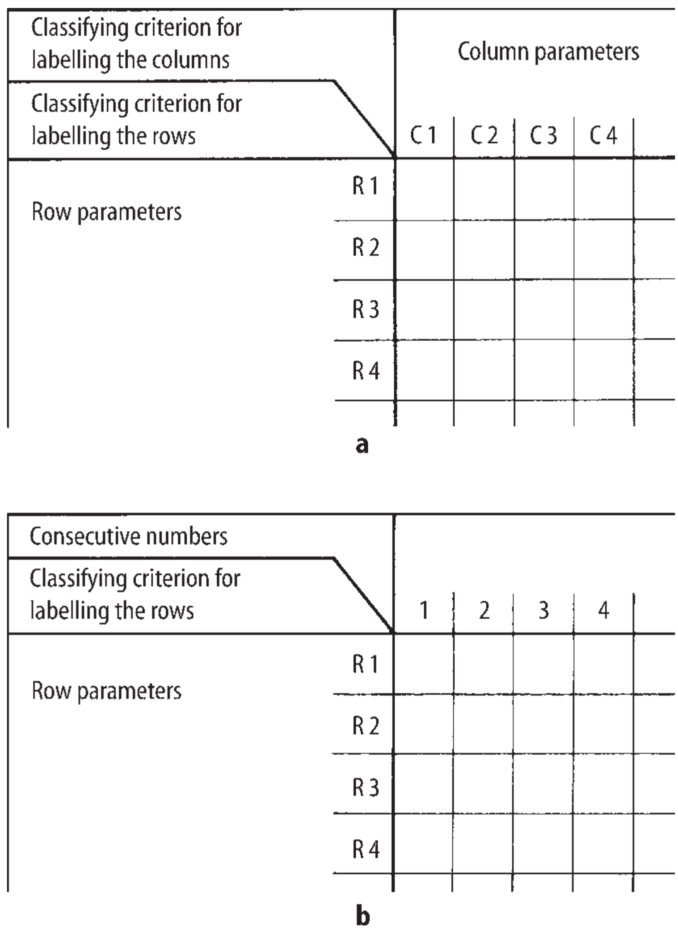


Figure 3.14. General structure of classification schemes. After [3.10]

The usual two-dimensional scheme consists of rows and columns of parameters used as classifying criteria. Figure 3.14 illustrates the general structure of classification schemes: (a) when parameters are provided for both the rows and the columns; and (b) when parameters are provided for the rows only, because the columns cannot be arranged in any apparent order. If necessary, the classifying criteria can be extended by a further breakdown of the parameters or characteristics (see Figure 3.15); this, however, often tends to confuse the general picture. By allocating the column parameters to the rows it is possible to trans-

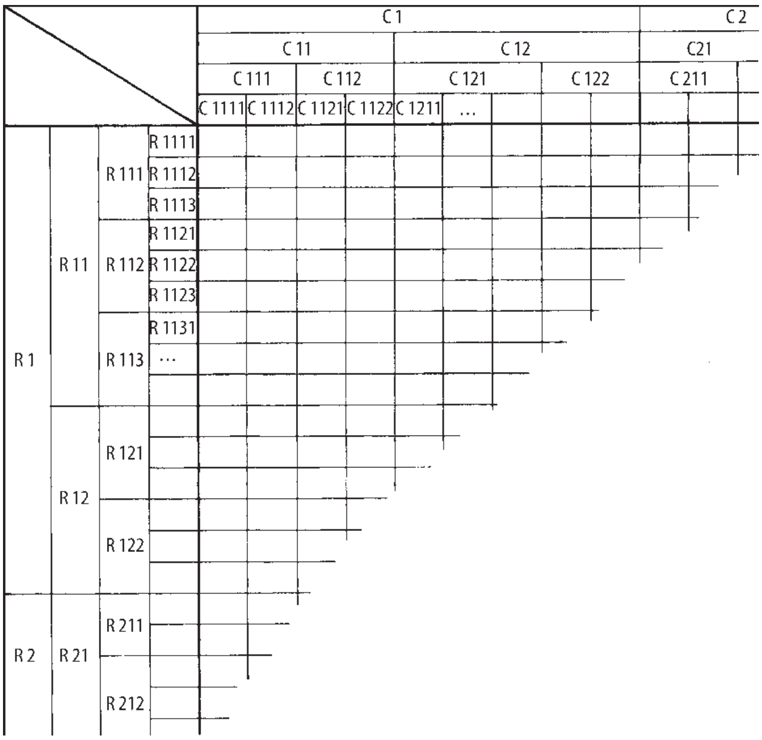


Figure 3.15. Classification scheme with further subdivision of parameters. After [3.10]

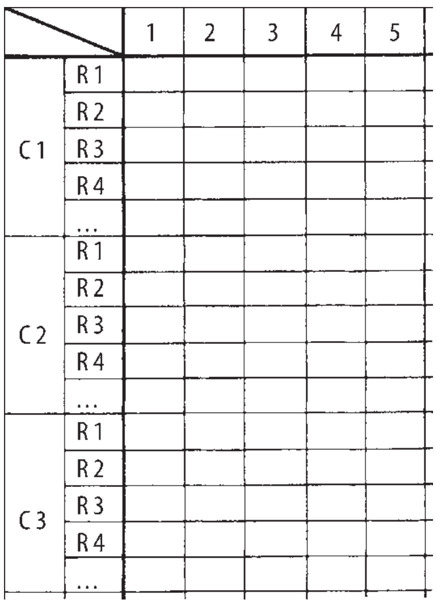


Figure 3.16. Modified classification scheme. After [3.10]

form every classification scheme based on row and column into a scheme in which only the row parameters are retained, and the columns are merely numbered (see Figure 3.16).

Such classification schemes help the design process in a great many ways. In particular, they can serve as design catalogues during all phases of the search for a solution, and they can also help in the combination of subsolutions into overall solutions (see Section 3.2.4). Zwicky [3.77] has referred to them as “morphological matrices”.

The choice of classifying criteria or their parameters is of crucial importance. In establishing a classification scheme it is best to use the following step-by-step procedure:

Step 1: Solution proposals are entered in the rows in random order.

Step 2: These proposals are analysed in the light of the main headings (characteristics), such as type of energy, working geometry, working motion, etc.

Step 3: They are classified in accordance with these headings.

The criteria and their parameters can also be obtained from an earlier use of intuitive methods to analyse known solutions or solution ideas.

This procedure not only helps with the identification of compatible combinations but, more importantly, encourages the opening up of the widest possible solution

Classifiing criteria:

Types of energy, physical effects and manifestations

Headings	*Examples*
Mechanical	Gravitation, inertia, centripetal force
Hydraulic	Hydrostatic, hydrodynamic
Pneumatic	Aerostatic, aerodynamic
Electrical	Electrostatic, electrodynamic, inductive, capacitative, piezo-electric, transformation, rectification
Magnetic	Ferromagnetic, electromagnetic
Optical	Reflection, refraction, diffraction, interference, polarisation, infra-red, visible, ultra-violet
Thermal	Expansion, bimetal effect, heat storage, heat transfer, heat conduction, heat insulation
Chemical	Combustion, oxidation, reduction, dissolution, combination, transformation, electrolysis, exothermic and endothermic reactions
Nuclear	Radiation, isotopes, source of energy
Biological	Fermentation, putrefaction, decomposition

Figure 3.17. Classifying criteria and headings (characteristics) for variation in the physical search area

fields. The classifying criteria and characteristics listed in Figures 3.17 and 3.18 can be useful when searching systematically for solutions and the variation of solution ideas for technical systems. They refer to types of energy, physical effects, manifestations, as well as the characteristics of the working geometry, working motions, and the basic material properties (see Section 2.1.4).

Figure 3.19 provides a simple example of searching for a solution to satisfy a subfunction. Here the answer was obtained by varying the type of energy against a number of working principles.

Classifying criteria

Working geometry, working motions and basic material properties

Working geometry

Headings	*Examples*
Type	Point, line, surface, body
Shape	Curve, circle, ellipse, hyperbola, parabola Triangle, square, rectangle, pentagon, hexagon, octagon Cylinder, cone, rhomb, cube, sphere Symmetrical, asymmetrical
Position	Axial, radial, tangential, vertical, horizontal Parallel, sequential
Size	Small, large, narrow, broad, tall, low
Number	Undivided, divided Simple, double, multiple

Working motions

Headings	*Examples*
Type	Stationary, translational, rotational
Nature	Uniform, non-uniform, oscillating Plane or three-dimensional
Direction	In x,y,z direction and/or about x,y,z axis
Magnitude	Velocity
Number	One, several, composite movements

Basic material properties

Headings	*Examples*
State	Solid, liquid, gaseous
Behaviour	Rigid, elastic, plastic, viscous
Form	Solid body, grains, powder, dust

Figure 3.18. Classifying criteria and headings (characteristics) for variation in the form design search area

Type of energy / Working principle	mechanical	hydraulic	electrical	thermal
1	Pot. energy, m, h	Liquid reservoir (pot. energy), h	Battery, V	Mass, m, s, $\Delta\vartheta$
2	Moving mass, v, m (transl.)	Flowing liquid	Capacitor (electr. field), C	Heated liquid
3	Flywheel, J, ω (rot.)			Superheated steam
4	Wheel on inclined plane, J, v, ω (rot. + transl. + pot.)			
5	Metal spring, d, F	Other springs (compression of fluid + gas), Δp; ΔV		
6		Hydraulic reservoir a. Bladder b. Piston c. Membrane (Pressure energy)		

Figure 3.19. Different working principles that satisfy the function "store energy" obtained by varying the type of energy

Figure 3.20 is an example of variation based on working motions.

Figure 3.21 shows the variation in the working geometry in the design of shaft–hub connections. Thanks to such arrangements, the multiplicity of solutions obtained, for instance by the method of forward steps (see Section 2.2.5 and Figure 2.21), can be put into order and completed.

To sum up, the following recommendations are given:

- Classification schemes should be built up step-by-step and as comprehensively as possible. Incompatibilities should be discarded, and only the most promising solution proposals pursued. In so doing, designers should try to determine which classifying criteria contribute to the discovery of a solution, and to examine further variations by modifying the parameters.
- The most promising solutions should be chosen and labelled using a special selection procedure (see Section 3.3.1).
- If possible, the most comprehensive classification schemes should be drawn up (those schemes intended for repeated use), but systems should never be built for the sake of systematics alone.

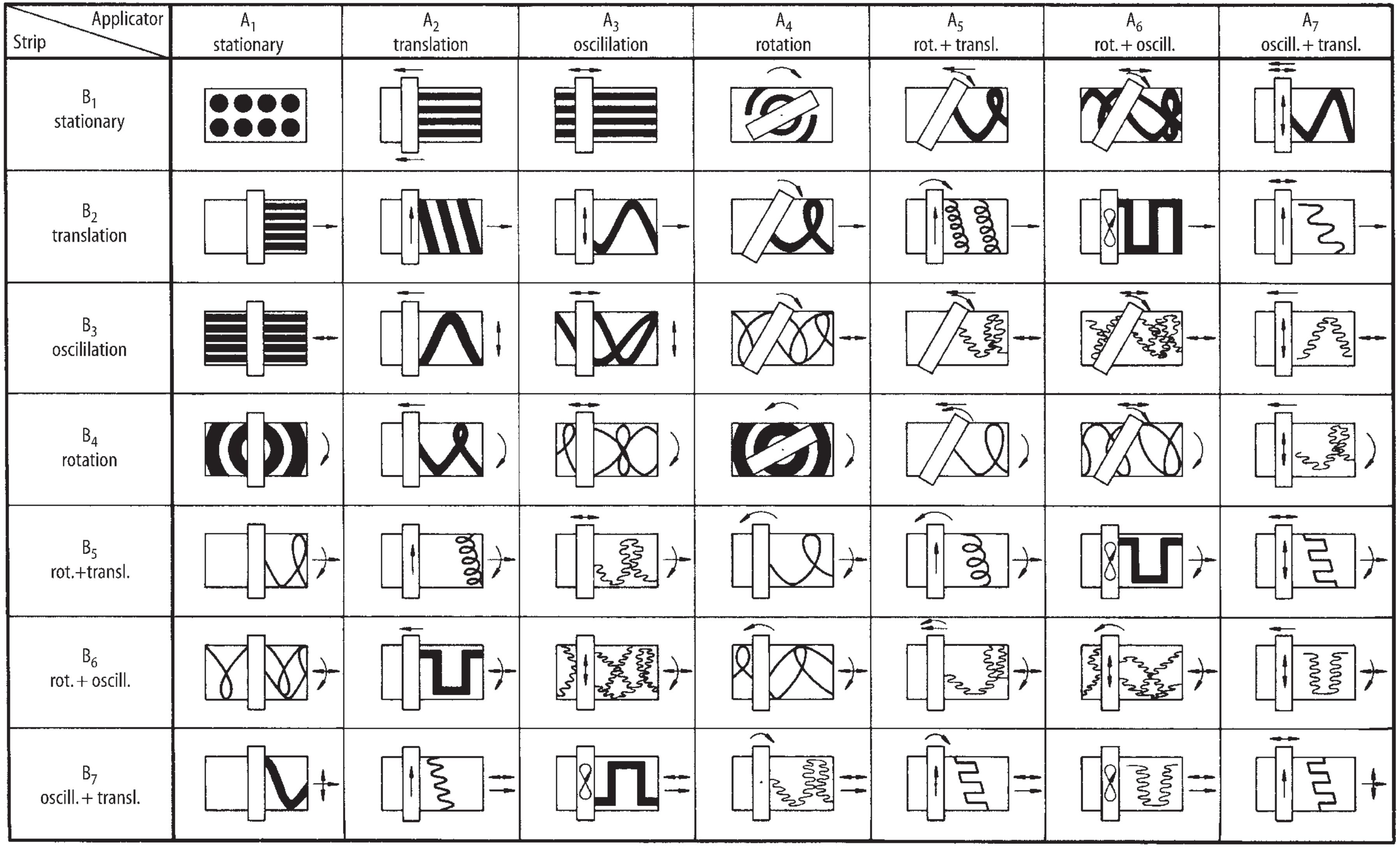


Figure 3.20. Means of coating the backs of carpets by combining the motions of the carpet (strip) and those of the applicator

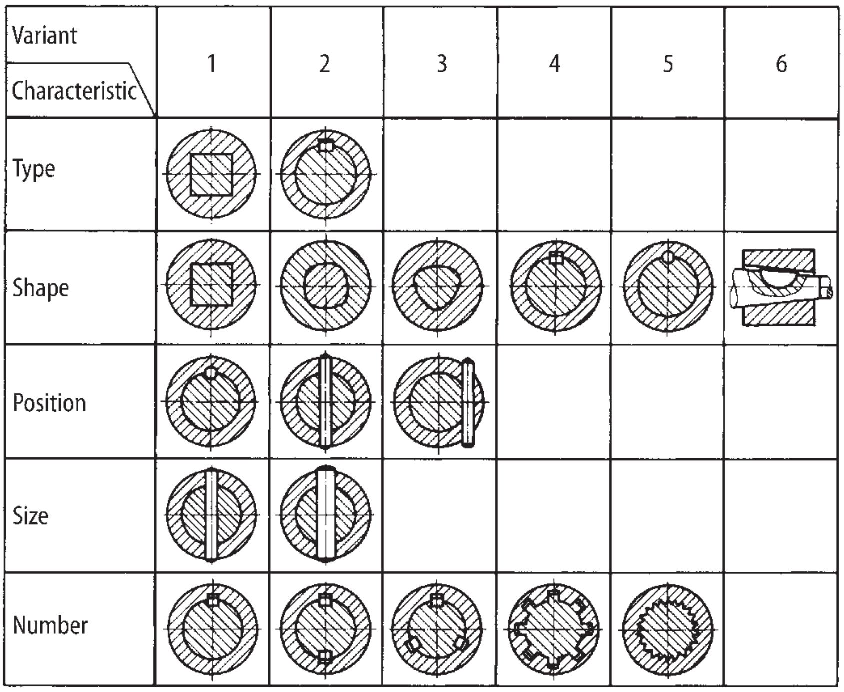


Figure 3.21. Variations in the working geometry for shaft–hub connections

3. Use of Design Catalogues

Design catalogues are collections of known and proven solutions to design problems. They contain data of various types and solutions at distinct levels of embodiment. Thus they may cover physical effects, working principles, principle solutions, machine elements, standard parts, materials, bought-out components, etc. In the past, such data were usually found in textbooks and handbooks, company catalogues, brochures and standards. Some of these contained, apart from purely objective data and suggested solutions, examples of calculations, solution methods and other design procedures. It is also possible to imagine catalogue-like collections for methods and procedures.

Design catalogues should provide:

- quicker, more problem-oriented access to the accumulated solutions or data
- the most comprehensive range of solutions possible, or, at the very least, the most essential ones, which can be extended later
- the greatest possible range of interdisciplinary applications
- data for conventional design procedures as well as for computer-aided methods.

The construction of design catalogues has been studied, above all, by Roth and collaborators [3.62]. Roth suggests that a design catalogue of the type shown in Figure 3.22 is most likely to satisfy all of the demands listed above.

The classifying criteria determine the structure of the catalogue. They influence the ease with which catalogues can be handled and reflect the level of complexity of particular solutions, as well as their degree of embodiment. In the conceptual design phase, for instance, it is advisable to select as classifying criteria the

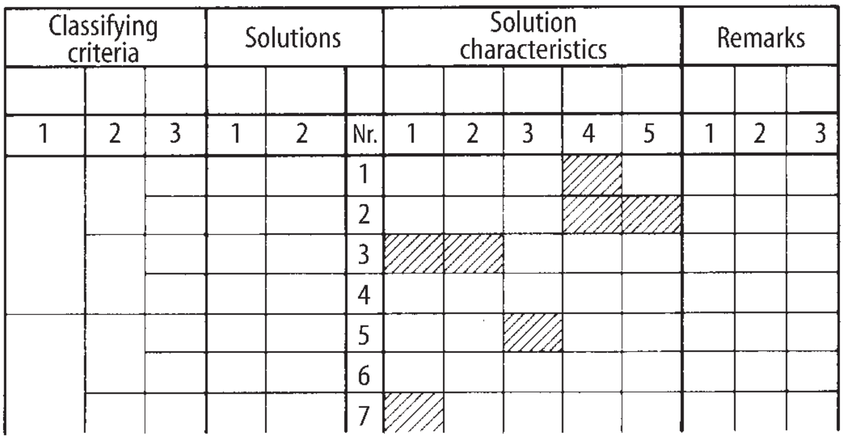


Figure 3.22. Basic structure of a design catalogue. After [3.62]

functions to be fulfilled by the solutions. This is because the conceptual design is based on the underlying subfunctions. When classifying characteristics it is best to choose generally valid functions (see Section 2.1.3), which help to elicit the most product-independent solutions.

Further classifying criteria might include the types and characteristics of energy (mechanical, electrical, optical, etc.), of material or signals, of working geometries, of working motions and of basic material properties. In the case of design catalogues intended for the embodiment design phase, useful classifying criteria include the properties of materials and the characteristics of particular machine elements, such as types of coupling.

The solution column is the main part of the catalogue and contains the solutions. Depending on the level of abstraction, the solutions are represented as sketches, with or without physical equations, or as more or less complete drawings or illustrations. The type and completeness of the information given once again depends on the intended application. It is important that all data is of the same level of abstraction and omits side issues.

The column covering the solution characteristics is important for the choice of solutions.

The remarks column can be used for information about the origin of the data and for additional comments.

The characteristics used for selection may involve a great variety of properties—for instance typical dimensions, reliability, response, number of elements, etc. They help designers in the preliminary selection and evaluation of solutions and, in the case of computer-based catalogues, they can also be used in the final selection and evaluation.

Another important requirement of design catalogues is that they should have uniform and clear definitions and symbols.

The more concrete and detailed the stored information, the more direct but also the more limited the application of a catalogue. With increasing degree of embodiment, data for a given solution become more comprehensive. However, the chances of arriving at a complete solution field decreases because the number of details, for example embodiment variants, increases rapidly. Thus, it may be

Table 3.2. Available design catalogues

Application	Object	Author and reference
General	Construction of catalogues	Roth [3.62]
	List of available catalogues and solutions	Roth [3.62]
Principle solutions	Physical effects	Roth [3.62]
	Solutions to functions	Koller [3.39]
Connections	Types of connections	Roth [3.62]
	Connections	Ewald [3.14]
	Fixed connections	Roth [3.62]
	Wedded joints for steel profiles	Wölse and Kastner [3.75]
	Riveted joints	Roth [3.62], Kopowski [3.41], Grandt [3.26]
	Adhesive joints	Fuhrmann and Hinterwalder [3.18]
	Clamping elements	Ersoy [3.13]
	Principles of threaded joints	Kopowski [3.41]
	Threaded fasteners	Kopowski [3.41]
	Elimination of backlash in threaded joints	Ewald [3.14]
	Elastic joints	Gießner [3.24]
	Shaft–hub connections	Roth [3.62], Diekhöner and Lohkamp [3.9], Kollmann [3.40]
Guides and bearings	Linear guides	Roth [3.62]
	Rotational guides	Roth [3.62]
	Plain and roller bearings	Diekhöner [3.8]
	Bearings and guides	Ewald [3.14]
Power generation, power transmission	Electric motors (small)	Jung and Schneider [3.32]
	Drives (general)	Schneider [3.65]
	Power generators (mechanical)	Ewald [3.14]
	Effects to generate power	Roth [3.62]
	Single-stage power multiplication	Roth [3.62], VDI 2222 [3.70]
	Lifting mechanisms	Raab and Schneider [3.57]
	Screw drives	Kopowski [3.41]
	Friction systems	Roth [3.62]
Kinematics, mechanisms	Solving motion problems using mechanisms	VDI 2727, part 2 [3.72]
	Chain drives and mechanisms	Roth [3.62]
	4-bar mechanisms	VDI 2222, part 2 [3.70]
	Logical inverse mechanisms	Roth [3.62]
	Logical conjunctive and disjunctive mechanisms	Roth [3.62]
	Mechanical flip-flops	Roth [3.62]
	Mechanical non-return safety devices	Roth [3.62], VDI 2222, part 2 [3.70]
	Lifting mechanisms	Raab and Schneider [3.57]
	Uniform-motion transmissions	Roth [3.62]
	Handling devices	VDI 2740 [3.73]
Gearboxes	Spur gears	VDI 2222, part 2 [3.70], Ewald [3.14]
	Mechanical single-stage gearboxes with constant gear ratio	Diekhoner and Lohkamp [3.9]
	Elimination of backlash in spur gears	Ewald [3.14]
Safety technology	Danger situations	Neudorfer [3.52]
	Protective barriers	Neudorfer [3.53]
Ergonomics	Indicators, controls	Neudorfer [3.51]
Production processes	Casting	Ersoy [3.13]
	Drop forging	Roth [3.62]
	Press forging	Roth [3.62]

Function	Input	Output	Physical effects						
E_{mech} → E_{mech}	Force, pressure, torque	Length, angle	Hooke (Tension/ compression/ bending)	Shear, torsion	Upthrust Poisson´s effect	Boyle-Mariotte	Coulomb I and II	...	...
		Speed	Energy Law	Conservation of momentum	Conservation of angular momentum	...	...	...	...
		Acceleration	Newton´s Law	...	...	...	...	...	...
	Length, angle	Force, pressure, torque	Hooke	Shear, torsion	Gravity	Upthrust	Boyle-Mariotte	Capillary	
			Coulomb I and II	...	...	...	...	...	...
	Speed	– – – –	Coriolis force	Conservation of momentum	Magnus-effect	Energy law	Centrifugal force	Eddy current	
	Acceleration	– – – –	Newton´s Law	...	...	...	...	...	...
E_{mech} → E_{hyd}	Force, length, speed, pressure	Speed, pressure	Bernoulli	Viscosity (Newton)	Torricelli	Gravitational pressure	Boyle-Mariotte	Conservation of momentum	...
E_{hyd} → E_{mech}	Speed	Force, length	Profile lift	Turbulence	Magnus-effect	Flow resistance	Back pressure	Reaction principle	...
E_{mech} → E_{therm}	Force, speed	Temperature, quantity of heat	Friction (Coulomb)	1st law	Thomson-Joule	Hysteresis (damping)	Plastic deformation	...	...
E_{therm} → E_{mech}	Temperature, heat	Force, pressure, lenght	Thermal expansion	Steam pressure	Gas Law	Osmotic pressure	...	...	...
E_{elektr} → E_{mech}	Voltage, current, magn. field	Force, speed, pressure	Biot-Savart-effect	Electro-kinetic effect	Coulomb I	Capacitance effect	Johnsen-Rhabeck-effect	Piezoeffect	...
E_{mech} → E_{elektr}	Force, length, speed, pressure	Voltage, current	Induction	Electro-kinetics	Electro-dynamic effect	Piezoeffect	Frictional electricity	Capacitance effect	...
E_{elektr} → E_{therm}	Voltage, current	Temperature, heat	Joule heating	Peltier-effect	Electric arc	Eddy current	...	...	...
E_{therm} → E_{elektr}	Temperature, heat	Voltage, current	Electr. conduction	Thermo-effect	Thermionic emission	Pyroelectricity	Noise-effect	Semiconductor, Superconductor	...
E_{mech} ↔ E_{mech}	Force, length, pressure, speed	Force, length, pressure, speed	Lever	Wedge	Poisson´s effect	Friction	Crank	Hydraulic effect	...
E_{hyd} ↔ E_{hyd}	Pressure, speed	Pressure, speed	Continuity	Bernoulli	...	...	...	...	...
E_{therm} ↔ E_{therm}	Temperature, heat	Temperature, heat	Heat conduction	Convection	Radiation	Condensation	Evaporation	Freezing	...
E_{elektr} ↔ E_{elektr}	Voltage, current	Voltage, current	Transformer	Valve	Transistor	Transducer	Thermogalvanometer	Ohm´s law	...
...	...	...	...	...	...	...	...	...	...

Figure 3.23. Design catalogue of physical effects based on [3.39, 3.48] for the generally applicable functions "change energy" and "vary energy component". Also applicable to flow of signals

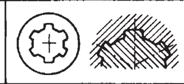
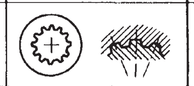
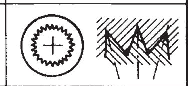
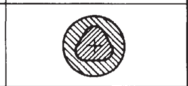
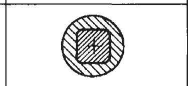
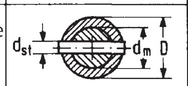
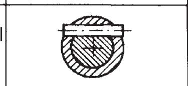
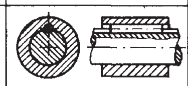
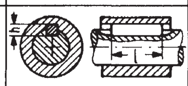
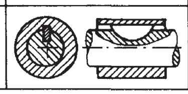
Classifying criteria		Solutions				Solution characteristics															Remarks		
Type of interface	Type of force transmission	Equation	Name	Configuration		Transferable torque	Torque transmission depending on	Axial forces	Stress concentration	Applicable for	Behaviour at overload	Centering possible	Unbalanced force	Axial displacement of hub	Hub moveable	Joint adjustable	Shaft diameter (mm)	Material	Manufacturing effort	Assembly effort	Standard (DIN)	Application examples	Remarks
1	2	1	2	3	Nr.	1	2	3		4	5	6	7	8	9	10	11	12	13	14	15	16	17
			spline shaft		1		shaft diameter,(form) factor, centering factor, material	–	large								10 – 150				5461/63 5471/72	toothed wheels	exterior, flank, interior centering possible
			involute spline shaft		2		h i l	–		pulsating or alternating		yes	no	clearance fit		no			high		5480, 5482		
	Direct		serrated shaft		3	large		–									150 – 500	shaft: 37 Cr 4 41 Cr 4 42 CrMo		small	5481		short hub possible
			3-sided polygon-shaft		4		e i l	–	medium	pulsating or alternating		self-centering	no	clearance fit without load	yes	yes for taper	10 – 100		small, special machinery necessary		—		used for short and thin hubs, conical shaft end possible, broaching or grinding necessary
			4-sided polygon-shaft		5			–									10 – 100				—		
Normal (form fit)		$T \le K\cdot\frac{d}{2}\cdot A_s\cdot\tau_{max}$ or $T \le K\cdot\frac{d}{2}\cdot A_p\sigma_{max}$	transverse pin		6		d_p D	are supported	large	—	fracture			—			0.5 – 50	pin: 4D, 5S 6S, 8G 9S, 20K St 50 K St 70 St 60		medium	1,7 1470 – 77 1481, 6324, 7346		taper and grooved pin possible
			tangential pin		7					—				—		yes for taper pin						power stretcher, machine tools, vehicles	
			in line pin		8	small	d_p l	–	medium	—		yes							medium				
	Indirect		key joint		9		h l b	–	large	—			yes	clearance fit	no	no	5 – 500	spring St 60		small	6885		
			Woodruff key		10			–		—								shaft, hub: GG,GS ST			6888		

Figure 3.24. Extract of a catalogue for shaft–hub connections. After [3.62]

possible to provide a full list of physical effects fulfilling the function "channel", but it would hardly be possible to list all of the potential embodiments of bearings (channelling a force from a rotating to a stationary system).

Table 3.2 lists the currently available design catalogues that satisfy the requirements and structure described above. Therefore, in what follows we include just a few examples of, or extracts from, available design catalogues.

Figure 3.23 shows a catalogue of physical effects associated with the functions "change energy" and "vary energy component". It is based on Koller [3.39] and Krumhauer [3.48]. The catalogue makes it possible to derive these effects from the classifying criteria, that is, the "inputs and outputs" of the functions. The characteristics on which the selection is based must be derived from the technical literature.

Figure 3.24 shows an extract of a catalogue for shaft–hub connections based on [3.62]. In this, unlike the previous catalogue, the solutions are concrete enough, thanks to the specification of the form design features, for the embodiment design phase to start with a scale layout drawing.

Computer-based systems are used to facilitate searching through catalogues, company brochures, supplier information and other documents. Hypermedia software provides a way of structuring, storing and retrieving the contents of such documents. It allows the flexible manipulation of chunks of information, and the representation and linking of objects and procedures in a specific knowledge domain, using different representation principles. This is called navigating in a hypermedia system [3.58]. To use distributed sources of information, a global network is required, such as the internet (www). Using the internet, so-called "virtual markets" or "virtual supply chains" can be created with which designers can communicate from their work places [3.4].

3.2.4 Methods for Combining Solutions

As described in Sections 2.1.3 and 2.2.5, it is often useful to divide problems, tasks and functions into subproblems, subtasks and subfunctions and to solve these individually (factorisation method) (see also Section 6.3). Once the solutions for subproblems, subtasks or subfunctions are available, they have to be combined in order to arrive at an overall solution.

The methods we have been describing, particularly those with an intuitive bias, may have led to the discovery of suitable combinations. However, there are special methods for arriving at such syntheses more directly. In principle, they must permit a clear combination of solution principles with the help of the associated physical and other quantities and the appropriate geometrical and material characteristics. When analysing combinations that involve software elements, it is important to identify and use appropriate solution characteristics.

The main problem with such combinations is ensuring the physical and geometrical compatibility of the solution principles to be combined, which in turn ensures the smooth flow of energy, material and signals, and avoids geometrical interference in mechanical systems. For information systems, the main problem is the compatibility requirements of the information flow.

A further problem is the selection of technically and economically favourable combinations of principles from the large field of theoretically possible combinations. This aspect will be discussed at greater length in Section 3.3.1.

1. Systematic Combination

For the purpose of systematic combination, the classification scheme to which Zwicky [3.77] refers as the "morphological matrix" (see Figure 3.25) is particularly useful. Here, the subfunctions, usually limited to the main functions, and appropriate solutions (solution principles) are entered in the rows of the scheme.

If this scheme is to be used for the elaboration of overall solutions, then at least one solution principle must be chosen for every subfunction (that is, for every row). To provide the overall solution, these principles (subsolutions) must then be combined systematically into an overall solution. If there are m_1 solution principles for the subfunction F_1, m_2 for the subfunction F_2, and so on, then after complete combination we have $N = m_1 \cdot m_2 \cdot m_3 \cdots m_n$ theoretically possible overall solution variants.

The main problem with this method of combination is to decide which solution principles are compatible; that is, to narrow down the theoretically possible search field to the practically possible search field.

The identification of compatible subsolutions is facilitated if:

- the subfunctions are listed in the order in which they occur in the function structure, if necessary separated according to flow of energy, material and signals
- the solution principles are suitably arranged with the help of additional column parameters, for example the types of energy
- the solution principles are not merely expressed in words but also in rough sketches

Subfunctions \ Solutions		1	2	…	j	…	m
1	F_1	S_{11}	S_{12}		S_{1j}		S_{1m}
2	F_2	S_{21}	S_{22}		S_{2j}		S_{2m}
⋮		⋮	⋮		⋮		⋮
i	F_1	S_{i1}	S_{i2}		S_{ij}		S_{im}
⋮		⋮	⋮		⋮		⋮
n	F_n	S_{n1}	S_{n2}		S_{nj}		S_{nm}

② ① Combinations of principles

Figure 3.25. Combining solution principles into combinations of principles: Combination 1: $S_{11} + S_{22} + \ldots S_{n2}$; Combination 2: $S_{11} + S_{21} \ldots S_{n1}$

- the most important characteristics and properties of the solution principles are recorded as well.

The verification of compatibilities, too, is facilitated by classification schemes. If two subfunctions to be combined—for instance, "change energy" and "vary mechanical energy component"—are entered respectively in the column and row headings of a matrix with their characteristics in the appropriate cells, then the compatibility of the subsolutions can be verified more easily than it could be if such examinations were to be confined to the designer's head. Figure 3.26 illustrates this type of compatibility matrix. Further examples of this method of combination will be found in Section 6.4.2 (Figures 6.15 and 6.19).

To sum up:

- Combine only compatible subfunctions.
- Pursue only such solutions as meet the demands of the requirements list and fall within the available resources (see selection procedures in Section 3.3.1).
- Concentrate on promising combinations and establish why these should be preferred above the rest.

In conclusion, it must be emphasised that what we have been discussing is a generally valid method of combining subsolutions into overall solutions. The method can be used for the combination of working principles during the conceptual phase, and of subsolutions or even of components and assemblies during the embodiment phase. Because it is essentially a method of information processing, it

Change energy / Vary mechan. energy component		Electric motor	Oscillating solenoid	Bimetal spiral in hot water	Oscillating hydraulic piston	...
		1	2	3	4	...
Four-bar linkage	A	if A capable of rotating	slow motion	yes	additional lever linkage but only for low piston speeds	...
Chain drive Spur gear drive	B	yes	slow rotation only through additional elements (free wheeling etc.), difficult to reverse direction	gear segments suffice, depending on angle of rotation	with a rack and swivel, but only for low piston speeds	...
Maltese drive	C	yes look out for shock loads	see B2	yes (when angle of rotation is small lever with sliding block)	lever with sliding block, but only for low piston speeds	...
Friction wheel drive	D	yes	see B2	large forces because of torque during slow movement, imprecise positioning	see D3	...
...	...	...	...	...	...	...

☒ very difficult to apply (do not pursue further)

☒ can only be applied under certain circumstances (defer)

Figure 3.26. Compatibility matrix for combination possibilities of the subfunctions "change energy" and "vary mechanical energy component". After [3.10]

is not confined to technical problems but can also be used in the development of management systems and in other fields.

2. Combining With the Help of Mathematical Methods

Mathematical methods and computers should only be used for the combination of solution principles if real advantages can be expected from them. Thus, at the relatively abstract conceptual phase, when the nature of the solution is not yet fully understood, a quantitative elaboration—that is, a mathematical combination along with an optimisation—is quite out of place and can be misleading. The exceptions are combinations of known elements and assemblies, for instance in variant or circuit design. In the case of purely logical functions, combinations can be performed with the help of Boolean algebra [3.17, 3.59] in, say, the layout of safety systems or the optimisation of electronic or hydraulic circuits.

In principle, the combination of subsolutions into overall solutions with the help of mathematical methods calls for knowledge of the characteristics or properties of the subsolutions that are expected to correspond with the relevant properties of the neighbouring subsolutions. These properties must be unambiguous and quantifiable. In the formation of principle solutions (for example working structures), data about the physical relationships may be insufficient, since the geometrical relationships may have a limiting effect and hence may, in certain circumstances, lead to incompatibilities. In that case, physical equations and geometrical structure must first be matched mathematically, and this is not generally possible except for systems of low complexity. For systems of higher complexity, in contrast, such correlations often become ambiguous, so that designers must once again choose between variants. We may, accordingly, speak of dialogue systems in which the process of combination consists of mathematical and creative steps.

This makes it clear that, with increasing physical realisation or embodiment of a solution, it becomes simpler to establish quantitative combination rules. However, the number of properties increases and with them the number of constraints and optimisation criteria, so that the mathematical effort becomes very great and requires computer support.

3.3 Selection and Evaluation Methods

3.3.1 Selecting Solution Variants

For the systematic approach, the solution field should be as wide as possible. By considering all possible classifying criteria and characteristics, designers are often led to a larger number of possible solutions. This profusion constitutes the strength and also the weakness of the systematic approach. The very great theoretically admissible, but practically unattainable, number of solutions must be reduced at the earliest possible moment. On the other hand, care must be taken not to eliminate valuable working principles, because it is often only through their combination with others that an advantageous working structure will emerge. While there is no absolutely safe procedure, the use of a systematic and verifiable selection procedure greatly facilitates the choice of promising solutions from a wealth of proposals [3.55].

This selection procedure involves two steps, namely *elimination* and *preference*.

First, all totally unsuitable proposals are eliminated. If too many possible solutions still remain, those that are patently better than the rest must be given preference. Only these solutions are evaluated at the end of the conceptual design phase.

If faced with a large number of solution proposals, the designer should compile a selection chart (see Figure 3.27). In principle, after every step—that is, even after establishing function structures—the only solution proposals pursued should:

- be compatible with the overall task and with one another (Criterion A)
- fulfil the demands of the requirements list (Criterion B)
- be realisable in respect of performance, layout, etc. (Criterion C)
- be expected to be within permissible costs (Criterion D).

Unsuitable solutions are eliminated in accordance with these four criteria applied in the above sequence. Criteria A and B are suitable for yes/no decisions and their application poses relatively few problems. Criteria C and D often need a more quantitative approach, which should only be used once Criteria A and B have been satisfied.

Since Criteria C and D involve quantitative considerations, they may lead not only to the elimination of proposed solutions that fail to meet the requirements, but also of those that exceed the requirements by an unnecessary margin.

A preference is justified if, among the very large number of possible solutions, there are some that:

- incorporate direct safety measures or introduce favourable ergonomic conditions (Criterion E)
- are preferred by the designer's company; that is, can be readily developed with the usual know-how, materials, procedures and under favourable patent conditions (Criterion F).

Additional selection criteria can be used if they help decisions to be made.

It must be stressed that selection based on preferential criteria is only advisable when there are so many variants that a full evaluation would involve too much time and effort.

If, in the suggested sequence, one criterion leads to the elimination of a proposal, then the other criteria need not be applied to it there and then. At first, only the solution variants that satisfy all of the criteria should be pursued. Sometimes, however, it is impossible to settle the issue because of lack of information. In the case of promising variants that satisfy Criteria A and B, the gap will have to be filled by a reevaluation of the proposal, which will ensure that no good solutions are passed over.

TH Darmstadt	SELECTION CHART for *Fuel gauge*	Page: *1*

Solution variants (Sv) evaluated by

SELECTION CRITERIA

(+) Yes
(–) No
(?) Lack of information
(!) Check requirements list

DECISION

Mark solution variants (Sv)

(+) Pursue solution
(–) Eliminate solution
(?) Collect information (re-evaluate solution)
(!) Check requirements list for changes

Enter solution variant (Sv):

A Compatibility assured
B Fulfils demands of requirements list
C Realisable in principle
D Within permissible costs
E Incorporates direct safety measures
F Preferred by designer's company
G Adequate information

Sv		A	B	C	D	E	F	G	Remarks (Indications, Reasons)	DECISION
1	1	+	+	+	?				*Number of measuring positions*	?
2	2	+	–						*Storing the mass*	–
3	3	–							*Radioactivity*	–
4	4	+	+	+	+		(+)		*(Further development of existing solutions)*	+
5	5	+	+	+	+					+
6	6	–							*Fluid not conducting*	–
7	7	+	+	+	+					+
8	8	+	+	+	+				*see Sv 7*	+
	9									
	10									
	11									
	12									
	13									
	14									
	15									
	16									
	17									
	18									
	19									
	20									

Date: *9.85*	Initials: *La*	

Figure 3.27. Systematic selection chart: 1, 2, 3, etc. are solution variants of the proposals made in Table 3.3. The column reserved for remarks lists reasons for lack of information or elimination

Table 3.3. Extract from a list of solutions for a fuel gauge

No.	Solution principle	Signal
	1. Measuring the quantity of fluid	
	1.1. Mechanical, static	
1.	Fix container at three points. Measure vertical forces (weight). (Measuring at one support may be sufficient)	Force
2.	Mutual attraction. The force is proportional to the masses and therefore to the fluid mass	Force
	1.2. Atomic	
3.	Distribution of radioactive material in the fluid	Concentration of radiation intensity
	2. Measuring the fluid level	
	2.1. Mechanical, static	
4.	Float with or without lever effect. Lever output: linear or angular displacement Potentiometer resistance represents fluid level within the container	Displacement
	2.2. Electrical	
5.	Resistance wire: hot in air, cold in fluid. Level of fluid determines: overall resistance, volume (dependent on temperature and length of wire)	Ohmic resistance
6.	Fluid as ohmic resistance (level-dependent). Changing the level of the (conducting) fluid changes the resistance	Ohmic resistance
	2.3. Optical	
7.	Photocells in the container. Fluid covers a certain number of photocells. The number of light signals is a measure of the fluid level	Light signal (discrete)
8.	Light transmission or light reflection. Transmission in the presence of fluid. Total reflection in presence of air	Light signal (discrete)

The criteria are listed in the order shown above as a labour-saving device, and not in order of importance.

The selection procedure has been systematised for easier implementation and verification (see Figure 3.27). Here, the criteria are applied in sequence and the reasons for eliminating any solution proposal is recorded. Experience has shown that the selection procedure we have described can be applied very quickly, that it gives a good picture of the reasons for selection, and that it provides suitable documentation in the form of a selection chart.

If the number of solution proposals is small, elimination may be based on the same criteria, but less formally recorded.

The example we have chosen concerns solution proposals for a fuel gauge in accordance with the requirements in Figure 6.4. An extract from the list of proposals is given in Table 3.3.

Further examples of selection charts can be found in Section 6.4.3 (see Figure 6.17) and Section 6.6.2 (see Figure 6.48).

3.3.2 Evaluating Solution Variants

The promising solutions that result from the selection procedure usually have to be firmed up before a final evaluation is made using criteria that are more detailed and possibly quantified. This evaluation involves an assessment of technical, safety, environmental and economic values. For this purpose, evaluation procedures have been developed that can be used to evaluate technical and nontechnical systems, and that can be applied in all phases of product development. Evaluation procedures are by their very nature more elaborate than selection procedures (see Section 3.3.1) and are therefore only applied at the end of the main working steps to determine the current value of a solution. This occurs, in general, when preparing for a fundamental decision concerning the direction of a solution path, or at the end of the conceptual and embodiment phases [3.61].

1. Basic Principles

An evaluation is meant to determine the "value", "usefulness" or "strength" of a solution with respect to a given objective. An objective is indispensable since the value of a solution is not absolute, but must be gauged in terms of certain requirements. An evaluation involves a comparison of concept variants or, in the case of a comparison with an imaginary ideal solution, a "rating" or degree of approximation to that ideal.

The evaluation should not be based on individual aspects such as production cost, safety, ergonomics or environment, but should, in accordance with the overall aim (see Section 2.1.7), consider all aspects in an appropriate balance.

Hence there is a need for methods that allow a more comprehensive evaluation, or in other words cover a broad spectrum of objectives (task-specific requirements and general constraints). These methods are intended to elaborate not only the quantitative but also the qualitative properties of the variants, thus making it possible to apply them during the conceptual phase, with its low level of embodiment and correspondingly low state of information. The results must be reliable, cost-effective, easily understood and reproducible. The most important methods to date are Cost–Benefit Analysis based on the systems approach [3.76], and the combined technical and economic evaluation technique specified in Guideline VDI 2225 [3.71], which essentially originates from Kesselring [3.36].

In what follows, we shall outline a basic evaluation procedure incorporating the concepts of Cost–Benefit Analysis and of Guideline VDI 2225. At the end the similarities and differences between both methods will be discussed.

Identifying Evaluation Criteria

The first step in any evaluation is the drawing up of a set of objectives from which evaluation criteria can be derived. In the technical field, such objectives are mainly derived from the requirements list and from general constraints (see guidelines in Section 2.1.7), which are identified while working on a particular solution.

A set of objectives usually comprises several elements that not only introduce a variety of technical, economic and safety factors, but that also differ greatly in importance.

A range of objectives should satisfy as far as possible the following conditions:

- The objectives must cover the decision-relevant requirements and general constraints as completely as possible, so that no essential criteria are ignored.

- The individual objectives on which the evaluation must be based should be as independent of one another as possible; that is, provisions to increase the value of one variant with respect to one objective must not influence its values with respect to the other objectives.
- The properties of the system to be evaluated must, if possible, be expressed in concrete quantitative or at least qualitative (verbal) terms.

The tabulation of such objectives depends very much on the purpose of the particular evaluation, that is, on the design phase and the relative novelty of the product.

Evaluation criteria can be derived directly from the objectives. Because of the subsequent assignment of values, all criteria must be given a positive formulation, i.e. such that a higher value indicates better, for example:

- "low noise" not "loudness level"
- "high efficiency" not "magnitude of losses"
- "low maintenance" not "maintenance requirements".

Cost–Benefit Analysis systematises this step by means of an objectives tree, in which the individual objectives are arranged in hierarchical order. The subobjectives are arranged vertically into levels of decreasing complexity, and horizontally into objective areas—for instance, technical, economic—or even into major and minor objectives (see Figure 3.28). Because of their required independence, subobjectives of a higher level may only be connected with an objective of the next lowest level. This hierarchical order helps the designer to determine whether or not all decision-relevant subobjectives have been covered. Moreover, it simplifies the assessment of the relative importance of the subobjectives. The evaluation criteria (called objective criteria in Cost–Benefit Analysis) can then be derived from the subobjectives of the stage with the lowest complexity.

Guideline VDI 2225, on the contrary, introduces no hierarchical order for the evaluation criteria, but derives a list of them from minimum demands and wishes and also from general technical properties.

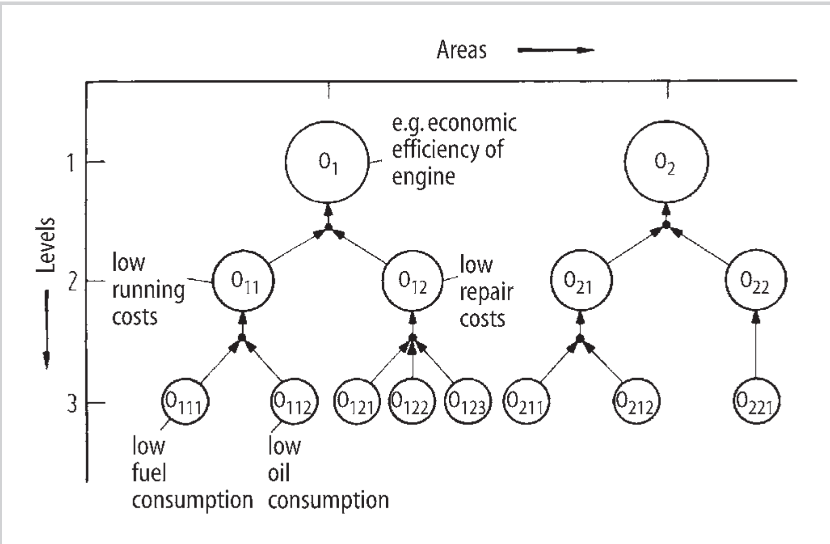


Figure 3.28. Structure of an objectives tree

Weighting Evaluation Criteria

To establish evaluation criteria, we must first assess their relative contribution (weighting) to the overall value of the solution, so that relatively unimportant criteria can be eliminated before the evaluation proper begins. The evaluation criteria retained are given "weighting factors" which must be taken into consideration during the subsequent evaluation step. A weighting factor is a real, positive number. It indicates the relative importance of a particular evaluation criterion (objective).

It has been suggested that such weightings should be assigned to the wishes when they are recorded in the requirements list [3.62, 3.63], but that is only possible if such wishes can be ranked in order of importance when the requirements list is first drawn up. That, however, rarely happens at this early stage—experience has shown that many evaluation criteria emerge during the development of the solution, and that their relative importance changes. It is nevertheless most helpful to include rough estimates of the importance of wishes when drawing up the requirements list, because, as a rule, all the persons concerned are available at that time (see Section 5.2.2).

In Cost–Benefit Analysis, weightings are based on factors ranging from 0 to 1 (or from 0 to 100). The sum of the factors of all evaluation criteria (subobjectives at the lowest level) must be equal to 1 (or 100) so that a percentage weighting can be attached to all of the subobjectives. The drawing up of an objectives tree greatly facilitates this process.

Figure 3.29 illustrates the procedure. Here the objectives have been set out on four levels of decreasing complexity and provided with weighting factors. The evaluation proceeds step-by-step from a level of higher complexity to the next lowest level. Thus the three subobjectives O_{11}, O_{12} and O_{13} of the second level are first weighted with respect to the objective O_1. In this particular case the

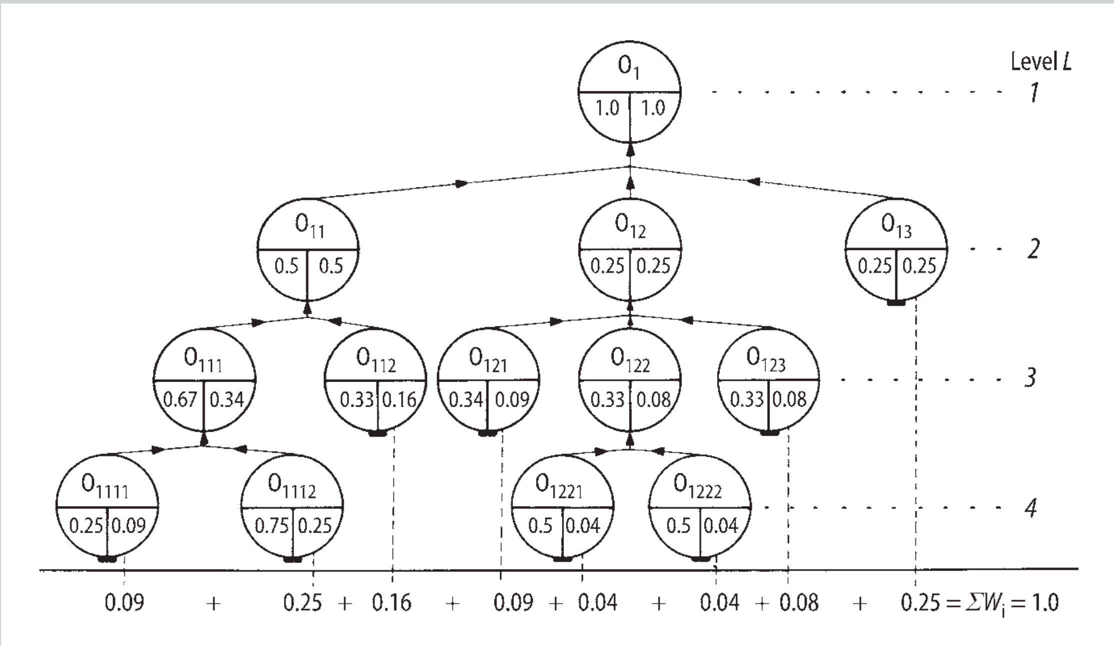


Figure 3.29. Objectives tree with weighting factors. After [3.76]

weightings are 0.5, 0.25 and 0.25. The sum of the weighting factors for any one level must always be equal to $\sum w_i = 1.0$. Next comes the weighting of the objectives of the third level with respect to the subobjectives of the second level. Thus the relative weights of O_{111} and O_{112} with respect to the higher objectives O_{11} were fixed at 0.67 and 0.33. The remaining objectives are treated in similar fashion. The relative weighting of an objective at a particular level with respect to the objective O_1 is found by multiplication of the weighting factor of the given objective level by the weighting factors of the higher objective levels. Thus the subobjective O_{1111}, which has a weighting of 0.25 with respect to the subobjective O_{111} of the next higher level, has a weighting of $0.25 \times 0.67 \times 0.5 \times 1 = 0.09$ with respect to O_1.

Such step-by-step weighting generally produces a realistic ranking because it is much easier to weight two or three subobjectives with respect to an objective on a higher level than to confine the weighting to one particular level only, especially the lowest. Figure 6.33 gives a concrete example of the recommended procedure.

Guideline VDI 2225 tries to dispense with weightings and relies instead on evaluation criteria of approximately equal importance. Weighting factors ($2\times, 3\times$) are, however, used for pronounced differences. Kesselring [3.36], Lowka [3.50] and Stahl [3.68] have examined the influences of such weighting factors on the overall value of the solution. Their conclusion was that they exert a significant influence whenever the variants to be evaluated have very distinct properties, and whenever the corresponding evaluation criteria have great importance.

Compiling Parameters

The setting up of evaluation criteria and the determination of their importance is followed, in the next step, by the assignment to them of known (or analytically determined) parameters. These parameters should either be quantifiable or, if that is impossible, be expressed by statements framed as concretely as possible. It has proved very useful to assign such parameters to the evaluation criteria in an evaluation chart before proceeding to the actual evaluation. Figure 3.30 shows an example of such a chart for an internal combustion engine, with appropriate magnitudes entered in the relevant variant columns. The reader will see that the verbal formulation of the evaluation criteria strongly resembles that of the parameters.

In Cost–Benefit Analysis these parameters are referred to as objective parameters (objective criteria) that are compiled with evaluation criteria in a chart. A concrete example is given in Figure 6.55.

In Guideline VDI 2225, in contrast, evaluation follows immediately upon the setting up of evaluation criteria (see Figure 6.41).

Assessing Values

The next step is the assessment of values and hence the actual evaluation. These "values" derive from a consideration of the relative scale of the previously determined parameters, and are thus more or less subjective in character.

The values are expressed by points. Cost–Benefit Analysis employs a range from 0 to 10; Guideline VDI 2225 a range from 0 to 4 (see Figure 3.31). The advantage of the wider range is that, as experience has shown, classification and evaluation

No.	Evaluation criteria	Wt.	Objective Parameters	Unit	Variant V_1 (e.g. Eng.$_1$) Magn. m_{i1}	Value v_{i1}	Weighted value WV_{i1}	Variant V_2 (e.g. Eng.$_2$) Magn. m_{i2}	Value v_{i2}	Weighted value WV_{i2}	…	Variant V_j Magn. m_{ij}	Value v_{ij}	Weighted value WV_{ij}	…	Variant V_m Magn. m_{im}	Value v_{im}	Weighted value WV_{im}
1	Low fuel consumption	0.3	Fuel consumption	$\frac{g}{kWh}$	240			300			…	m_{1j}			…	m_{1m}		
2	Light weight construction	0.15	Mass per unit power	$\frac{kg}{kW}$	1.7			2.7			…	m_{2j}			…	m_{2m}		
3	Simple production	0.1	Simplicity of components	–	low			average			…	m_{3j}			…	m_{3m}		
4	Long service life	0.2	Service life	km	80 000			150 000			…	m_{4j}			…	$m4_{m}$		
⋮	⋮	⋮	⋮	⋮	⋮			⋮				⋮				⋮		
i		W_i			m_{i1}			m_{i2}			…	m_{ij}			…	m_{im}		
⋮	⋮	⋮	⋮	⋮	⋮			⋮				⋮				⋮		
n		W_n			m_{n1}			m_{n2}			…	m_{nj}			…	m_{nm}		
		$\sum_{i=1}^{n} W_i = 1$																

Figure 3.30. Correlation of evaluation criteria and parameters in an evaluation chart

are greatly facilitated by the use of a decimal system that reflects percentages. The advantage of the smaller range is that, in dealing with what are so often no more than inadequately known characteristics of the variants, rough evaluations are sufficient and, indeed, may be the only meaningful approach. They involve the following assessments:

- far below average
- below average
- average
- above average
- far above average.

It is useful to begin with a search for variants with extremely good and bad qualities for a particular criterion and to assign appropriate points to them. Points 0 and 4 (or 10) should only be awarded if the characteristics are really extreme, that is, unsatisfactory or very good (ideal). Once these extreme points have been assigned, the remaining variants are relatively easy to fit in.

Before points can be assigned to the parameters of the variants, the evaluator must at least be clear about the assessment range and the shape of the so-called "value function" (see Figure 3.32). A value function connects values and parameter

Value scale			
Use-value analysis		Guideline VDI 2225	
Pts.	Meaning	Pts.	Meaning
0	absolutely useless solution	0	unsatisfactory
1	very inadequate solution		
2	weak solution	1	just tolerable
3	tolerable solution		
4	adequate solution	2	adequate
5	satisfactory solution		
6	good solution with few drawbacks	3	good
7	good solution		
8	very good solution	4	very good (ideal)
9	solution exceeding the requirement		
10	ideal solution		

Figure 3.31. Points awarded in use-value analysis and guideline VDI 2225

magnitudes, and its characteristic shape is determined either with the help of the known mathematical relationship between the value and the parameter or, more frequently, by means of estimates [3.28].

It is useful to draw up a chart in which the parameter magnitudes are correlated step-by-step with the value scale. Figure 3.33 shows such a scheme, incorporating the point systems of Cost–Benefit Analysis and VDI 2225.

All in all, therefore, the assignment of a value, the selection of a value function and the setting up of an assessment scheme may involve strong subjective influences. Cases with a clear, or even experimentally verified, correlation between the values and the parameters are few and far between. One such exception is the evaluation of machine noise, where the correlation between the value (that is, the protection of the human ear) and the parameter (noise level in dB) is clearly defined by ergonomics.

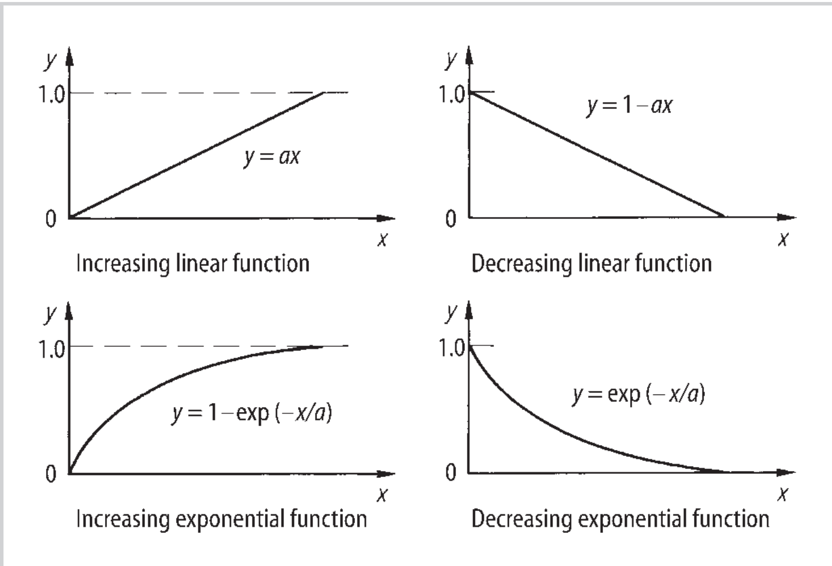


Figure 3.32. Common value functions, after [3.76]; $x \triangleq m_{ij}, y \triangleq v_{ij}$

Value Scale		Parameter magnitudes			
Use-value analysis Pts	VDI 2225 Pts	Fuel consumption g/kWh	Mass per unit power kg/kW	Simplicity of components	Service life
0	0	400	3.5	extremely complicated	$20 \cdot 10^3$
1		380	3.3		30
2	1	360	3.1	complicated	40
3		340	2.9		60
4	2	320	2.7	average	80
5		300	2.5		100
6	3	280	2.3	simple	120
7		260	2.1		140
8	4	240	1.9	extremely simple	200
9		220	1.7		300
10		200	1.5		$500 \cdot 10^3$

Figure 3.33. Chart correlating parameter magnitudes with value scales

Evaluation criteria			Parameters		Variant V_1 (e.g. *Eng.*$_1$)			Variant V_2 (e.g. *Eng.*$_2$)			…	Variant V_j			…	Variant V_m		
No.		Wt.		Unit	Magn. m_{i1}	Value v_{i1}	Weighted value WV_{i1}	Magn. m_{i2}	Value v_{ij}	Weighted value WV_{ij}		Magn. m_{ij}	Value v_{ij}	Weighted value WV_{ij}		Magn. m_{im}	Value V_{im}	Weighted value WV_{im}
1	Low fuel consumption	0.3	Fuel consumption	$\frac{g}{kWh}$	240	8	2.4	300	5	1,5	…	m_{1j}	V_{1j}	WV_{1j}	…	m_{1m}	V_{1m}	WV_{1m}
2	Lightweight construction	0.15	Mass per unit power	$\frac{kg}{kW}$	1.7	9	1.35	2.7	4	0,6	…	m_{2j}	V_{2j}	WV_{2j}	…	m_{2m}	V_{2m}	WV_{2m}
3	Simple production	0.1	Simplicity of components	-	compli-cated	2	0.2	average	5	0,5	…	m_{3j}	V_{3j}	WV_{3j}	…	m_{3m}	V_{3m}	WV_{3m}
4	Long service life	0.2	Service life	km	80 000	4	0.8	150 000	7	1,4	…	m_{4j}	V_{4j}	WV_{4j}	…	$m4_m$	V_{4m}	WV_{4m}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮		⋮	⋮	⋮
i		W_i			m_{i1}	V_{i1}	WV_{i1}	m_{i2}	V_{i2}	WV_{i2}	…	m_{ij}	V_{ij}	WV_{ij}	…	m_{im}	V_{im}	WV_{im}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮		⋮	⋮	⋮
n		W_n			m_{n1}	V_{n1}	WV_{n1}	m_{n2}	V_{n2}	WV_{n2}	…	m_{nj}	V_{nj}	WV_{nj}	…	m_{nm}	V_{nm}	WV_{nm}
		$\sum_{i=1}^{n} W_i = 1$				OV_1 R_1	OWV_1 WR_1		OV_2 R_2	OWV_2 WR_2			OV_j R_j	OWV_j WR_j			OV_m R_m	OWV_m WR_m

Figure 3.34. Completed evaluation chart with values (see Figure 3.30)

The values v_{ij} of every solution variant established in respect to every evaluation criterion are added to the list shown in Figure 3.30 in order to produce Figure 3.34.

Whenever the evaluation criteria have a different importance to the overall value of a solution, the weighting factors determined during the second step must also be taken into consideration. To that end, subvalue v_{ij} is multiplied by the weighting factor w_i ($wv_{ij} = w_i \cdot v_{ij}$). Figure 6.55 gives a practical example. The Cost–Benefit Analysis refers to the unweighted values as objective values and the weighted ones as benefit values.

Determining Overall Value

When the subvalues for every variant have been determined, the overall value must now be calculated.

In the evaluation of technical products, the summation of subvalues has become the usual method of calculation but can only be considered accurate if the evaluation criteria are independent. However, even when this condition is only satisfied approximately, the assumption that the overall value has an additive structure seems to be justified.

The overall value of a variant j can then be determined as follows:

$$\text{Unweighted: } OV_j = \sum_{i=1}^{n} v_{ij}$$

$$\text{Weighted: } OWV_j = \sum_{i=1}^{n} w_i \cdot v_{ij} = \sum_{i=1}^{n} wv_{ij}$$

Comparing Concept Variants

On the basis of the summation rule it is possible to assess variants in several ways.

Determining the maximum overall value: In this procedure the variant is judged to be the best if it has the largest overall value:

$$OV_j \to \max \quad \text{or} \quad OWV_j \to \max$$

What we have here is a relative comparison of the variants. This fact is made use of in Cost–Benefit Analysis.

Determining the rating: If a relative comparison of the variants is considered to be insufficient and the absolute rating of a variant has to be established, then the overall value must be referred to an imaginary ideal value which results from the maximum possible value as follows:

$$\text{Unweighted: } R_j = \frac{OV_j}{v_{\max} \cdot n} = \frac{\sum_{i=1}^{n} v_{ij}}{v_{\max} \cdot n}$$

$$\text{Weighted: } WR_j = \frac{OWV_j}{v_{\max} \cdot \sum_{i=1}^{n} w_i} = \frac{\sum_{i=1}^{n} w_i \cdot v_{ij}}{v_{\max} \cdot \sum_{i=1}^{n} w_i}$$

If the available information about all the concept variants allows cost estimates, then it is advisable to proceed to a separate determination of the *technical rating* R_t and the *economic rating* R_e. The technical rating is calculated in accordance with the rule we have given—that is, by division of the technical overall value of the given variant by the ideal value—and the economic rating is calculated similarly, but by reference to comparative costs. The latter procedure is suggested in VDI 2225, which relates the manufacturing costs determined for a variant to the comparative manufacturing costs C_o. In that case, the economic rating becomes $R_e = (C_o/C_{variant})$. It is possible to put, say, $C_o = 0.7 \times C_{admissible}$ or $C_o = 0.7 \times C_{minimum}$ for the cheapest variant. If the technical and economic ratings have been determined separately, then the determination of the "overall rating" of a particular variant may prove interesting. For that purpose, Guideline VDI 2225 suggests a so-called s-diagram (strength diagram) with the technical rating R_t as the abscissa and the economic rating R_e as the ordinate (see Figure 3.35). Such diagrams are particularly useful in the appraisal of variants during further developments, because they show up the effects of design decisions very clearly.

In some cases it is useful to derive the overall rating from these partial ratings and to express it in numerical form, for instance for computer processing. To that end, Baatz [3.1] has proposed two procedures, namely:

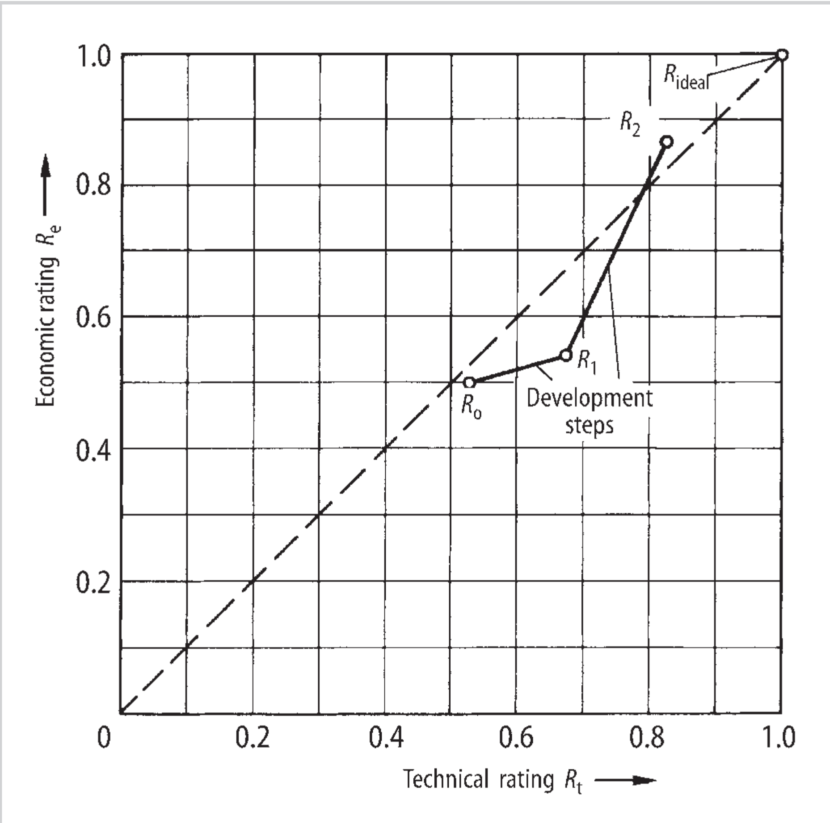


Figure 3.35. Rating diagram. After [3.36, 3.71]

- the straight-line method, based on the arithmetic mean:

$$R = \frac{R_t + R_e}{2} \tag{3.5}$$

and

- the hyperbolic method, which involves multiplying both ratings and then reducing to values between 0 and 1:

$$R = \sqrt{R_t \times R_e} \tag{3.6}$$

The two methods have been combined in Figure 3.36.

Where there are large differences between the technical and economic ratings, the straight-line method might compute a higher overall rating than is the case with lower but balanced partial ratings. Because balanced solutions should be preferred, however, the hyperbolic method is the better of the two; it helps to balance large differences in rating by its progressive reduction effect. The greater the imbalance, the greater the reduction effect on the overall value.

Rough comparison of solution variants: The method we have described relies on differentiated value scales. It is useful whenever the "objective" parameters can be stated with some accuracy and whenever clear values can be assigned to

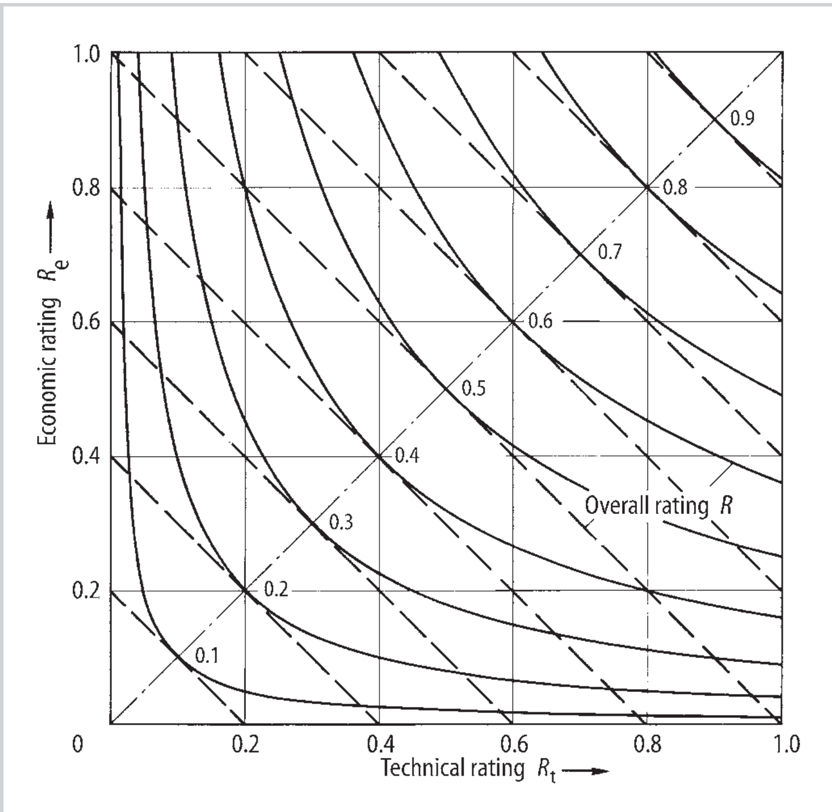


Figure 3.36. Determining the overall rating by the straight-line and hyperbolic methods. After [3.1]

		Variant						
		1	2	3	4	5	6	7
In comparison with Variant	1	-	1	0	1	0	1	0
	2	0	-	0	1	0	0	0
	3	1	1	-	1	0	1	0
	4	0	0	0	-	0	0	0
	5	1	1	1	1	-	1	1
	6	0	1	0	1	0	-	0
	7	1	1	1	1	0	1	-
Sum		3	5	2	6	0	4	1
Rank		4	2	5	1	7	3	6

1 ≙ better   0 ≙ not better

Figure 3.37. Binary evaluation of solution variants. After [3.15]

them. If these conditions cannot be satisfied, relatively fine evaluations based on a differentiated value scale constitute a questionable and expensive method. The alternative here is a rough evaluation involving the application of a particular evaluation criterion to two variants at a time and the selection of the best in each case. The results are entered into a so-called *dominance matrix* [3.15] (see Figure 3.37). From the sum of the columns it is possible to establish a ranking order. If such matrices of individual criteria are combined into an overall matrix, an overall ranking order can be established, either by addition of the preference frequencies or by addition of all the column sums. While this method is comparatively easy and quick, it is not nearly as informative as the other procedures we have discussed.

Estimating Evaluation Uncertainties

The possible errors or uncertainties of the proposed evaluation methods fall into two main groups, namely subjective errors and procedure-inherent shortcomings.

Subjective errors can arise through:

- Abandonment of the neutral position, that is, through bias and partiality. The bias may be hidden from designers, for instance when they compare their own designs with those of others. Hence an evaluation by several persons, if possible from various departments, is always advisable. It is equally important to refer to the different variants in neutral terms, for instance as A, B, C rather than as "Smith's Proposal", etc., since otherwise unnecessary identifications and emotional overtones may be introduced. Systematisation of the procedure also helps to reduce subjective influences.

- Comparison of variants by application of (the same) evaluation criteria not equally suited to all the variants. Such mistakes arise even during the determination of the parameters and their association with the evaluation criteria. If it is impossible to determine the parameter magnitudes of individual variants for certain evaluation criteria, then these criteria must be reformulated or dropped in case they lead to mistaken evaluations of the individual variants.
- Evaluation of variants in isolation instead of successively by application of the established evaluation criteria. Each criterion must be applied to all the variants in turn (row-by-row in the evaluation chart) to eliminate any bias in favour of a particular variant.
- Pronounced interdependence of the evaluation criteria.
- Choice of unsuitable value functions.
- Incompleteness of evaluation criteria. This defect can be minimised if one of the checklists for design evaluation appropriate to the relevant design phase is followed (see Figures 6.22 and 7.148).

Procedure-inherent shortcomings of the recommended evaluation methods are the result of the almost inevitable "prognostic uncertainty" arising from the fact that the predicted parameter magnitudes and also the values are not precise, but subject to uncertainty and to random variation. These mistakes can be greatly reduced by estimates of the mean error.

With regard to prognostic uncertainty, it is therefore advisable not to express the parameters in figures unless this can be done with some accuracy. It is preferable to use verbal estimates (for instance high, average, low) which do not claim to be precise. Numerical values, by contrast, are dangerous because they introduce a false sense of certainty.

Uncertainties in the evaluation are not only caused by prognostic uncertainty, but also through uncertainties in the formulation of requirements and solution descriptions. To be able to process such vague information in a quantitative way, fuzzy logic, and its extension into fuzzy-MADM (multi-attribute decision making), can be used [3.49]. These procedures use so-called fuzzy sets to describe these imprecise numbers and ranges and calculate their combined averages. The result is a fuzzy overall value for every solution variant.

A more detailed analysis of evaluation procedures for the purpose of judging their reliability and also for comparative purposes has been carried out by Feldmann [3.15] and Stabe [3.67]. The latter also provides an extensive bibliography. If there is an adequate number of evaluation criteria, and if the subvalues of a particular variant are fairly balanced, then the overall value will be subject to a balancing statistical effect, and partly too optimistic and partly too pessimistic individual values will more or less balance out.

Searching for Weak Spots

Weak spots can be identified from below average values for individual evaluation criteria. Careful attention must be paid to them, particularly in the case of promising variants with good overall values, and they ought if possible to be eliminated during further development. The identification of weak spots may be facilitated by graphs of the subvalues—for instance, by the so-called value profiles illustrated in Figure 3.38. In it, the lengths of the bars correspond to the values and the thicknesses to the weightings. The areas of the bars then indicate the weighted subvalues, and the cross-hatched area the overall weighted value of a solution variant. It is clear that, in order to improve a solution, it is essential to improve those subvalues that provide a greater contribution to the overall value than the rest. This is the case with the evaluation criteria that have an above average bar thickness (great importance) but a below average bar length. Apart from a high overall value, it is important to obtain a balanced value profile, with no serious weak spots. Thus, in Figure 3.38, variant 2 is better than variant 1, although both have the same overall weighted value.

There are also cases in which a minimum permissible value is stipulated for all sub-values; that is, any variant that does not fulfil this condition has to be rejected, and all variants that do fulfil it are developed further. In the literature this procedure is described as the "determination of satisfactory solutions" [3.76].

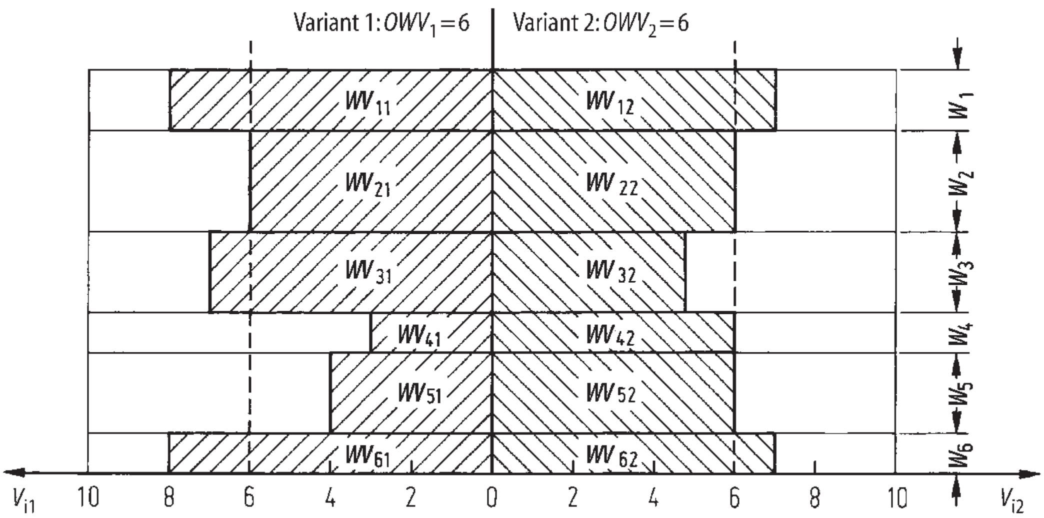


Figure 3.38. Value profiles for the comparison of two variants ($\sum w_i = 1$)

2. Comparison of Evaluation Procedures

Table 3.4 lists the individual steps in the evaluation procedures we have described and also the similarities and differences between Cost–Benefit Analysis and Guideline VDI 2225, which are based on similar principles.

Table 3.4. Individual steps in evaluation, and comparison between use-value analysis and Guideline VDI 2225

Step		Cost–Benefit Analysis	VDI Guideline 2225
1	*Identification of objectives* or *evaluation criteria* for the evaluation of concept variants with the aid of the requirements list and a checklist	Construction of a hierarchically related system of design objectives (objectives tree) based on the requirements list and other general requirements	Compilation of important technical characteristics and also of the minimum demands and wishes of the requirements list
2	*Analysis of the evaluation criteria* for the purpose of determining their *weighting* to the overall value of the solution. If necessary, determination of weighting factors	Step-by-step weighting of the objective criteria (evaluation criteria) and if necessary elimination of unimportant criteria	Determination of weighting factors only if evaluation criteria differ markedly in importance
3	*Compilation of parameters* applicable to the concept variants	Construction of an objective parameter matrix	Not generally included
4	*Assessment of the parameter magnitudes* and assignment of values (0–10 or 0–4 points)	Construction of objective value matrix, with the help of a points system or value functions; 0–10 points	Assessment of characteristics by points (0–4 points)
5	*Determination of the overall value* of the individual concept variants, generally by reference to an ideal solution (rating)	Construction of a use-value matrix with due regard to the weightings; determination of overall values by summation	Determination of a technical rating by summation, with or without weightings based on an ideal solution. If necessary determination of an economic rating based on manufacturing costs
6	*Comparison of concept variants*	Comparison of overall use-values	Comparison of the technical and economic ratings. Construction of an s-(strength) diagram
7	*Estimation of evaluation uncertainties*	Estimation of objective parameter scatter and use–value distribution	Not explicitly included
8	*Search for weak spots* for the purpose of improving selected variants	Construction of use-value profiles	Identification of characteristics with a few points only

The individual steps of Cost–Benefit Analysis are more highly differentiated and more clear-cut but involve more work than those of Guideline VDI 2225. The latter is more suitable when there are relatively few and roughly equivalent evaluation criteria, which is frequently the case during the conceptual phase, and also for the evaluation of certain form design areas during the embodiment phase.

The essence of evaluation procedures has been described on the basis of existing evaluation methods. However, these methods have been consolidated and the terms clarified. Specific suggestions for the use of these methods during the conceptual phase are given in Section 6.5.2, and during the embodiment phase in Section 7.6.

4 Product Development Process

In the previous chapters we examined the fundamentals on which design work should be built to best advantage. They form the basis of a systematic approach which practising designers can follow, regardless of their speciality. The approach is not based on one method but applies known and less well known methods where they are most suitable and useful for specific tasks and working steps.

4.1 General Problem Solving Process

An essential part of our own problem solving method involves step-by-step *analysis* and *synthesis*. In it we proceed from the *qualitative* to the *quantitative*, each new step being more concrete than the last.

In the following sections we propose plans and procedures that should be regarded as mandatory for the general problem solving process of planning and designing technical products, and as guidance for the more concrete phases of the design process. These plans and procedures assist in identifying what, in principle, has to be done, but of course they must be adapted to specific problem situations.

All procedural plans proposed in this book have to be considered as *operational guidelines for action* based on the pattern of technical product development and the logic of stepwise problem solving. According to Müller [4.17], they are process models that are suitable for describing in a rational way the approach necessary to make complex processes comprehensible and transparent.

Thus, these procedural plans are not descriptions of *individual thinking processes* as described in Section 2.2.1, and are not determined by personal characteristics. In a practical application of these procedural plans, the operational guidelines for action blend with individual thinking processes. This results in a set of individual planning, acting and controlling activities based on general procedures, specific problem situations and individual experiences.

As discussed in Section 2.2.1, the suggested procedural plans are meant to be guidelines and not rigid prescriptions. However, they have to be regarded as essentially sequential because, for example, a solution cannot be evaluated before it has been found or elaborated. On the other hand, the procedural plans have to be adapted to specific situations in a flexible manner. It is, for example, possible to leave out certain steps or order them in another sequence. It may be necessary or useful to repeat certain steps at a higher information level. Furthermore, special procedures (adapted from the more general plans proposed here) may be appropriate in specific product domains.

Given the complexity of the product development process and the many methods that have to be applied, not adopting a procedural plan would leave designers with an unmanageable number of possible approaches. It is therefore necessary for designers to learn about the design process and the application of individual methods, as well as the working and decision making steps proposed in the procedural plans.

The activity of planning and designing was described in Section 2.2.3 as information processing. After each information output, it might become necessary to improve or increase the value of the result of the last working step. That is, to repeat the working step at a higher information level, or to execute other working steps until the necessary improvements have been achieved.

Repeating working steps is the process of *iteration* by which one approaches a solution step-by-step until the result seems satisfactory. The so-called iteration loop can also be observed in the basic thinking processes, for example in the TOTE model (see Section 2.2.1). Such iteration loops are almost always required and occur continuously within and between steps. The reasons for this are that the interrelationships are often so complex that the desired solution cannot be achieved in one step and that information is frequently needed from a subsequent step. The iteration arrows in procedural plans clearly indicate this fact. In subsequent chapters, strategies for reducing, or even avoiding, such iteration loops are presented. It is therefore important that the procedural plans proposed are not considered rigid and purely sequential.

A systematic approach aims to keep the iteration loops as small as possible in order to make design work effective and efficient. It would be a disaster, for example, if the design team had to start again at the beginning having reached the end of a product development. This would correspond to an iteration loop covering the whole of the product development process.

The division in working and decision making steps ensures necessary and permanent links between *objectives*, *planning*, *execution* (organisation) and *control* [4.3, 4.29]. With these links, we can, following Krick [4.15] and Penny [4.21], construct a basic scheme for the general problem solving process (see Figure 4.1).

Every task involves an initial *confrontation* of the problem, which involves elucidating what is known or not (yet) known. The intensity of this confrontation depends on the knowledge, ability and experience of the designers, and on the particular field in which they are engaged. In all cases, however, more detailed *information* about the task itself, about the constraints, about possible solution principles and about known solutions for similar problems is extremely useful since it clarifies the precise nature of the requirements. This information can also reduce confrontation and increase confidence that solutions can be found.

Next comes the *definition* phase, where the essential problems (the crux of the task) are defined on a more abstract plane, in order to set the objectives and main constraints. Such solution-neutral definitions open the way to an unconstrained

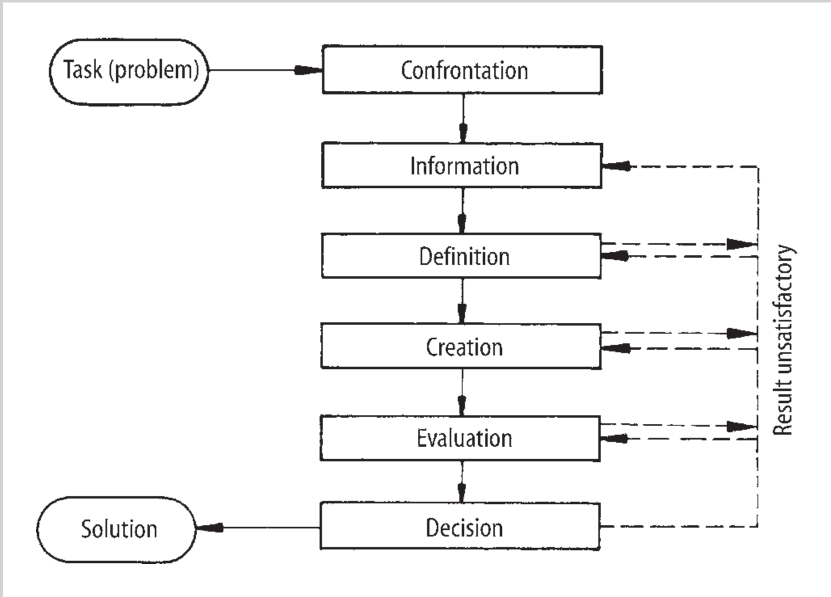


Figure 4.1. General problem solving process

search for solutions because this abstract definition encourages a search for more unconventional solutions.

The next step is *creation*, where solutions are developed by various means and then varied and combined using methodical guidelines. If the number of variants is large, there must also be an *evaluation* which is then used to select the best variant through a *decision*. Because each step of the design process must be evaluated, evaluation serves as a check on progress towards the overall objective.

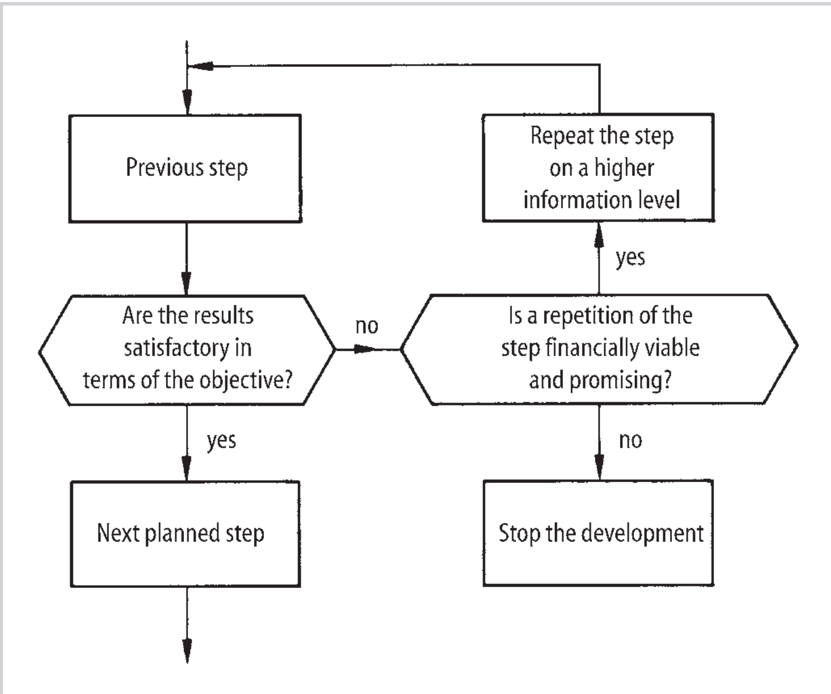


Figure 4.2. General decision making process

Decisions involve the following considerations (see Figure 4.2):

- If the result from the previous step meets the objective, the next step can be started.
- If the results are incompatible with the objective, the next step should not be taken.
- If resources permit repetition of the previous step (or if necessary several preceding steps), and good results can be expected, the step must be repeated on a higher information level.
- If the answer to the previous question is no, the development must be stopped.

Even if the results of a particular step do not meet the objectives, they might nevertheless prove interesting if the objectives are wholly or partly changed. In this case, there should be an investigation to see whether the objectives can be changed or if the results can be used for other applications. This whole process, leading from confrontation through creation to decision, must be repeated in each successive, increasingly concrete, phase of the design process.

4.2 Flow of Work During the Process of Designing

Today's conditions for product design and development demand careful planning of:

- the required activities for the proposed project
- the timing and scheduling of these activities
- the project and product costs.

The activities and their durations strongly depend on the type of task, in particular whether the task is for an original, adaptive or variant design.

4.2.1 Activity Planning

The flow of work during the process of designing has been described in both general terms as well as domain and product-specific terms in VDI Guidelines 2221 and 2222 [4.24, 4.25] (see Figure 1.9). In line with these guidelines, the next sections provide an extensive description of this flow of work, focused on mechanical engineering. The description is essentially based on the fundamentals of technical systems (see Section 2.1), the fundamentals of the systematic approach (see Section 2.2), and the general problem solving process (see Section 4.1). The aim is to adapt the general statements to the requirements of the mechanical engineering design process and to incorporate the specific working and decision making steps for this domain. In principle, the planning and design process proceeds from the planning and clarification of the task, through the identification of the required functions, the elaboration of principle solutions, the construction of modular structures, to the final documentation of the complete product [4.18].

In addition to the planning of the specific tasks described in the guidelines mentioned above, it is useful and common to divide the planning and design process into the following *main phases*:

- Planning and task clarification: specification of information
- Conceptual design: specification of principle solution (concept)
- Embodiment design: specification of layout (construction)
- Detail design: specification of production.

As we will see later on, it is not always possible to draw a clear borderline between these main phases. For example, aspects of the layout might have to be addressed during conceptual design, or it might be necessary to determine some production processes in detail during the embodiment phase. Neither is it possible to avoid backtracking, for example during embodiment design when new auxiliary functions may be discovered for which principle solutions have to be found. Nevertheless, the division of the planning and control of a development process into main phases is always helpful.

The working steps proposed for each of the main phases are termed the *main working steps* (see Figure 4.3). The results of these main working steps provide the basis for the subsequent working steps. Many lower level working steps are required to realise these results, such as collecting information, searching for solutions, calculating, drawing and evaluating. Each of these working steps is accompanied by indirect activities such as discussing, classifying and preparing. The *operational main working steps* listed in the procedural plans proposed in this chapter are considered to be the most useful strategic guidelines for a technical domain. Guidelines that are not listed include, for example, those related to basic problem solving, collecting information and verifying results. This is because they can usually only be recommended in relation to a specific problem and a particular designer. Recommendations for such elementary working steps will, where possible, be given in the sections describing individual methods and those dealing with practical applications.

After the main phases, and some of the more important main working steps, *decision making steps* are required. The decision making steps listed are the main ones—those that end a main phase or working step, which after an appropriate assessment of the results, allow the main flow of work to proceed. It is also possible, because the result of a decision making step was unsatisfactory, that certain steps will have to be repeated. The smallest possible iteration loop is desirable.

Again, the individual test and decision making steps (see for example the TOTE model in Section 2.2.1) that are required for every single action have not been listed separately. This would have been impossible because such decisions are determined by the approach of individual designers and by particular problem situations.

The decision to stop a development that ceases to be viable, as discussed in Section 4.1, is not mentioned explicitly in the individual decision making steps of the procedural plans. One should, however, always explicitly consider this possibility because an early and clear decision to halt a hopeless situation will, in the end, minimise disappointment and cost.

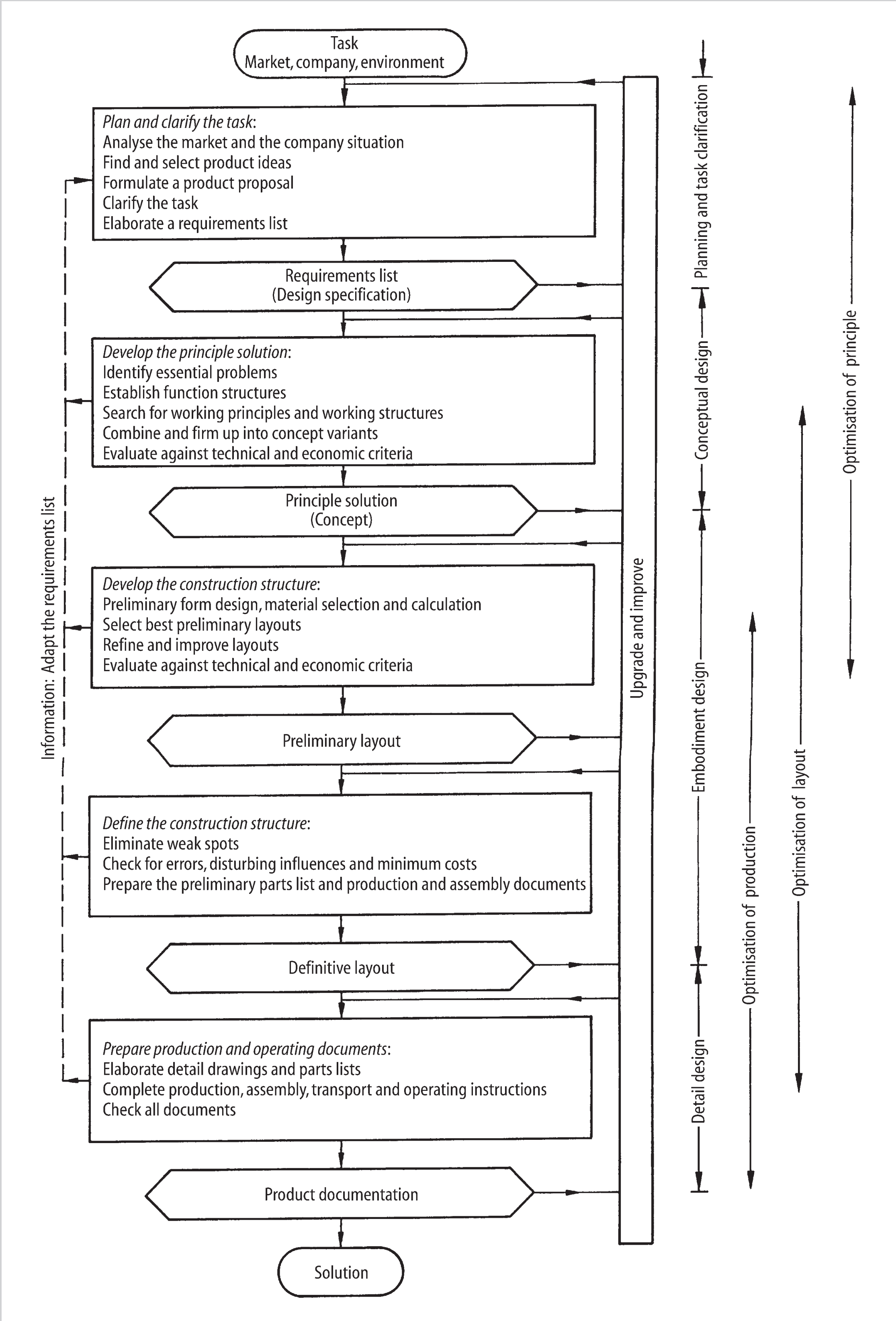


Figure 4.3. Steps in the planning and design process

In all cases procedural plans should be applied in a flexible manner and adapted to the particular problem situation. At the end of each main working and decision step, the overall approach should be assessed and adjusted if necessary.

The four main phases are outlined below.

1. Planning and Task Clarification

The product development task is given to the engineering department by the marketing department, or a special department responsible for product planning, see also Sections 3.1 and 5.1.

Irrespective of whether the task is based on a product proposal stemming from a *product planning* process or on a specific customer order, it is necessary to clarify the given task in more detail before starting product development. The purpose of this *task clarification* is to collect information about the requirements that have to be fulfilled by the product, and also about the existing constraints and their importance.

This activity results in the *specification of information* in the form of a *requirements list* that focuses on, and is tuned to, the interests of the design process and subsequent working steps (see Section 5.2). The conceptual design phase and subsequent phases should be based on this document, which must be updated continuously (this is indicated by the information feedback loop in Figure 4.3).

2. Conceptual Design

After completing the task clarification phase, the conceptual design phase determines the principle solution. This is achieved by abstracting the essential problems, establishing function structures, searching for suitable working principles and then combining those principles into a working structure. Conceptual design results in the specification of a *principle solution* (concept).

Often, however, a working structure cannot be assessed until it is transformed into a more concrete representation. This concretisation involves selecting preliminary materials, producing a rough dimensional layout, and considering technological possibilities. Only then, in general, is it possible to assess the essential aspects of a solution principle and to review the objectives and constraints (see Section 2.1.7). It is possible that there will be several principle solution variants.

The representation of a principle solution can take many forms. For existing building blocks, a schematic representation in the form of a function structure, a circuit diagram or a flow chart may be sufficient. In other cases a line sketch might be more suitable, and sometimes a rough scale drawing is necessary.

The conceptual design phase consists of several steps (see Chapter 6), none of which should be skipped if the most promising principle solution is to be found. In the subsequent embodiment and detail design phases it is extremely difficult or impossible to correct fundamental shortcomings of the solution principle. A lasting and successful solution is more likely to spring from the choice of the most appropriate principles than from exaggerated concentration on technical details.

This claim does not conflict with the fact that problems may emerge during the detail design phase, even in the most promising solution principles or combinations of principles.

The solution variants that have been elaborated must now be evaluated. Variants that do not satisfy the demands of the requirements list have to be eliminated; the rest must be judged by the methodical application of specific criteria. During this phase, the chief criteria are of a technical nature, though rough economic criteria also begin to play a part (see Sections 3.3.2 and 6.5.2). Based on this evaluation, the best concept can now be selected.

It may be that several variants look equally promising, and that a final decision can only be reached on a more concrete level. Moreover, various form designs may satisfy one and the same concept. The design process now continues on a more concrete level referred to as embodiment design.

3. Embodiment Design

During this phase, designers, starting from a concept (working structure, principle solution), determine the construction structure (overall layout) of a technical system in line with technical and economic criteria. Embodiment design results in the specification of a *layout*.

It is often necessary to produce several *preliminary layouts* to scale simultaneously or successively in order to obtain more information about the advantages and disadvantages of the different variants.

After sufficient elaboration of the layouts, this design phase also ends with an evaluation against technical and economic criteria. This results in new knowledge on a higher information level. Frequently, the evaluation of individual variants may lead to the selection of one that looks particularly promising but which may nevertheless benefit from, and be further improved by, incorporating ideas and solutions from the others. By appropriate combination and the elimination of weak spots, the best layout can then be obtained.

This *definitive layout* provides a means to check function, strength, spatial compatibility, etc., and it is also at this stage (at the very latest) that the financial viability of the project must be assessed. Only then should work start on the detail design phase.

4. Detail Design

This is the phase of the design process in which the arrangement, forms, dimensions and surface properties of all of the individual parts are finally laid down, the materials specified, production possibilities assessed, costs estimated, and all the drawings and other production documents produced [4.28] (see also [4.26]). The detail design phase results in the *specification of information* in the form of *production documentation*.

It is important that designers should not relax their vigilance at this stage, otherwise their ideas and plans might change out of all recognition. It is a mistake to think that detail design poses subordinate problems lacking in importance or interest. As we said earlier, difficulties frequently arise from lack of attention to detail. Quite often corrections must be made during this phase and the preceding steps repeated, not so much with the overall solution in mind, as to improve assemblies and components as well as reduce costs.

5. Overall Design Process

In the flow diagram (see Figure 4.3), the main themes are:

- optimisation of principle
- optimisation of layout
- optimisation of production.

Clearly the description above is a generalisation of actual processes. In practice a clear distinction between the working steps and their results cannot always be made, nor is it necessary to do so. However, it is useful for designers to be aware of the main process flow and tasks described in order to plan their work and to avoid forgetting something.

Figure 4.3 does not include the production of models and prototypes because the information they supply may be needed at any point in the design process and so cannot be fitted into any particular slot. In many cases, it is even necessary to develop models and prototypes during the conceptual phase, particularly when they are intended to clarify fundamental questions in, say, the precision engineering, electronics and mass production industries. Due to the one-off nature of heavy and process engineering, the cost and time required to produce prototypes normally makes them uneconomic or infeasible. However, it is possible to test parts of the proposed plant or equipment by building partial prototypes within existing plant and equipment or by using specific test facilities. In batch production it is common to produce prototypes well before production starts and also to undertake a pre-production run to ensure that production will run smoothly. These pre-production products can still be sold.

Figure 4.3 also does not indicate when work has to be subcontracted, because this depends upon the type of product.

The execution of orders is usually part of product development, although in the case of size ranges and modular products it can take place quite late in the process.

If on receiving an order, only existing documents are used, and only production instructions, subcontractor orders, parts lists, etc., need to be compiled, no product development is required. So apart from tender drawings, layout drawings and assembly plans, no further design work is needed, and in many cases these drawings and plans can be generated automatically using variant design software.

Upon looking at Figure 4.3, and after reading about the methods described in the following chapters, practising designers may well object to the process on the basis that they lack the time to go through every one of the many steps. They should bear in mind that:

- Most of the steps are performed in any case—albeit unconsciously—although they are often carried out too quickly, leading to unforeseen consequences.

- This deliberate step-by-step procedure, on the other hand, ensures that nothing essential has been overlooked or ignored, and is therefore indispensable in the case of original designs.
- In the case of adaptive designs, it is possible to resort to time-tested approaches and to reserve the procedure described for where it offers special benefits; for example, when improving a specific detail, in which case the steps should be undertaken focusing on this detail.
- If designers are expected to produce better results, then they must be given the extra time the systematic approach demands, although experience has shown that only a little extra time is needed for a stepwise procedure.
- Scheduling becomes more accurate if the step-by-step method is followed rigorously.

4.2.2 Timing and Scheduling

Products will only be successful when they:

- satisfy the customer needs (requirements)
- reach the market at the right time
- are sold at the right price.

This section focuses on the second prerequisite, because designers often underestimate the importance of time-to-market and are not familiar with the methods and tools used for timing and scheduling. We only introduce the basic approaches. Details have to be obtained from the literature.

Two constraints determine the planning difficulty:

- the project or design result must be finished at a certain point in time, and intermediate results are required on specific dates
- not every task can be executed by every member of the team, i.e. there is a resource constraint.

Network planning is one of the most important planning tools [4.7,4.8]. A network plan is used to estimate the overall project length and resource requirements. The graphical representation shows the logical links between the required project tasks and the resources assigned to these tasks.

Creating a network plan involves performing three main steps:

- Structure analysis to identify and describe the links and dependencies between the project tasks.
- Time analysis to identify the necessary duration for each task along with a feasible starting date for each of the main steps.
- Resource analysis to allocate the various tasks to individual team members. In the first instance this should be based on their competences, followed by their

Table 4.1. Procedure for creating a network plan

	Activity	Explanation
1.	Determine product structure	In general the structure of an existing similar product is adapted
2.	Determine the tasks necessary to create the individual product elements	For every product element and for the overall product, the tasks include the following to an appropriate level: ● solution finding ● investigation ● embodiment ● calculation
3.	Establish logical and temporal dependencies between individual tasks	Dependencies between tasks have to be identified and documented as unambiguous IF–THEN statements: e.g. IF the shaft diameter is determined, THEN the shaft–hub connection can be fixed
4.	Establish the duration of the tasks	● interview those with relevant experience ● compare with similar tasks ● document completed tasks ● estimate
5.	Fix milestones (these are used to check whether the work and schedule have been achieved; a milestone trend analysis allows the prediction of the success or failure of a project)	**Types of milestones:** **Event-driven:** The content of a milestone has to be defined precisely. A milestone is reached when the available working results meet the defined content of that milestone **Application:** Mostly used as the final milestone for the design of an assembly **Time-driven:** The milestone is reached at a certain point in time or after a certain time interval has elapsed **Application:** For large tasks when it is not possible to define clear intermediate results **Point of no return:** Event or point in time after which the results achieved must not be changed further **Application:** Securing intermediate results, e.g. against customers changing requirements **Review:** Point in time at which clearly defined results have to be explicitly signed off or released **Application:** The embodiment of expensive and complex assemblies or components is signed off by production or, in some cases, by the customer
6.	Determine necessary and possible float times for the tasks	Float times serve to manage risk in order to avoid endangering the project plan when delays occur and are applied, in particular, for novel tasks
7.	Create network plan (usually using special software tools: e.g. Microsoft Project, Super-Project-Expert)	A network plan shows in graphical and tabular form all of the dependencies between the tasks and milestones, and is used to determine the course of a project
8.	Create project calendar	The project calendar shows the exact number of working days available for the duration of the project
9.	Select resources and allocate them to tasks in the network plan	The selection is based on the required competencies and the availability of resources during the planned period of the project
10.	Create a resource calendar and allocate to the network plan	For every employee an individual calendar is created showing his or her available working time during the duration of the project: holidays, training days, etc. must be taken into account

Table 4.1. (continued)

	Activity	Explanation
11.	Run through the plan	After the resources and the individual calendars are allocated to the network plan, the first run-through is undertaken
12.	Evaluate the plan	• Can the project milestone be achieved? • What is the critical path? (i.e. the sequence of tasks with no float times that determines the overall duration of the project)
13.	Optimise the plan	The plan can be optimised and corrected by: • increasing the resources available • moving deadlines • reducing the number of tasks • changing the sequence of tasks • altering the content of the tasks
14.	Sign off the plan	The project plan is released through the signature of the manager responsible, and, where appropriate, by the customer
15.	Monitor the project	All project parameters, such as deadlines, costs and risks are continuously monitored and reported

availability, which can be limited because they may be absent due to training courses, illness, holidays, etc., or because they have been allocated to other projects.

In general the product structure is used as the basis for planning the task structure. The product structure determines the main assembly groups and components that have to be designed and, as a consequence, the majority of the tasks.

Table 4.1 shows the procedure for creating a network plan and the individual steps. Figure 4.4 shows part of a network plan, in this case a Gantt chart. The individual tasks are represented by bars. Their dependencies result from logical or possible working sequences, e.g. input/output requirements, where one task must have been completed before the next can be started.

A network plan not only shows project duration, resource requirements and allocation of team members to tasks, but also float times and the critical path of the project. The float times indicate how much the start or end of a task or series of tasks can be delayed without jeopardising the overall lead time of the project. The critical path contains those tasks that have no float times and therefore determine the overall duration of the project.

4.2.3 Planning Project and Product Costs

The cost price is the basis for determining the selling price and is therefore crucial to the success of the product. The cost price is influenced by the production costs and the associated project costs. Design and development are the costliest items contributing to the project cost, so engineering departments carry a great responsibility.

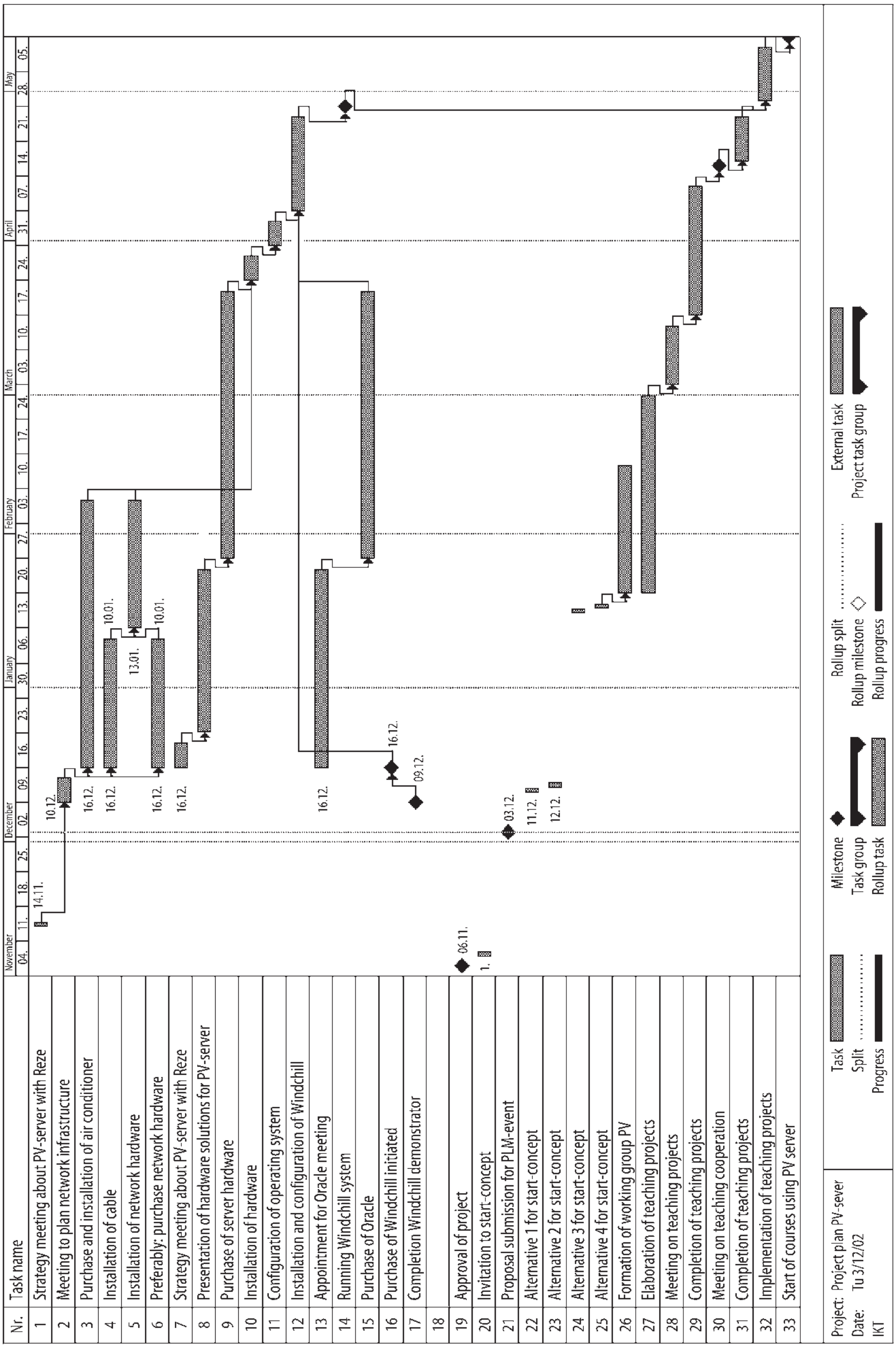


Figure 4.4. Example of a network plan

In order to meet the target cost price, engineering departments not only have to keep production costs to a minimum (see Chapter 11 for more details), but also the design and development costs. Depending on the batch size, the latter costs can represent a large share of the cost price.

To estimate the design and development costs, a network plan can also be used because the main costs incurred by engineering departments are staff costs. Support costs, such as facilities, CAD systems, external consultants, etc., are usually much lower. Using the network plan, costs can be assigned to the allocated resources using the appropriate hourly rate. The distribution of the costs with time can be represented by a cost plan [4.9], which is important when estimating the project budget.

4.3 Effective Organisation Structures

4.3.1 Interdisciplinary Cooperation

Designers cannot work independently of their environment—they depend on the results produced by others and others depend on their results. They are members of their departments, which in turn are parts of the company. Only the coordinated activities of all participants will lead to a satisfactory overall result [4.11, 4.22]. To achieve this, the responsibilities, tasks, etc., for every individual are specified by the organisational and operational structures:

- The *organisational structure* specifies the responsibilities and tasks for individuals, departments and standing committees, and relates these in a hierarchy.
- The *operational structure* specifies the various procedures.

The design and development process is made more efficient through the following actions:

- reducing inner iterations, i.e. repetition of the same activity within a working step
- reducing outer iterations, i.e. jumping back to a working step that has already been completed or even repeating a design phase
- omitting working steps
- executing working steps in parallel.

In particular, the last action has the potential to reduce the overall project time significantly. To achieve these four actions, the following prerequisites must be met:

- The product has to be structured in such a way that the properties of its systems, subsystems and system elements can be modelled precisely and unambiguously during every step of the process. Chapter 9 proposes some possible product structures.
- The interfaces between the process steps have to be defined precisely and unambiguously.
- The process steps have to be independent.

When these prerequisites have been met and interdisciplinary teams formed, then *Simultaneous or Concurrent Engineering* can be introduced. Simultaneous or Concurrent Engineering involves goal-oriented, interdisciplinary (interdepartmental) collaboration and parallel working throughout the development of the product, the production process and the sales strategy. It covers the total product life cycle and requires firm project management [4.1]. Experiences of its application in industry can be found in [4.12, 4.14]. Figure 1.4 highlights the intensive information flows that occur between departments. In a simultaneous engineering process the activities of the various departments run in parallel or at least have significant overlap. Intensive contacts with customers are encouraged, many suppliers are integrated in the process, see Figure 4.5 [4.5, 4.13, 4.23], and the product is monitored until the end of its working life.

For the duration of the project, a team is formed consisting not only of members of the design and development department but also those from other departments involved in the product creation process. This team, which is formed as early as possible, is led by a project manager, works independently, but has to report directly to the Board of Management or Head of Development. Departmental boundaries are thereby transcended. The team can operate as a virtual team; that is, without a visible organisational form. Characteristics of team structures and their importance can be found in [4.6, 4.27]. The objectives of this type of organisation and working procedure are:

- shorter development times
- faster product realisation
- reduction of product and product development costs
- improved quality.

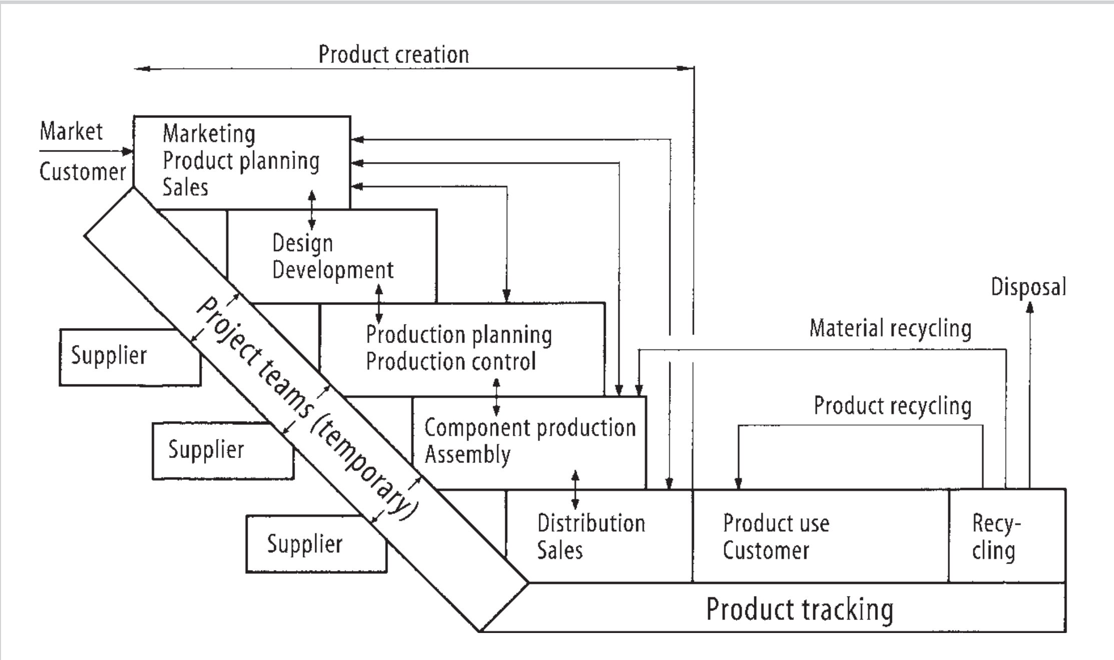


Figure 4.5. Product creation and tracking processes using Simultaneous Engineering, showing the overlapping activities of different disciplines, the formation of a project team and close contact with customers and suppliers

Simultaneous or Concurrent Engineering changes a designer's work as follows [4.20]:

- Working in an interdisciplinary team requires the adaptation of language and terminology.
- A closer, more direct exchange of information takes place through the early involvement of other departments and disciplines.
- More electronic information and communication technologies are used, e.g. data processing systems, CAD, multimedia, etc.
- A project management process with schedules and milestones is imposed so that design work has to be structured more systematically.
- Activities are run in parallel and therefore have to be coordinated accordingly.
- Individual responsibility for the assigned problems and tasks has to be accepted in line with team decisions.
- Contact with suppliers and customers becomes more intense.

It is useful to build a small core team with the experts who are responsible for design, production planning, marketing and sales. The composition of the team depends on the particular problem and type of product. This core team is complemented by experts from quality, assembly, electronics, software, recycling, etc., as and when needed and who may only participate for short periods of time. In such a team the knowledge and experience from neighbouring disciplines (see Figure 4.6) are more or less automatically incorporated into the project. This integration of a wide range of expertise significantly improves the realisation of the project goals and the ability to meet the constraints, as discussed in Section 2.1.7 and in accordance with Figure 2.15.

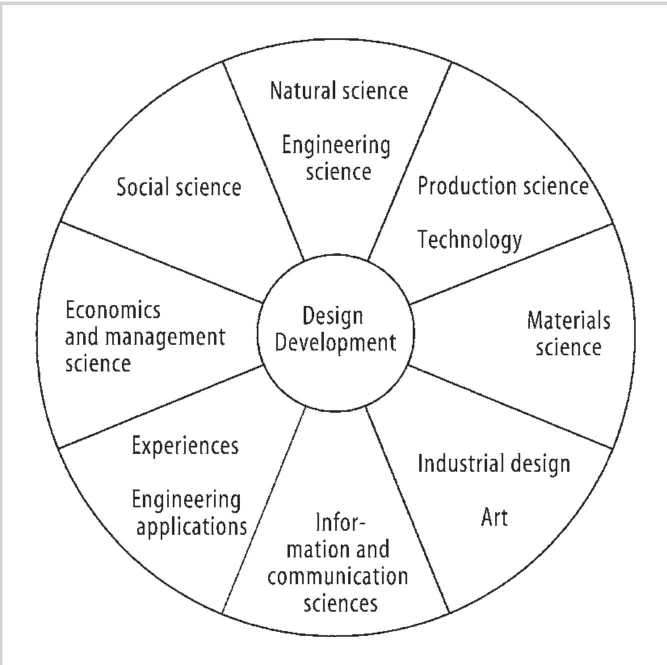


Figure 4.6. Related knowledge domains that support design and development

The advantages of an interdisciplinary team are:

- increased availability of knowledge and mutual stimulation
- better control of the product and the process—achieved by questioning issues and identifying contradictions
- increased motivation through direct participation and information sharing
- immediate responses to situations without the need to seek and wait for approval from higher levels in the hierarchy.

When the focus is on lean production, information and decision chains must become shorter. To facilitate this it is often necessary to build temporary project groups whose members are released from the departmental hierarchy for the duration of the project. The designer who previously worked within the confines of his discipline-based department, where he or she could easily call upon colleagues for advice and support, now has to work much more independently and within less familiar surroundings. To work in such project teams, a number of skills are required that go beyond the usual discipline-based ones [4.19,4.20] (see Figure 4.7). These issues must be taken into account when selecting the project leader.

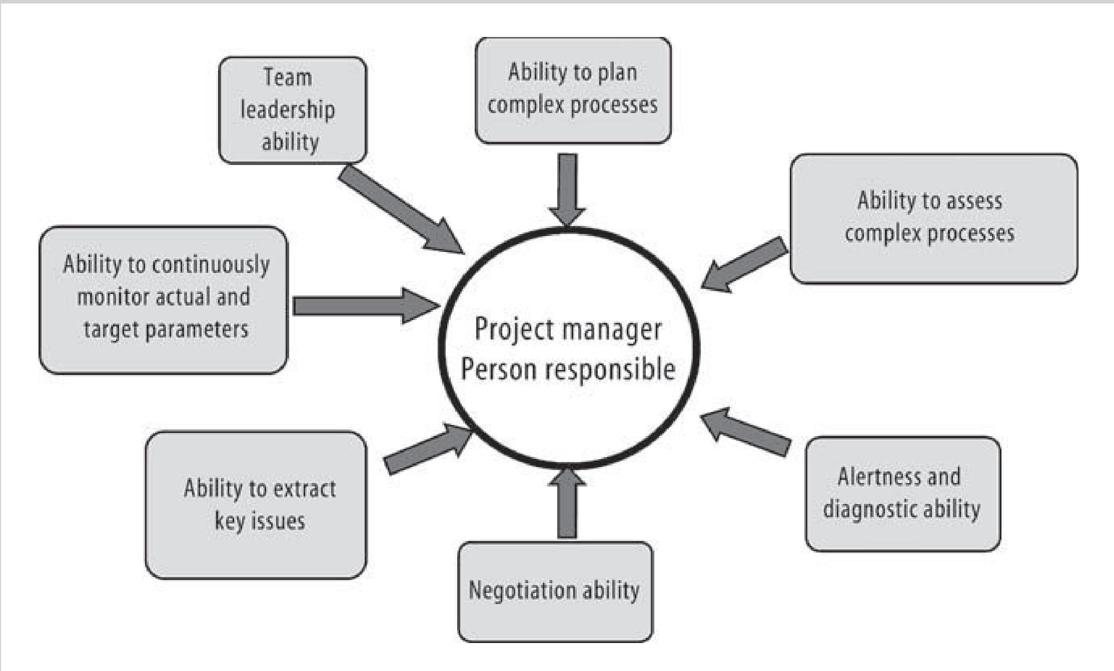


Figure 4.7. Abilities required of project managers

4.3.2 Leadership and Team Behaviour

Strong project leadership is necessary when developing new products in teams that are independent of departmental structures. Project leaders must have a good knowledge of the relevant technology and design methods as well as the characteristics of good problem solvers (see Section 2.2.2). Only then are they able to lead a team of experts from different fields to achieve the project goals and to cope with the tasks assigned to them [4.20].

Project leaders and their teams can use the systematic approach presented in this book as an effective means of support. They can use it to initiate and check their approach, to select suitable methods, define decision steps (milestones), and apply established design principles. Depending on the problem, project leaders and teams need to be willing to adapt their approaches and methods on the basis of importance and urgency. Project leaders must not be dogmatic in their leadership style, must utilise the manifold skills in the team, must provide every team member with individual freedom of action, and must demonstrate decisiveness when it matters. Leadership involves:

Providing timely information by:

- pointing out deviations from the project plan as early as possible
- managing information in a balanced and uniform manner.

Steering individual activities carefully in line with a systematic approach by:

- planning the main project parameters such as deadlines, costs and resources
- pursuing these project targets
- estimating the effort and consequences of any changes
- updating the project plan when necessary.

Representing the team effectively by:

- managing reporting and documentation
- taking personal responsibility for team presentations, etc.

Fostering team building and mutual trust by:

- making and encouraging decisions in difficult situations.

If project leaders cannot fulfil these requirements, then the simultaneous engineering approach will be difficult to adopt.

Team behaviour also plays an essential role. Teamwork benefits product development and individual team members (see Section 4.3.1), however it can also give rise to the following problems [4.2]:

- groups or teams that work together for a long time tend to oversimplify
- control of team effectiveness can decline
- teams begin to conform, which can lead to the protection of competences and the overestimation of capabilities
- groups who have worked together successfully for a long time develop a self-confidence that is not always justified
- within a team one may find opinionated individuals who dominate others and who need careful management
- some team members may sit back and not pull their weight.

In addition to adopting an understanding leadership style, these problems can be addressed specifically by creating small teams, encouraging an open dialogue, and, if necessary, removing or adding team members. Ideally, teams should be dissolved when their project goals have been achieved.

Dörner and Badke-Schaub [4.2, 4.10] have written about the effectiveness of groups and teams in comparison with individuals. Although general statements are difficult to make, it appears that group opinions settle at a relatively high level. This means that results are never as good as those of the best individuals, but also never as bad as those of the worst individuals. An idea or the work of an individual can stand out from that of the team, but can also be significantly worse.

This implies that surprising proposals from individuals should not be suppressed. On the contrary, these should be developed to a point where a clear comparison with the team result is possible. In a team one cannot rely on, or even expect, valuable individual original contributions to arise, so opportunities should be created to encourage them. Team building does not automatically guarantee good solutions. Company culture and leadership style remain fundamental for effective teamwork and successful individual work.

5 Task Clarification

5.1 Importance of Task Clarification

The design task is generally presented to the design and development department in one of the following forms:

- as a development order (from outside or from the product planning department in the form of a product proposal)
- as a definite order
- as a request based on, for instance, suggestions and criticism by sales, research, test or assembly staff, or originating in the design department itself.

The task description contains not only statements about the product, such as its functionality and performance, but also statements about deadlines and cost targets. The design and development department now faces the problem of identifying the requirements that determine the solution and embodiment and formulating and documenting these quantitatively as far as possible. In order to achieve this, the following questions need to be answered in close cooperation with the client or proposer:

- What are the objectives that the intended solution is expected to satisfy?
- What properties must it have?
- What properties must it not have?

The result or this process is a *requirements list*. This document thus represents the specification against which the success of the design project can be judged.

In so far as this has not already been done in product planning (see Section 3.1), the design and development department should undertake the situation analysis described in 3.1.4 in order to specify the product situation and to identify future developments.

A useful method used to support the preparation of the requirements list is *Quality Function Deployment* (QFD) (see Section 10.5). QFD helps to translate customer wishes into product requirements.

5.2 Setting Up a Requirements List (Design Specification)

The main working steps required to set up a requirements list are shown in Figure 5.1. The procedure involves two stages. In the first stage the obvious requirements are defined and recorded. In the second stage these requirements are refined and extended using special methods.

The following sections describe the contents and format of a requirements list, along with the individual working steps.

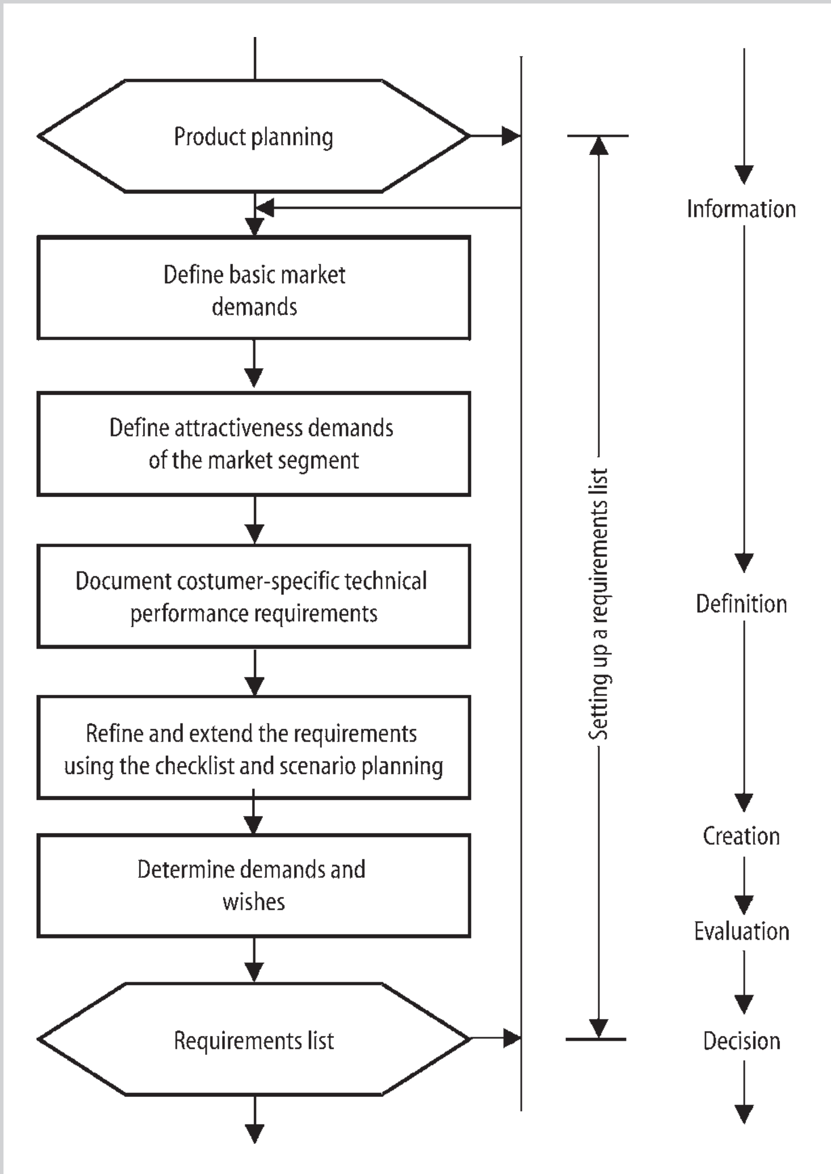


Figure 5.1. Main working steps required to set up a requirements list

5.2.1 Contents

When preparing a detailed requirements list it is essential to clearly elaborate the goals and the circumstances under which they have to be met. The resulting requirements must be identified either as demands or wishes.

Demands are requirements that must be met under all circumstances; in other words, if any of these requirements are not fulfilled the solution is unacceptable (for instance such qualitative demands as "suitable for tropical conditions", "splashproof", etc.). Minimum demands must be formulated as such (for example $P > 20$ kW; $L < 400$ mm).

Wishes are requirements that should be taken into consideration whenever possible, perhaps with the stipulation that they only warrant limited increases in cost, for example, central locking, less maintenance, etc. It is advisable to classify wishes as being of major, medium or minor importance [5.4].

The distinction between demands and wishes is also important at the evaluation stage, since selection (see Section 3.3.1) depends on the fulfilment of demands, while evaluation (see Section 3.3.2) is only performed on variants that already meet the demands.

Even before a certain solution is adopted, a list of demands and wishes should be set up and the quantitative and qualitative aspects tabulated. Only then will the resulting information be adequate:

- Quantity: All data involving numbers and magnitudes, such as number of items required, maximum weight, power output, throughput, volume flow rate, etc.
- Quality: All data involving permissible variations or special requirements, such as waterproof, corrosionproof, shockproof, etc.

Requirements should, if possible, be quantified and, in any case, defined in the clearest possible terms. Special indications of important influences, intentions or procedures may also be included in the requirements list, which is thus an internal digest of all the demands and wishes expressed in the language of the various departments involved in the design process. As a result, the requirements list not only reflects the initial position but, since it is continually reviewed, also serves as an up-to-date working document. In addition, it is a record that can, if necessary, be presented to the Management Board and the sales department so that they may make their objections known before the actual work is started.

5.2.2 Format

The requirements list should at least contain the following information in a structured format (see also Figure 5.2).

- user: company or department
- project or product name
- requirements labelled as demands or wishes
- person responsible for each requirement
- date of issue for the overall requirements list
- date of last change
- version number and/or index number
- page number.

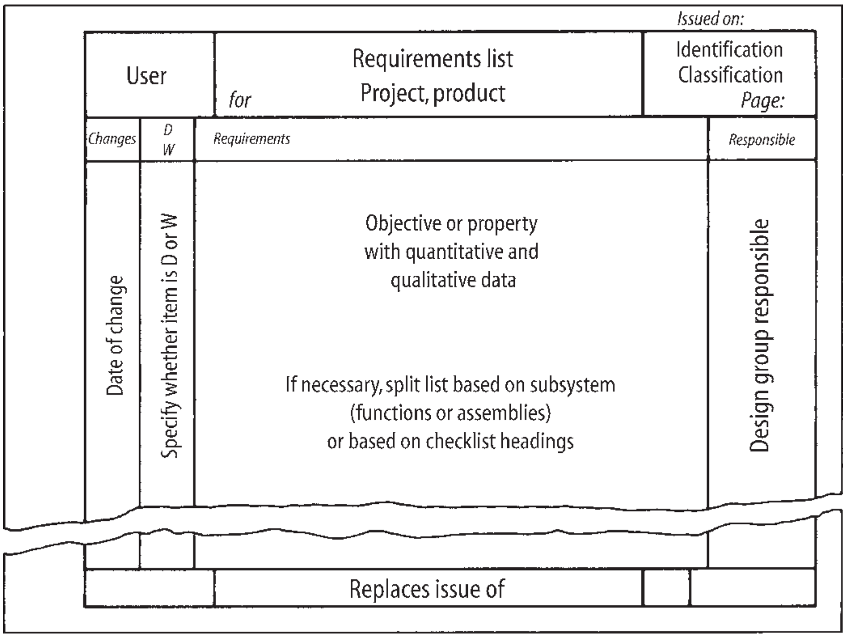


Figure 5.2. Layout of a requirements list

It is helpful if the format of the requirements list becomes a company standard so that it can be used, elaborated and adopted by as many departments as possible. Figure 5.2 is thus no more than a suggestion that can, of course, be modified at will.

It may prove useful to set up the requirements list based on subsystems (functions or assemblies) where these can be identified, or else based on checklist headings (see Figure 5.3). With established solutions, where the assemblies to be developed or improved are already determined, the requirements list must be arranged in accordance with these: special design groups are usually put in charge of the development of each assembly. With motor cars, for instance, the requirements list can be subdivided into engine, transmission and bodywork development.

It is extremely useful to *record the source* of demands and wishes. It is then possible to go back to the proposers of requirements and to enquire into their actual motives. This is particularly important when the question arises of whether or not the demands can be changed in the light of subsequent developments.

Any *changes in*, and *additions to*, the original task that may result from a better understanding of solution possibilities or from possible changes in emphasis must always be entered into the requirements list, which will then reflect the progress of the project at any particular time.

Responsibility for this work is placed on the chief designer. The updated requirements list should be circulated among all departments concerned with the development of the product (management, sales, accounts, research, etc.). The requirements list should only be changed or extended due to a decision of those in charge of the overall project and by following a formal change management procedure.

Main headings	Examples
Geometry	Size, height, breadth, length, diameter, space requirement, number, arrangement, connection, extension
Kinematics	Type of motion, direction of motion, velocity, acceleration
Forces	Direction of force, magnitude of force, frequency, weight, load, deformation, stiffness, elasticity, inertia forces, resonance
Energy	Output, efficiency, loss, friction, ventilation, state, pressure, temperature, heating, cooling, supply, storage, capacity, conversion.
Material	Flow and transport of materials. Physical and chemical properties of the initial and final product, auxiliary materials, prescribed materials (food regulations etc)
Signals	Inputs and outputs, form, display, control equipment.
Safety	Direct safety systems, operational and environmental safety.
Ergonomics	Man-machine relationship, type of operation, operating height, clarity of layout, sitting comfort, lighting, shape compatibility.
Production	Factory limitations, maximum possible dimensions, preferred production methods, means of production, achievable quality and tolerances, wastage.
Quality control	Possibilities of testing and measuring, application of special regulations and standards.
Assembly	Special regulations, installation, siting, foundations.
Transport	Limitations due to lifting gear, clearance, means of transport (height and weight), nature and conditions of despatch.
Operation	Quietness, wear, special uses, marketing area, destination (for example, sulphurous atmosphere, tropical conditions).
Maintenance	Servicing intervals (if any), inspection, exchange and repair, painting, cleaning.
Recycling	Reuse, reprocessing, waste disposal, storage
Costs	Maximum permissible manufacturing costs, cost of tooling, investment and depreciation.
Schedules	End date of development, project planning and control, delivery date

Figure 5.3. Checklist for setting up a requirements list

5.2.3 Identifying the Requirements

As a rule, the first requirements list undertaken is the most difficult to set up. Experience greatly facilitates the compilation of subsequent ones. After a relatively short period, several examples become available that can be used as the starting point for subsequent requirements lists.

The main issue associated with setting up a requirements list is the quantity and quality of the documents and data that are supplied with the design task. Depending on the branch of engineering, not all the expected product properties will have been defined and documented. The rest of them are expected by the customers but not made explicit, i.e. they are implicit requirements. The following questions therefore need to be answered:

- What is the problem really about?
- Which implicit wishes and expectations are involved?
- Do the specified constraints actually exist?
- What paths are open for development?

It is therefore important for the design and development department to understand the customers or the market segment involved. The basis of the requirements list is often a contract that has been signed with a customer. This contract usually includes the agreed product properties and performance data, product liability regulations, and the guidelines that have to be applied.

In a first exploratory step, the statements and requirements in the contract are translated into product-relevant parameters that designers and engineers can apply. This is relatively straightforward to do because the product specification in a contract involves explicit requirements. A bigger problem is how to deal with the implicit requirements; although they are not expressed they still have a very negative impact if they are not fulfilled. What effects, for example, do statements such as "simple maintenance" have on the embodiment of the product, and how can such statements be formulated as specific requirements? How difficult it is to formulate a requirements list depends on the type of customer and, in principle, two types can be distinguished:

- *anonymous customers:* these include a particular market segment, those identified by the sales department without a customer order, and those identified by the product planning department.
- *specific customers:* these not only include individual customers who place an order, but also market segments that are served by many companies with similar products in which requirements have become standardised, e.g. those for "compact cars" and "family cars". Although the actual customers in such cases are anonymous, they can, in effect, be treated as specific customers.

According to Kramer [5.3], some specific types of requirement can be formulated for each type of customer.

Basic requirements are always implicit requirements, i.e. they are not articulated by the customer. Their fulfilment is self-evident and vital for the customer. Success or failure of a product is determined by these requirements. For example, for a follow-on product the customer generally expects energy consumption and operating costs to be reduced. It is essential for the design and development department to recognise the importance of these implicit requirements. The sales department or product management must supply information about these requirements, along with the thoughts and expectations of the customers.

Technical performance requirements are explicit requirements. They are articulated by the customer and can usually be specified precisely. For example, a new engine may have to have 15 kW of power and weigh not more than 40 kg. Such concrete values are used by customers when comparing competing products. The importance of the individual parameters is determined by the customers themselves.

Attractiveness requirements are again implicit requirements. Customers are usually not aware of these; however, they are used to differentiate between competing products. In general customers are not willing to pay higher prices for these additional product properties. Consider an example from a motor car where the number of standard colours and the available combinations of external and internal colour schemes are such requirements.

5.2.4 Refining and Extending the Requirements

Two methods have been developed to refine and extend the requirements list defined thus far:

- follow a checklist
- create scenarios.

The checklist shown in Figure 5.3 is a generic one based on ideas described in Section 2.1.7. The items in this list are checked against the existing task and its requirements in order to obtain further requirements. A further checklist can be found in Ehrlenspiel's book [5.1].

When creating scenarios, the product life from production to disposal is considered and sketched out. For every stage, a scenario is developed and the following questions asked:

- What might happen to the product? Examples: What kind of state might it find itself in? How might it be treated and used? Who might use it or come into contact with it? Where might it be used?
- How should the product react? Examples: What level of tolerance to failure should be built in? How should dangerous situations be avoided?

The answers to these questions are used to formulate further product requirements. Most of these requirements will not be very specific, i.e. they cannot be translated into the product parameters that determine solutions or embodiments. For example, the previously mentioned statement "simple maintenance" needs to be specified in more detail. Kramer [5.3] proposes the following three-step procedure to achieve this.

First step (statement)

- Customer's need: simple maintenance.

Second step (development)

- Customer's requirements:
  1. Provide long maintenance interval

2. Enable simple maintenance
3. Make maintenance procedures easy to learn.

Third step (refinement)

- Provide long maintenance interval:
  1. Maintenance interval at least 5 000 operating hours
  2. Grease cams every 10 000 operating hours.
- Enable simple maintenance:
  1. Fit maintenance access covers with manual locks
  2. Fit cams with lubricating points that fit standard grease guns
  3. Leave space for oil drip tray
  4. Provide locating features to assist when refitting access covers.
- Make maintenance procedures easy to learn:
  1. Add a separate section in the operating manual describing maintenance procedures
  2. Provide labels indicating the locks that need to be undone for maintenance
  3. Indicate the directions of the maintenance operations with etched arrows.

The results of the third step are then added to the requirements list.

When clarifying the task, one should start by collecting the essential functions and the existing task-specific constraints with respect to the energy, material and signal transformations. When all of the information is available, it must be grouped, ordered and labelled.

In 5.2.1 we pointed out the essential differences between demands and wishes. In many cases it is clear from the outset whether a requirement is a demand or a wish. However, a definitive assignment is required before the requirements list is released. If necessary, further information should be collected. Wishes should be formulated such that their weighting can be established. Initially it is often useful to express such weightings qualitatively rather than quantitatively, because the estimates often change as the understanding of the task develops.

5.2.5 Compiling the Requirements List

In the light of arguments advanced in this chapter, the following general method of compiling a requirements list can now be recommended:

1. Identify the requirements:
   - Check the customer contract or the sales documents for technical requirements and define and document them.
   - Refer to the items of the checklist (Figure 5.3) and determine the quantitative and qualitative data.
   - Create scenarios that consider all stages in the product's life and thus derive further requirements.

- Refine by asking:
  - What objectives must the solution satisfy?
  - What properties must it have?
  - What properties must it not have?
- Collect further information.
- Specify demands and wishes clearly.
- If possible, rank wishes as being of major, medium or minor importance.

2. Arrange the requirements in a clear order:
   - Define the main objective and the main characteristics.
   - Split into identifiable subsystems, functions, assemblies, etc., or in accordance with the main headings of the checklist.
3. Enter the requirements list on standard forms and circulate among interested departments, licensees, directors, etc.
4. Examine objections and amendments and, if necessary, incorporate them into the requirements list.

Once the task has been adequately clarified, and the relevant departments are satisfied that the listed requirements are technically and economically attainable, the way is clear for the conceptual design phase.

5.2.6 Examples

Figure 5.4 shows a requirements list for a printed circuit board positioning machine, illustrating the main characteristics of the content and the format of requirements lists. It has been structured according to the main characteristics given in Figure 5.3. The requirements have been split into demands and wishes, and, where possible and necessary, quantified. Modifications and amendments with their dates are also shown. The latter were the result of an intensive discussion of a first draft of the requirements list (first version 21st April 1988).

Requirements lists based on the above recommendations are provided in Figures 6.4, 6.27 and 6.43 as further examples.

5.3 Using Requirements Lists

5.3.1 Updating

In principle, requirements lists should be binding and complete. However, initially a requirements list is always provisional because, as the design process proceeds, it grows and changes. Any attempt to formulate all possible requirements at the start of a project will fail and would cause considerable delays. Looking at the inputs

Siemens		Requirements list for a printed circuit board positioning machine	Issued on 27/04/88 Page: 1
Changes	D W	Requirements	Responsible
		1. Geometry: dimensions of the test sample	Langner's group
		Circuit board:	
	D	Length = 80 – 650 mm	
	D	Breadth = 50 – 570 mm	
	W	Height = 0.1 – 10 mm	
	D	Required hight = 1.6 – 2 mm	
	W	Clearance between basic grid boards ≤ 120 mm	
	D	'Clamping area' ≤ 2 mm (edges of the board)	
		2. Kinematics:	
27/04/88	D	Precise positioning of the test sample	
27/04/88	D	Minimum of 2 mm displacement of the test sample normal to the board	
27/04/88	D	Feedback to transfer position	
	W	separate stations for input and output	
	D	Design of clearance zone	
	W	Minimum handling time (as fast as possible)	
		3. Forces:	
	D	Weight of the test sample ≤ 1.7 kg	
27/04/88	W	Maximum weight of the test sample ≤ 2.5 kg	
		4. Energy:	
	D	Electrical and /or pneumatic (6 – 8 bar)	
		5. Material:	
	D	Free from rust	
	D	Isolation between test sample and testing device	
27/04/88	W	Thermal expansion of testing device adjusted to expansion of printed circuit	
27/04/88	D	Consideration of influence of temperature	
27/04/88	D	Temperature range: 15–40 °C	
27/04/88	D	Humidity: 65 %	
27/04/88	W	Circuit boards: epoxy-fiberglass sheet	
27/04/88	D	No condensation	
		6. Safety:	
27/04/88	D	Operator Safety	
		7. Production:	
		Consideration of tolerance build up	
		8. Operation:	
	D	No contamination inside the testing device	
	D	Destination: production line	
		9. Maintenance:	
	W	Maintenance intervall > 10^6 test operations	
		10. Schedule:	
	D	Embodiment finished by July 1988	
		Replaces issue of 21/04/88	

Figure 5.4. Requirements list for a printed circuit board positioning machine (Siemens AG)

and outputs of individual working steps in the design process, the reasons become clear. For example, in the final stages of designing a motor car, the thicknesses of all of the individual layers of body paint need to be known. However, to develop the concept, these data are not relevant. The paintwork requirements therefore do not have to be specified until much later in the process.

Thus, working with binding yet provisional requirements lists takes into account the fact that not all of the data and requirements are known or have to be known at the beginning of a design process. Only those requirements that are absolutely necessary in order to be able to proceed to the next working step need to be documented. At the start of a project, it is important to specify those parameters and properties that:

- define the particular concept
- influence the product structure
- determine the overall embodiment of the product.

The contents of a requirements list therefore depend on the state of the product design and the stage of the design process. The list has to be continuously amended and extended. Managing requirements lists in this way avoids having to deal with questions and requirements before they can be adequately answered and specified.

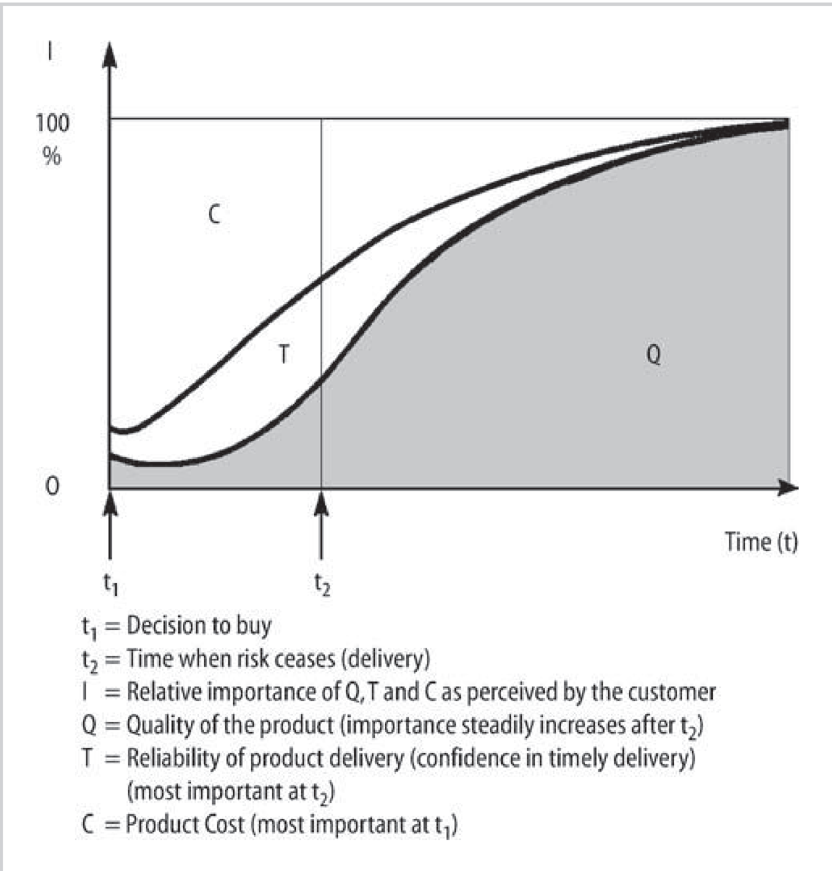


Figure 5.5. Changing appreciation of product quality by the customer

Product requirements frequently change with time, during both product creation and product use.

During Product Creation, customers often change their demands and wishes. This happens when customers gain new knowledge and understanding and when the planned application has been extended. This is typical of capital equipment because of its long development process. For railway rolling stock, for example, it is possible that during the development process the rail network has been extended with the consequence that specified powers and capacities are no longer sufficient.

During product use, customer appreciation of the product can change and, as a consequence, the requirements and their relative importance can also change. For example, the longer a product is in service, the more important quality issues such as long maintenance intervals and reliability become, see Figure 5.5 [5.3].

The fact that requirements can change must be considered when setting up a requirements list. A mutually satisfactory requirements management process is therefore essential in order to ensure good and lasting relations between a company and its customers.

5.3.2 Partial Requirements Lists

It is often beneficial for specialist areas or departments of a company to prepare so-called "partial requirements lists" documenting only their particular requirements. This removes the need for the design and development department to spend time collecting more information and data than are strictly relevant to them, see Figure 5.6 [5.2].

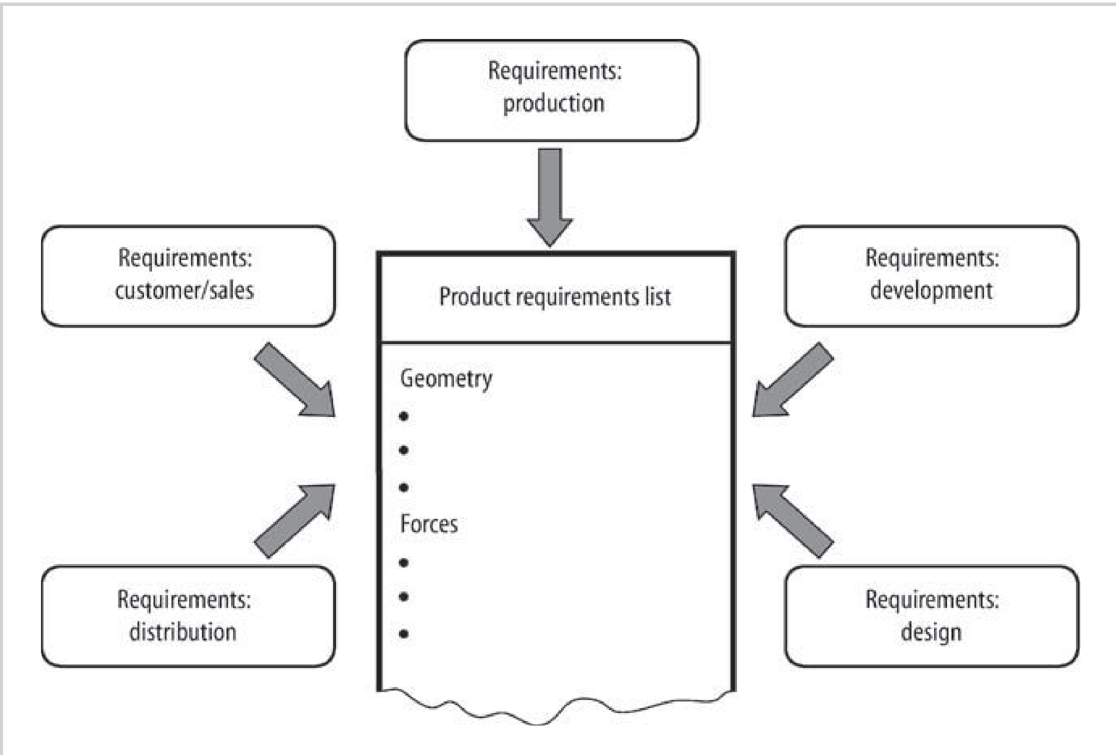


Figure 5.6. Product requirements list compiled from partial requirements lists

The product requirements list is a compilation of all of the partial requirements lists. An important role of project management is to ensure that the partial requirements lists cover all areas and are compatible with each other. Modern engineering data management systems [5.5] support their efficient administration and editing.

5.3.3 Further Uses

Even when the design is not original, and the solution principle and layout are fixed so that nothing more than *adaptations* or *dimensional changes* need to be made in a familiar area, orders should nevertheless be executed on the basis of requirements lists, which can then take the form of templates or questionnaires. These should be constructed in such a way that information for electronic data processing and quality control can be read off directly. As a result, requirements lists become sources of information for direct action.

Beyond that, requirements lists, once compiled, are an invaluable *source of information* about the required or desired properties of the product, and hence extremely helpful for further developments, negotiations with suppliers, etc. Setting up requirements lists for existing products can also provide a very valuable source of information for the subsequent development and rationalisation of those products.

The examination of a requirements list during project meetings or before assessing various designs is an extremely useful procedure. All of those involved are placed in possession of all of the available information and all salient evaluation criteria are brought home to them.

Requirements lists are an important basis for knowledge management systems. Stored in such systems, requirements lists provide a very valuable source of knowledge about previous projects that can often be reused.

5.4 Practical Application of Requirements Lists

In the last few years it has been shown that, at least for original designs, the formulation of a requirements list is a very efficient method for solution development and has been broadly adopted by industry. When used in practice, however, the following issues often arise:

- *Obvious requirements*, such as low-cost production, ease of assembly, are often not included in the requirements list. One should take care that these issues are both addressed and expressed precisely.
- In an early stage of the project it is not always possible to make *precise statements* in the requirements list. The statements have to be amended or corrected during the design and development process.
- A *stepwise development* of the requirements list is very useful when tasks are poorly defined. In these cases the requirements should be formulated more precisely as soon as possible.

- During the formulation of requirements lists or related discussions, *functions* or *solution ideas* are often mentioned. This is not wrong. They can encourage a clearer formulation of the requirements and even lead to the identification of new requirements. The solution ideas or proposals generated should be recorded so that they can be used later in the systematic search for solutions. However, they should not enter—and possibly bias—the requirements list.
- The identification of *deficiencies* and *failures* can initiate requirements that must then be formulated in a solution-neutral way. Failure analysis is often the starting point for a requirements list.
- For *adaptive* or *variant designs*, designers should still make requirements lists for themselves, even when the task is small.
- Setting up requirements lists should not be formalised too strictly. *Guidelines* and *forms* are only a *means* to prevent important issues from being forgotten and to provide a supporting structure. If one deviates from the recommendations in this book, one should a least consider the main characteristics and distinguish between demands and wishes.

6 Conceptual Design

Conceptual design is the part of the design process where—by identifying the essential problems through abstraction, establishing function structures, searching for appropriate working principles and combining these into a working structure—the basic solution path is laid down through the elaboration of a solution principle. Conceptual design *specifies the principle solution*.

From Figure 4.3 we can see that the conceptual phase is preceded by a decision. The purpose of this decision is to answer the following questions based on the requirements list agreed upon during task clarification:

- Has the task been clarified sufficiently to allow the development of a solution in the form of a design?
- Is a conceptual elaboration really needed, or do known solutions permit direct progress to the embodiment and detail design phases?
- If the conceptual stage is indispensable, how and to what extent should it be developed systematically?

6.1 Steps of Conceptual Design

According to the procedural plan outlined in Section 4.2, the conceptual design phase follows the clarification of the task. Figure 6.1 shows the steps involved, correlated in such a way as to satisfy the principles of the general problem solving process set out in Section 4.1.

The reasons for the individual steps have been examined in Section 4.2 and need not be discussed further here. It should, however, be mentioned that refinements of any one of the steps by reiteration on a higher information level should be made whenever necessary. The loops involved have been omitted from Figure 6.1 for the sake of greater clarity.

The individual steps and the appropriate working methods for the conceptual design phase will now be examined in detail.

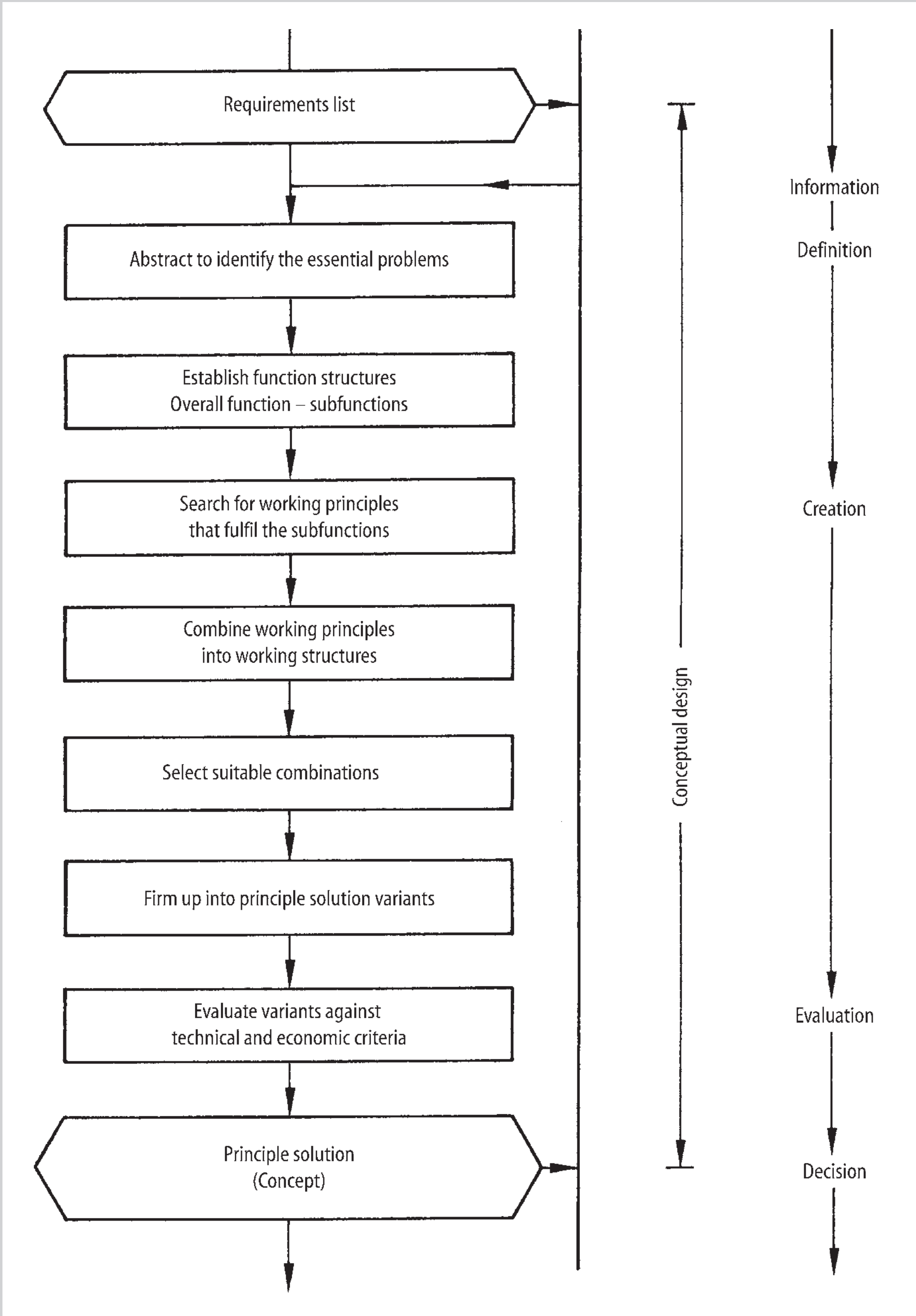


Figure 6.1. Steps of conceptual design

6.2 Abstracting to Identify the Essential Problems

6.2.1 Aim of Abstraction

Solution principles or designs based on traditional methods are unlikely to provide optimum answers when new technologies, procedures, materials, and also new scientific discoveries, possibly in new combinations, hold the key to better solutions.

Every industry and every design office is a store of experiences as well as of prejudices and conventions which, coupled to the wish to minimise risks, stand in the way of better and more economic but unconventional solutions. The client, customer or product planning group might have included specific proposals for a solution in the requirements list. It is also possible that during the discussion of individual requirements, ideas and suggestions for realising a solution will emerge. In the unconscious, at least, certain solutions might exist. Perhaps concrete ideas already exist, however these could be based on fixed ideas and fictitious constraints.

In their search for optimum solutions, designers, far from allowing themselves to be influenced by fixed or conventional ideas, must therefore examine very carefully whether novel and more suitable paths are open to them. In order to solve the problem of fixation and sticking with conventional ideas, *abstraction* is used. This means ignoring what is particular or incidental and emphasising what is general and essential. Such generalisation leads straight to the *crux of the task*. If it is properly formulated, then the overall function and the essential constraints become clear without prejudicing the choice of a particular solution in any way.

As an example, consider the improvement of a labyrinth seal in a high-speed turbine in accordance with a set of requirements. The task is described in detail by means of a requirements list and the formulation of the goal to be achieved. In the abstracting approach, the crux of the task would not so much be the design of a labyrinth seal as that of a shaft seal without physical contact, with due regard being paid to certain operating and spatial constraints, and also to cost limits and delivery times. Specifically, the designer should ask whether the crux is:

- to improve the technical functions, e.g. the sealing quality or safety
- to reduce weight or space
- to significantly lower costs
- to significantly shorten delivery times
- to improve production methods.

All of these questions might have to be satisfied by the overall solution, but their importance may differ from case to case. Nevertheless, due regard must be paid to each of them, since any one of them is likely to provide the impetus for the discovery of a new and better solution principle. New developments involving a proven solution principle, coupled to modifications in production methods, are often imposed by the need to lower costs and shorten delivery times.

Thus, if an improvement in the sealing properties were the crucial requirement in the example we have mentioned, new sealing systems would have to be found.

This would mean studying the flow of fluids in narrow passages and, from the knowledge acquired, providing for better sealing properties, while also satisfying the other subproblems we have mentioned.

If, on the other hand, cost reduction were the crucial point then, after an analysis of the cost structure, one would have to see whether the same physical effects could be produced through the use of cheaper materials, by reducing the number of components or by using a different production process. It is also possible to search for new concepts to achieve a better or at least similar sealing performance for lower cost.

It is the identification of the crux of the task with the functional connections and the task-specific constraints that throws up the essential problems for which solutions have to be found. Once the crux of the task has been clarified, it becomes much easier to formulate the overall task in terms of the essential subproblems as they emerge [6.2, 6.6, 6.13].

6.2.2 Broadening the Problem Formulation

This is the best point in the process to bring in those designers who are actually going to be responsible for the project. Having identified the crux of the task by correct problem formulation, a step-by-step enquiry is now initiated to discover if an extension of, or even a change in, the original task might lead to promising solutions.

An excellent illustration of this procedure has been given by Krick [6.5]. The task he used as an example was an improved method of filling, storing and loading bags of animal feed. An analysis gave the situation shown in Figure 6.2. It would have been a grave mistake to begin immediately by thinking of possible improvements to the existing situation. By proceeding in this way one is likely to ignore other, more useful and more economic solutions. Using abstraction and the systematic extension of what is already known about the task, the following problem formulations are possible, each representing a higher level of abstraction than the last:

1. Filling, weighing, closing and stacking bags of feed.
2. Transferring feed from the mixing silo to stacked bags in the warehouse.
3. Transferring feed from the mixing silo to bags on the delivery truck.
4. Transferring feed from the mixing silo to the delivery truck.
5. Transferring feed from the mixing silo to a delivery system.
6. Transferring feed from the mixing silo to the consumer's storage bins.
7. Transferring feed from ingredient containers to consumer's storage bins.
8. Transferring feed ingredients from their source to the consumer.

Krick has incorporated some of these formulations into a diagram (see Figure 6.3).

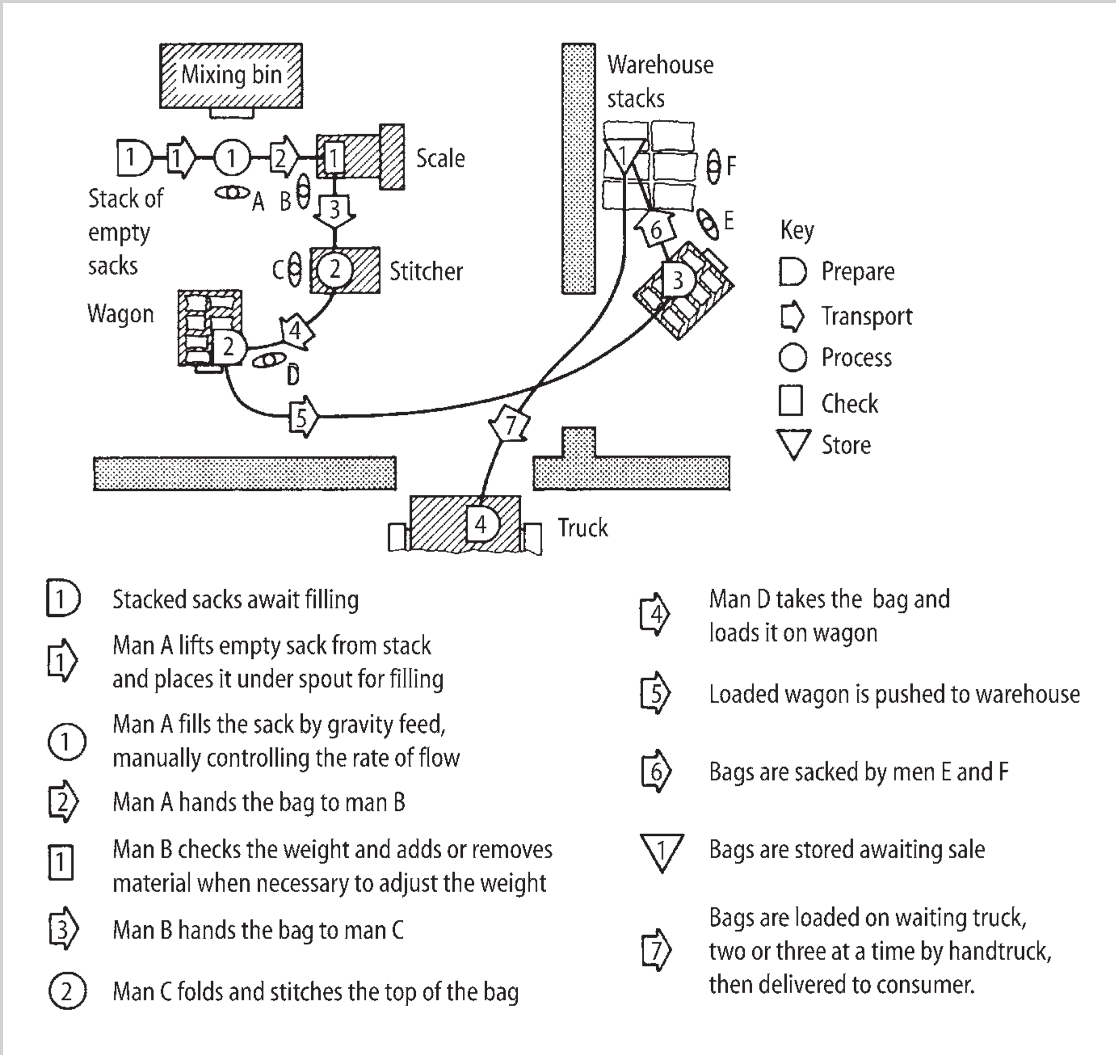


Figure 6.2. The present method of filling, storing and loading bags of feed. After [6.5]

It is characteristic of this approach that the problem formulation is made as broad as possible in successive steps. In other words, the current or obvious formulation is not accepted at face value but *broadened systematically*. Although this may conflict with decisions already taken, it opens up new perspectives. Thus, formulation 8 above is the broadest, the most general and the least circumscribed.

The crux of the task, in fact, is the transport of the correct quantity and quality of feed from the producer to the consumer and not, for instance, the best method of closing or stacking bags, or moving them into the warehouse. With a broader formulation, solutions may appear that render the filling of bags and storing them in the warehouse unnecessary.

How far this process of abstraction is continued depends on the constraints. In the case under consideration, Formulation 8 must be rejected on technical, seasonal and meteorological grounds: the consumption of feed is not confined to harvest time; for various reasons consumers will not want to store feed for a whole year; moreover, they may be reluctant to mix the required ingredients themselves. However, the transfer of feed on demand, for instance, with delivery trucks taking

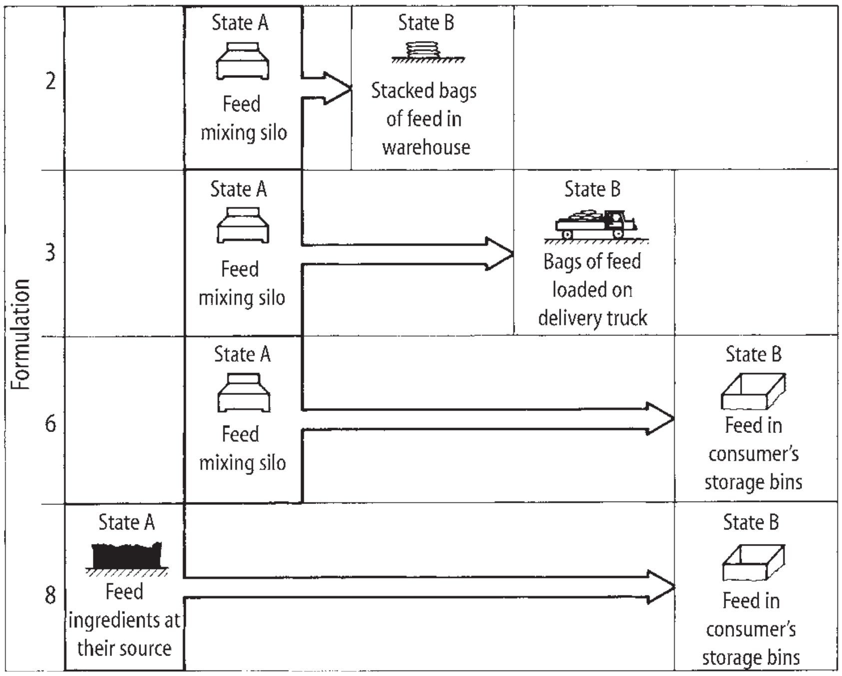


Figure 6.3. Alternative formulations of the feed distribution problem, illustrating progressively broader formulations of a problem. After [6.5]. A = initial state; B = final state

it directly from the mixing silo to consumers' storage bins (Formulation 6), is more economical than intermediate storage in a warehouse and the transport of smaller quantities in bags. In this connection, the reader might recall a development in a different field which culminated in the delivery of ready-mixed concrete direct to building sites in special vehicles.

We have tried to show how comprehensive problem formulation on an abstract plane opens the door to better solutions. This approach, furthermore, helps to raise the influence and responsibility of designers by giving them an overview of the problem and thus involving them in, for instance, environmental protection and recycling. It is useful to analyse the requirements list as set out in the next section.

6.2.3 Identifying the Essential Problems from the Requirements List

The clarification of the task with the help of a requirements list will have helped to focus attention on the problems involved and will have greatly increased the particular level of information (see Chapter 5). Elaborating the requirements list may thus be said to have prepared the way for following steps.

Here the task is to *analyse the requirements list* with respect to the required function and essential constraints in order to confirm and refine the crux of the problem. Roth [6.11] advises that the functional relationships contained in the requirements list should be formulated explicitly and arranged in order of their importance.

That analysis, coupled to the following step-by-step abstraction, will reveal the general aspects and essential problems of the task, as follows:

Step 1. Eliminate personal preferences.

Step 2. Omit requirements that have no direct bearing on the function and the essential constraints.

Step 3. Transform quantitative into qualitative data and reduce them to essential statements.

Step 4. As far as it is purposeful, generalise the results of the previous step.

Step 5. Formulate the problem in solution-neutral terms.

Depending on either the nature of the task or the size of the requirements list (or both), certain steps may be omitted.

Table 6.1 illustrates abstraction based on these steps using the requirements list for a motor vehicle fuel gauge shown in Figure 6.4. The general formulation makes it clear that, with respect to the functional relationships, the problem is the measurement of quantities of liquid, and that this is subject to the essential conditions that the quantity of liquid is changing continuously and that the liquid is in containers of unspecified size and shape.

This analysis thus leads to a definition of the objective on an abstract plane without laying down any particular solution.

In principle, all paths must be left open until such time as it becomes clear which solution principle is the best. Thus designers must question all the constraints they are given and work out with the client or proposer whether or not they should be retained as genuine restrictions. In addition, designers must learn to discard fictitious constraints that they themselves have come to accept, and to that end ask critical questions and test all their presuppositions. Abstraction helps to identify fictitious constraints and to eliminate all but genuine restrictions.

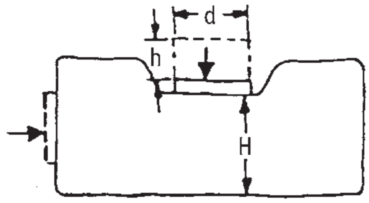
We shall conclude this section with a few examples of purposeful abstraction and problem formulation:

- Do not design a garage door, but look for means of securing a garage in such a way that a car is protected from thieves and the weather.
- Do not design a keyed shaft, but look for the best way of connecting a gear wheel and shaft.
- Do not design a packing machine, but look for the best way of despatching a product safely or, if specific constraints really exist, of packing a product safely, compactly and automatically.
- Do not design a clamping device, but look for a means of keeping a workpiece firmly fixed.

From the above formulations, and this is very helpful for the next step, the final formulation can be derived in a way that does not prejudice the solution, i.e. is *solution-neutral*, and at the same time turns it into a *function*:

- "Seal shaft without contact", not "Design a labyrinth seal".
- "Measure quantity of fluid continuously", not "Gauge height of liquid with a float".
- "Measure out feed", not "Weigh feed in sacks".

TH Darmstadt	Requirements list for fuel gauge	3rd issue 10/07/85 Page 1

Changes	D W	Requirements	Responsible
		• <u>Container</u>	
			
		<u>Geometry</u> H = 100 mm – 600 mm Volume: 20 – 160 litres	
	W	2 – 630 litres	
		Shape fixed but unspecified (rigid)	
	W	Container flexible or only partially rigid	
		<u>Material:</u> steel, plastic <u>Connection to container</u> Bayonet socket, clamped connections, top or side:	
	W	d = ∅71 mm, h = 20 mm	
		Tank not pressurised (ventilated) Pressure test for container 0.3 bar	
		<u>Contents, temperature range</u>	

Liquid	Operating Range °C	Storage enviroment °C
Petrol	– 25 to + 65	
Diesel	– 24 to + 65	– 40 to + 100
Engine oil (W)	up to + 140	– 40 to + 100

Changes	D W	Requirements	Responsible
	W	• <u>Display</u> System with electric input signal – Moving magnet instrument (catalogue) – Bimetallic instrument (catalogue) – Board computer	
		Availabel source of energy: DC at 12 V, 24 V Voltage variation – 10% to + 25% of nominal voltage Current consumption max. 300 mA	

	Replaces 2nd issue of 27/06/1973	

Figure 6.4. Requirements list: motor vehicle fuel gauge

TH Darmstadt	Requirements list for fuel gauge		3rd issue 10/07/85 Page 2
Changes	D W	Requirements	Responsible
		• System to be developed	
		Geometry Consider connection constraints to container	
	W	Kinematics No moving parts	
		Energy (see display)	
		Material (see container)	
		Signal	
		∘ Input Minimum measurable content: 3 % of maximum value Reserve tank contents by special signal	
	W	Signal unaffected by angle of liquid surface Possibility of signal calibration	
	W	Possiblity of signal calibration with full container	
		∘ Output	
	W	Output of transmitter: electric signal Output signal accuracy at max. value $\pm 3\%$	
	W	$\pm 2\%$ (together with indicator error $\pm 5\%$) Under normal conditions, horizontal level, $v = \text{const.}$ Able to withstand shocks of normal driving Response sensitivity: 1 % of maximum output signal	
	W	0.5 % of maximum output signal	
		∘ Connection between input and output Distance container – display: $\neq$ zero m; 3 m–4 m	
	W	1 m–20 m Separate power possible	
		Production large-scale production	
	Replaces 2nd issue of 27/06/1973		

Figure 6.4. (continued)

TH Darmstadt	Requirements list for fuel gauge	3rd issue 10/07/85 Page 3

Changes	D W	Requirements	Responsible
		Test requirements	
		Operating conditions of vehicle	
		Forward acceleration $\pm$ 10 m/s^2	
		Sideways acceleration $\pm$ 10 m/s^2	
		Upward acceleration (vibration), up to 30 m/s^2	
	W	Shocks in forward direction without damage, up to 30 m/s^2	
		Forward tilt up to $\pm 30°$	
		Sideways tilt max 45°	
		Salt spray tests for inside and outside components according to client's requirements (DIN 90905 to be considered)	
		Must conform with heavy vehicle regulations	
		Operation, Maintenance	
	W	Installation by non-specialist	
		Life expectancy 10^4 level changes (full/empty)	
		Minimum of 5 years service life	
		Fuel gauge replaceable	
		Fuel gauge maintenance-free	
	W	Fuel gauge simply modified to suite different container sizes	
		Regulations	
		No regulations relating to explosion safety	
		Quantity	
		10000/day of adjustable type	
		5000/day of the most popular type	
		Costs	
		Manufacturing costs $\leq$ DM 6.00 each (without display)	

	Replaces 2nd issue of 27/06/1973	

Figure 6.4. (continued)

Table 6.1. Procedure during abstraction: motor vehicle fuel gauge based on requirements list given in Figure 6.4

Result of Steps 1 and 2

- Volumes: 20 to 160 litres
- Shape of container: fixed or unspecified (rigid)
- Top or side connection
- Height of container: 100 mm to 600 mm
- Distance between container and indicator: $\neq 0$ m, 3 m to 4 m
- Petrol and diesel, temperature range: $-25\,^{\circ}C$ to $65\,^{\circ}C$
- Output of transmitter: unspecified signal
- External energy: DC at 12 V, 24 V. Variation -15% to $+25\%$
- Output signal accuracy at maximum $\pm 3\%$ (together with indicator error $\pm 5\%$)
- Response sensitivity: 1% of maximum signal output
- Possibility of signal calibration
- Minimum measurable content: 3% of maximum value

Result of Step 3

- Various volumes
- Various container shapes
- Various connections
- Various contents (liquid levels)
- Distance between container and indicator: $\neq 0$ m
- Quantity of liquid varies with time
- Unspecified signal
- (with outside energy)

Result of Step 4

- Various volumes
- Various container shapes
- Transmission over various distances
- Measure (continuously changing) quantities of liquid
- (with outside energy)

Result of Step 5 (Problem formulation)

- Measure continuously changing quantities of liquid in containers of unspecified size and shape and indicate the measurements at various distances from the containers.

6.3 Establishing Function Structures

6.3.1 Overall Function

According to Section 2.1.3, the requirements determine the function that represents the intended overall relationship between the inputs and the outputs of a plant, machine or assembly. In Section 6.2 we explained that problem formulation obtained by abstraction does much the same. Hence, once the crux of the overall problem has been formulated, it is possible to indicate an *overall function* that, based on the *flow of energy, material and signals* can, with the use of a *block diagram*, express the solution-neutral relationship between *inputs* and *outputs*. That relationship must be specified as precisely as possible (see Figure 2.3).

In our example of a fuel gauge (see Figure 6.4), quantities of liquid are introduced into and removed from a container, and the problem is to measure and indicate the quantity of liquid found in the container at any one time. The result, in the liquid system, is a flow of material with the function "store liquid" and, in the measuring system, a flow of signals with the function "measure and indicate quantity of liquid". The second is the overall function of the specific task under consideration, that is, the development of a fuel gauge (see Figure 6.5). That overall function can be broken down into subfunctions in a further step.

6.3.2 Breaking a Function Down into Subfunctions

Depending on the complexity of the problem, the resulting overall function will in turn be more or less complex. By complexity we mean that the transparency of the relationships between inputs and outputs is relatively poor, that the required physical processes are relatively intricate, and that the number of assemblies and components involved is relatively large.

Just as a technical system can be divided into subsystems and elements (see Section 2.1.3), so a complex or *overall function* can be broken down into *subfunctions* of lower complexity. The combination of individual subfunctions results in a *function structure* representing the overall function.

The aims of breaking down complex functions are to:

- determine subfunctions that facilitate the subsequent search for solutions
- combine these subfunctions into a simple and unambiguous function structure.

Let us return to the example of the fuel gauge (see Sections 6.2.3 and 6.3.1). The starting point is the problem formulation for the overall function given in Figure 6.5.

The flow of signals has been treated as the main flow. Associated subfunctions are developed in several steps. As a first step, the contents of the container have to be measured and the resulting signal received. This signal has to be channelled and

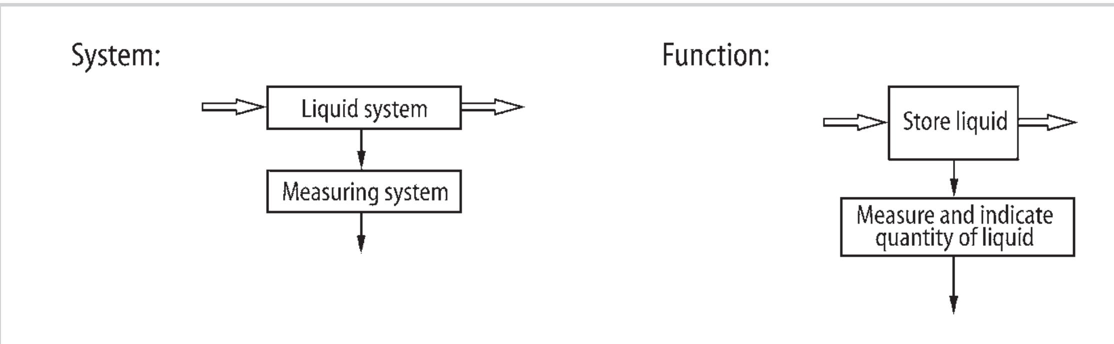


Figure 6.5. Overall functions of the systems involved in measuring the contents of a container. After Figure 6.4 and Table 6.1

finally displayed to the driver to indicate the contents of the container. Thus, three important direct main functions have been identified. Possibly the signal needs to be changed before it can be channelled. Figure 6.6 shows the development and variation of a function structure in accordance with the suggestions set out in this section.

Since the requirements list also provides for measurements in containers of different sizes holding varying initial quantities of liquid, an adjustment of the signal to the respective size of the container is expedient, and is accordingly introduced as an auxiliary function. Measurements in containers of various unspecified shapes will, in certain circumstances, demand the correction of the signal as another auxiliary function. The measuring operation may require a supply of external energy, which must then be introduced as a further flow. Finally, consider the system boundary. If existing indicating instruments are to be used, the device will have to emit an electric output signal. If they are not, then the subfunctions "channel signal" and "indicate signal" must be included in the search for solutions. In this way, a function structure with suitable subfunctions can be developed. The individual subfunctions are of a lower complexity than the overall function and, furthermore, it will become clear which subfunction provides the most useful starting point for the search for solutions.

In our example, this important solution-determining subfunction, that has the working principle upon which the others clearly depend, is "receive signal" (see Figure 6.6). The initial search for solutions should therefore focus on this subfunction. The solution selected for this will largely decide to what extent individual subfunctions can be changed round or omitted. It also allows for better judgement of whether to use existing channelling and display solutions or whether to seek a new solution for these subfunctions, i.e. an extension of the system boundary.

Further recommendations for identifying and formulating subfunctions are now described.

It is useful to start by determining the *main flow* in a technical system, if this is clear. The *auxiliary flows* should only be considered later. When a basic function structure, including the most important links, has been found, it is easier to undertake the next step; that is, to consider the auxiliary flows with their subfunctions and to achieve a further subdivision of complex subfunctions. For this step it is helpful to create a temporary working structure or a solution for the *basic function structure*, without, however, prejudicing the final solution.

The optimum method of breaking down an overall function—that is, the number of subfunction levels and also the number of subfunctions per level—is determined by the relative novelty of the problem and also by the method used to search for solutions. In the case of *original designs*, neither the individual subfunctions nor their relationships are generally known. In that case, the search for and establishment of an optimum function structure constitute some of the most important steps of the conceptual design phase. In the case of *adaptive designs*, on the other hand, the general structure with its assemblies and components is much more well-known, so that a function structure can be obtained by analysing the existing product. Depending on the special demands of the requirements list, that function

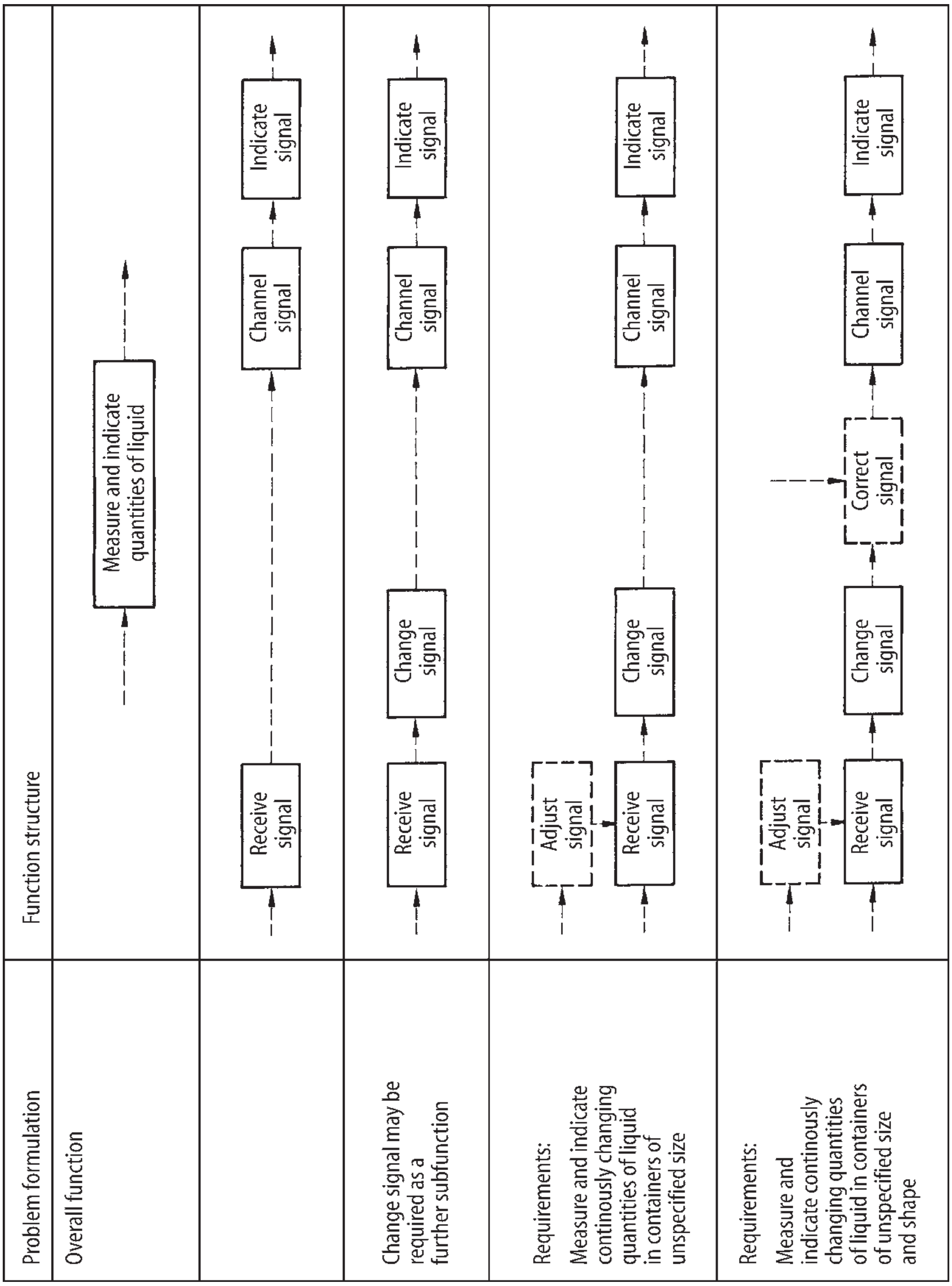


Figure 6.6. Development of a function structure for a fuel gauge

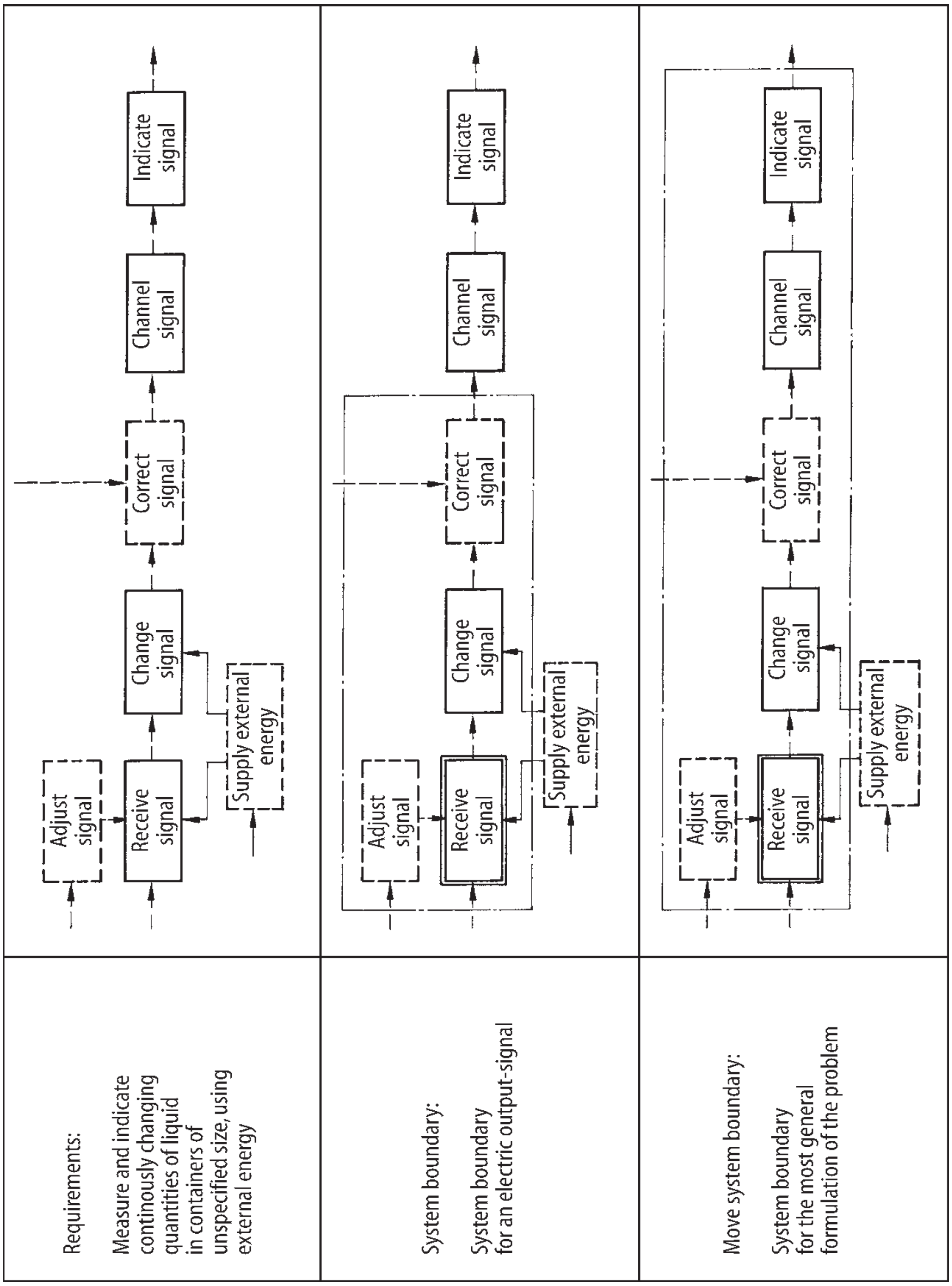


Figure 6.6. (continued)

structure can be modified by the variation, addition or omission of individual subfunctions or by changing the way that they are combined.

Function structures are of great importance in the development of modular systems. For this type of *variant design*, the physical structure—that is, the assemblies and individual components used as building blocks and also their interfaces—must be reflected in the function structure (see also Section 9.2.1).

A further advantage of setting up a function structure is that it permits the *clear definition* of existing subsystems or of those to be newly developed, so that they can be *dealt with separately*. If existing assemblies can be assigned directly as complex subfunctions, the subdivision of the function structure can be discontinued at a fairly high level of complexity. In the case of new assemblies or those requiring further development, however, the division into subfunctions of decreasing complexity must be continued until the search for solutions seems promising. By adapting function structures to the novelty of the task or the subsystem, the use of function structures can save a great deal of time and money.

Apart from aiding in the search for solutions, function structures or their subfunctions can also be used for purposes of classification. Examples are the "classifying criteria" of classification schemes (see Section 3.2.3) and the subdivision of design catalogues.

It may prove expedient not only to set up task-specific functions, but also to elaborate the function structure from *generally valid subfunctions* (see Figure 2.7). The latter recur in technical systems, and may be helpful when searching for a solution since they may lead to the discovery of task-specific subfunctions or because design catalogues may list solutions for them. Defining generally valid functions can also be of use when varying function structures, for example to optimise the energy, material and signal flows. The following list and examples should be helpful in this regard.

Conversion of energy:

- Changing energy (e.g. electrical into mechanical energy)
- Varying energy components (e.g. amplifying torque)
- Connecting energy with a signal (e.g. switching on electrical energy)
- Channelling energy (e.g. transferring power)
- Storing energy (e.g. storing kinetic energy)

Conversion of material:

- Changing matter (e.g. liquefying a gas)
- Varying material dimensions (e.g. rolling sheet metal)
- Connecting matter with energy (e.g. moving parts)
- Connecting matter with signal (e.g. cutting off steam)
- Connecting different types of materials (e.g. mixing or separating materials)
- Channelling material (e.g. mining coal)
- Storing material (e.g. keeping grain in a silo)

Conversion of signals:

- Changing signals (e.g. changing a mechanical into an electrical signal, or a continuous into an intermittent signal)
- Varying signal magnitudes (e.g. increasing a signal's amplitude)
- Connecting signals with energy (e.g. amplifying measurements)
- Connecting signals with matter (e.g. marking materials)
- Connecting signals with signals (e.g. comparing target values with actual values)
- Channelling signals (e.g. transferring data)
- Storing signals (e.g. in databases)

In many cases in industry it may not be expedient to build up a function structure from generally valid subfunctions, because they are, in fact, too general and thus do not provide a sufficiently concrete picture of the relationships to aid the subsequent search for solutions. In general, a clear picture only emerges after adding more task-specific details (see Section 6.3.3).

To illustrate the approach some expamples follow. Figures 6.7 and 6.8 show the function structure of a tensile testing machine with a relatively complex flow of energy, material and signals. In this type of overall function, the function structure is built up step-by-step from subfunctions, with attention initially focused on essential main functions. Thus, on a first functional level, only the subfunctions

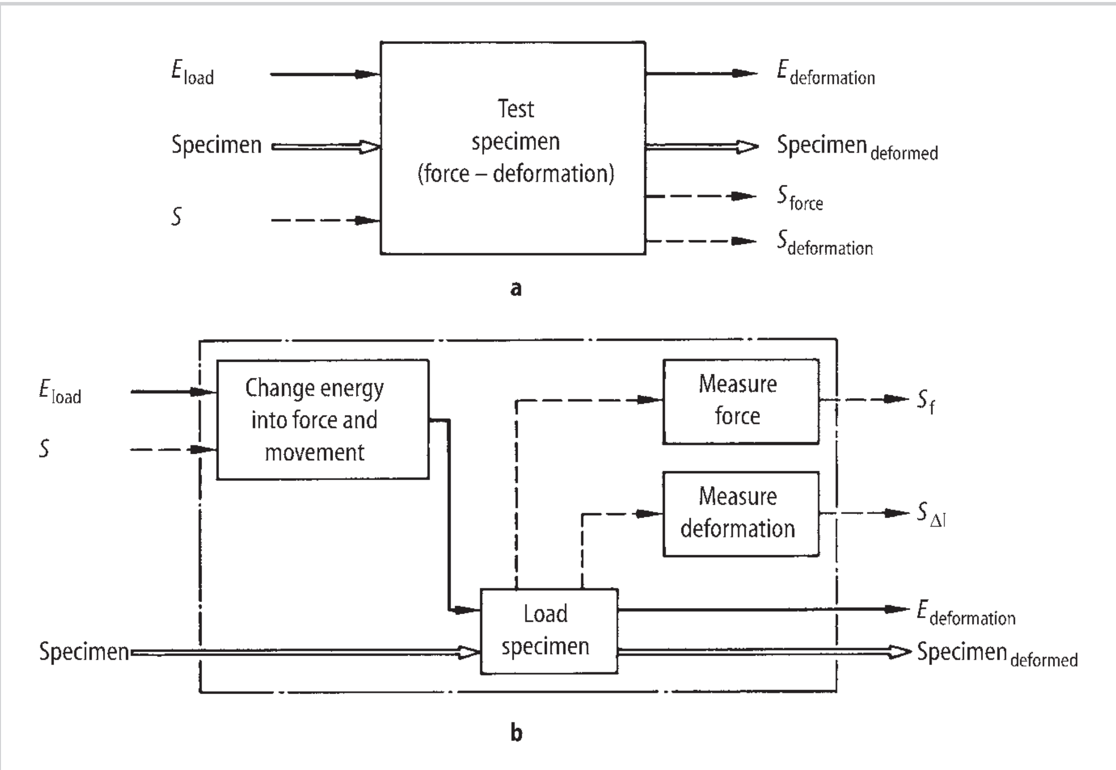


Figure 6.7. Overall function **a** and subfunctions (main functions) **b** of a tensile testing machine

that directly satisfy the required overall function are specified (see Figure 6.7). These are formulated as complex subfunctions, such as "change energy into force and movement" and "load specimen" in our example. Starting with complex subfunctions helps to establish a simple function structure.

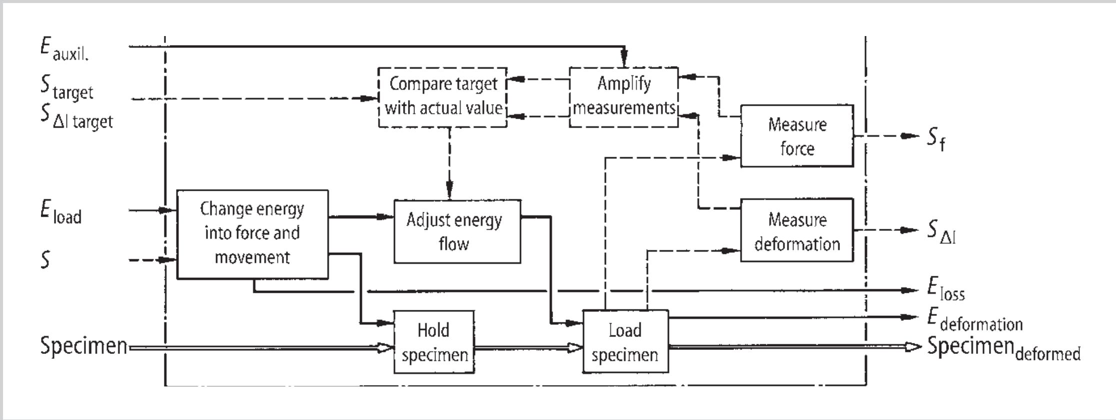


Figure 6.8. Completed function structure for the overall function set out in Figure 6.7

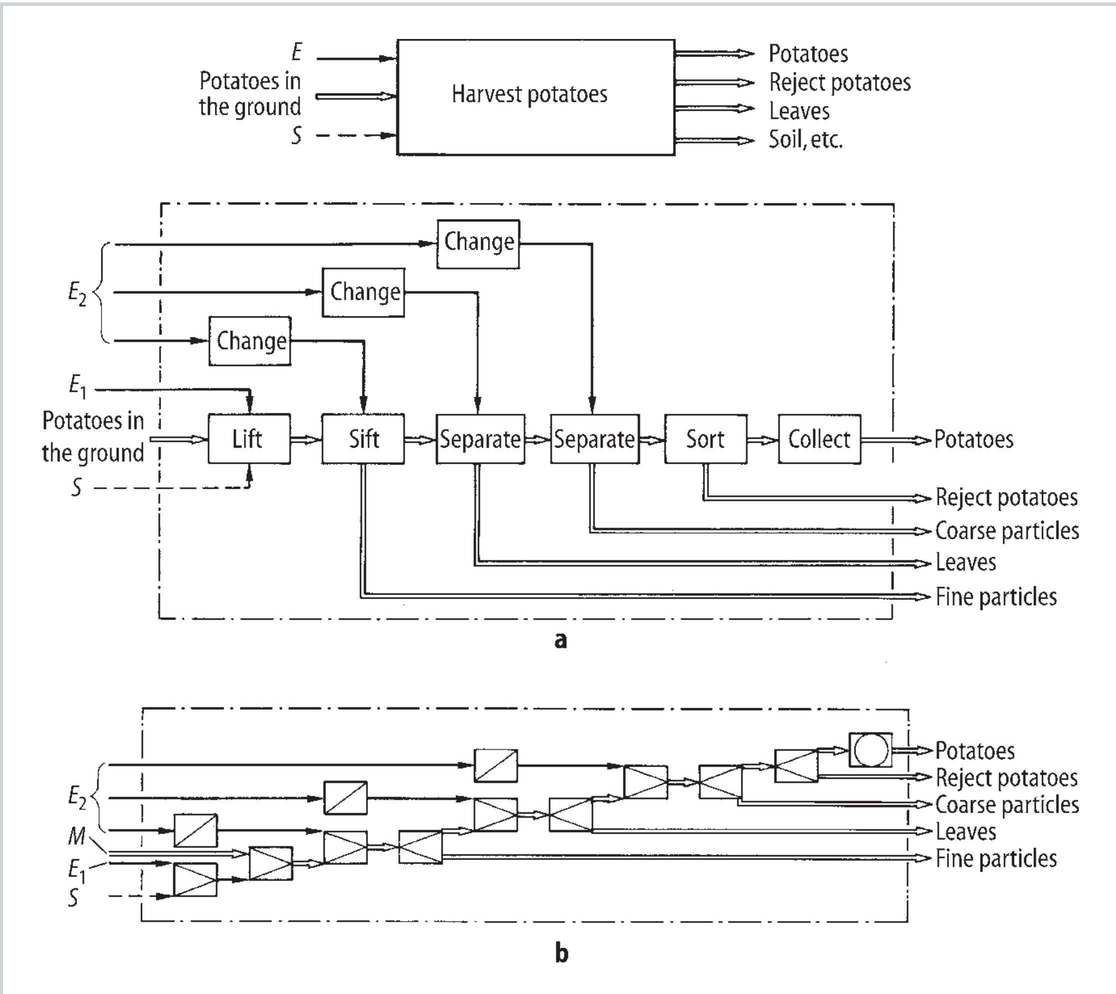


Figure 6.9. a Function structure of a potato harvesting machine **b** For comparison: diagram with generally valid functions based on [6.1], Figure 2.7

In the problem under consideration, the energy and signal flows are of roughly equivalent importance in the search for solutions, while the flow of material—the exchange of specimens—is only essential for the holding function added in Figure 6.8. In this figure, an adjusting function for the load magnitudes and, at the output of the system, the energy lost during the energy flow were also added because both clearly affect the design. The energy required to deform the specimen is lost with the material flow when the specimen is exchanged. Moreover, the auxiliary functions "amplify measurements" and "compare target with actual values" proved indispensable for the adjustment of the energy level.

There are, however, some problems in which variation of the main flow alone cannot lead to a solution, because *auxiliary flows* have a *crucial* bearing on the design and are solution-determining. As an example, let us consider the function structure of a potato harvesting machine. Figure 6.9a shows the overall function and the function structure based on the flow of material (the main flow) and the auxiliary flows of energy and signals. In Figure 6.9b, by comparison, the function structure is represented by means of generally valid functions, in order to emphasise the clear interrelationship of the different flows.

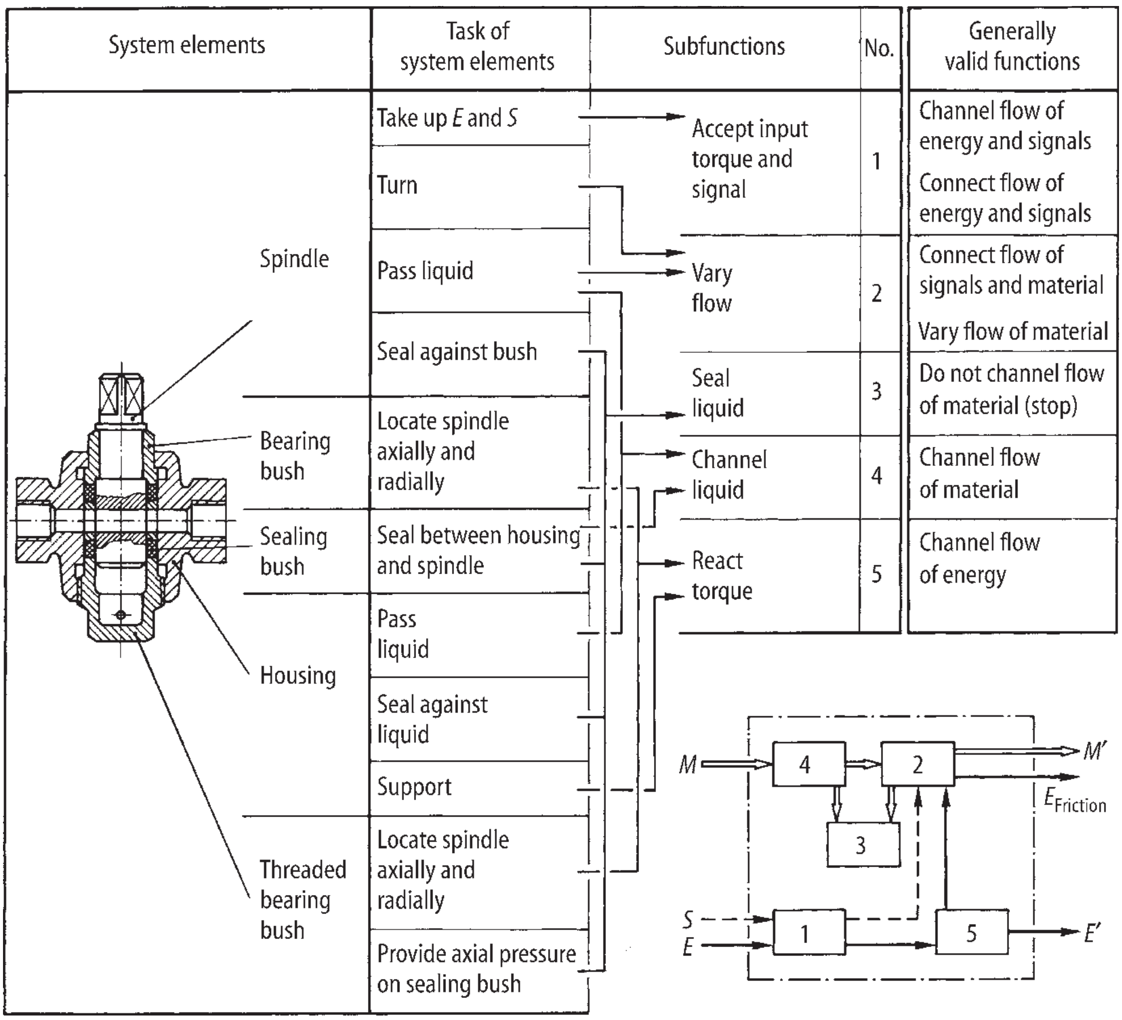


Figure 6.10. Analysis of a flow control valve with respect to its function structure

When generally valid functions are used, the separation into subfunctions is generally more pronounced than it is in the case of task-specific subfunctions. Thus, in the present example, the subfunction "separate" is replaced with the generally valid functions "connect energy with material mix" and "separate material mix" (the reverse of "connect"). The representation, however, is on such an abstract level that it is not easy to understand and requires further interpretation.

Our final example illustrates the derivation of function structures by the *analysis of existing systems*. This method is particularly suitable for developments in which at least one solution with the appropriate function structure is known, and the main problem is the discovery of better solutions. Figure 6.10 shows the steps used in the analysis of a flow control valve (a typical on–off switch), showing the individual tasks of the various elements and the subfunctions satisfied by the system. The function structure can be derived from the subfunctions and then varied in order to improve the product.

The function structure for the one-handed mixing tap examined in Section 6.6 clearly shows that the study of function structures may prove extremely useful, even after the physical effect has been selected, for determining the behaviour of the system at a very early stage of its development, and hence for identifying the structure that best suits the problem under consideration.

6.3.3 Practical Applications of Function Structures

When establishing function structures, we must distinguish between original and adaptive designs. In the case of *original designs*, the basis of a function structure is the *requirements list* and the *abstract formulation of the problem*. Among the demands and wishes, we are able to identify functional relationships, or at least the subfunctions at the inputs and outputs of a function structure. It is helpful to write out the functional relationships arising from the requirements list in the form of sentences and to arrange these in the order of their anticipated importance or in some other logical order [6.11].

In the case of *adaptive designs*, the starting point is the *function structure of the existing solution* obtained by analysing its elements. It helps to develop variants in order to open the path for other solutions, for subsequent optimisation and for the development of modular products. The identification of functional relationships can be facilitated by asking the right questions.

In modular systems, the function structure has a decisive influence on the modules and their arrangement (see Section 9.2). Here, the function structure and that of the assembly is affected not only by functional considerations, but also, and increasingly so, by production needs.

Function structures are intended to facilitate the discovery of solutions: they are not ends in themselves. The degree of detail used depends very much on the novelty of the task and the experience of the designers.

Moreover, it should be remembered that function structures are seldom completely free of physical or formal presuppositions, which means that the number of possible solutions is inevitably restricted to some extent. Hence, it is perfectly legitimate to conceive a preliminary solution and then abstract this by developing and completing the function structure by a process of iteration.

Anyone setting up a function structure ought to bear the following points in mind:

1. First derive a rough function structure with a few subfunctions from what functional relationships you can identify in the requirements list, and then break this rough structure down, step-by-step, by resolving complex subfunctions. This is much simpler than starting out with more complicated structures. In certain circumstances, it may be helpful to substitute a first solution idea for the rough structure and then, by analysing that first idea, to derive other important subfunctions. It is also possible to begin with subfunctions whose inputs and outputs cross the assumed system boundary. From these, we can then determine the inputs and outputs for the neighbouring functions; in other words, we work from the system boundary inwards.

2. If no clear relationship between the subfunctions can be identified, the search for a first solution principle may, under certain circumstances, be based on the mere *enumeration of identified subfunctions* without logical or physical relationships, but if possible these should be arranged according to the extent to which they have been realised.

3. *Logical relationships* may lead to function structures through which the logical elements of various working principles (mechanical, electrical, etc.) can be anticipated.

4. Function structures are not complete unless the existing or expected flows of energy, material and signals can be specified. Nevertheless, it is useful to begin by focusing attention on the *main flow* because, as a rule, it determines the design and is more easily derived from the requirements. The auxiliary flows then help with the further elaboration of the design, coping with faults, and when dealing with problems of power transmission, control, etc. The complete function structure, comprising all flows and their relationships, can be obtained by iteration; that is, by looking first for the structure of the main flow, completing that structure by taking the auxiliary flows into account, and then establishing the overall structure.

5. When setting up function structures it is useful to know that, in the conversion of energy, material and signals, several *subfunctions recur* in most structures and should therefore be introduced first. Essentially, these are the generally valid functions of Figure 2.7, and they can prove extremely helpful in the search for task-specific functions.

6. For the application of microelectronics, it is useful to consider signal flows as shown in Figure 6.11 [6.6]. This results in a function structure that clearly suggests the modular use of elements to detect (sensors), to activate (actuators), to operate (controllers), to indicate (displays) and, in particular, to process signals using microprocessors.

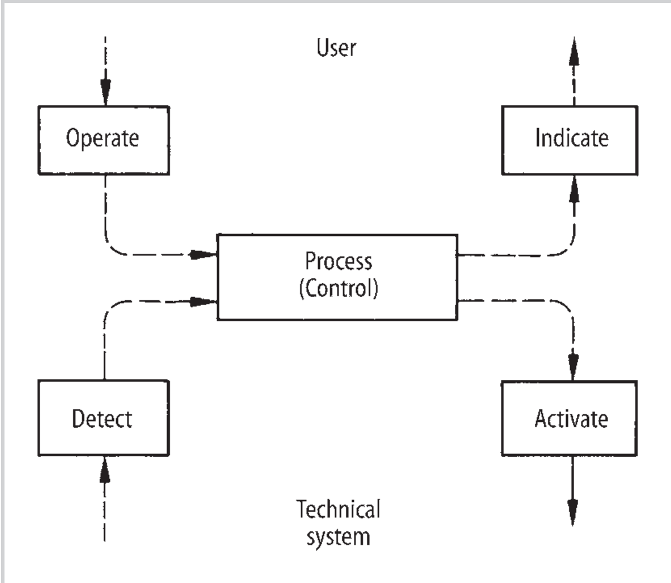


Figure 6.11. Basic signal flow functions for modular use in microelectronics. After [6.6]

7. From a rough structure, or from a function structure obtained by the analysis of known systems, it is possible to derive further *variants* and hence to optimise the solution, by:
   - breaking down or combining individual subfunctions
   - changing the arrangement of individual subfunctions
   - changing the type of switching used (series switching, parallel switching or bridge switching)
   - moving the system boundary.

   Because varying the function structure introduces distinct solutions, the setting up of function structures constitutes a first step in the search for solutions.
8. Function structures should be kept as *simple* as possible, in order to encourage simple and economical solutions. To this end, it is also advisable to aim at the combination of functions for the purpose of obtaining integrated function carriers. There are, however, some problems in which discrete functions must be assigned to discrete function carriers; for instance, when the requirements demand clarity in the solution, or when there is a need for extreme loading and quality. In this connection, the reader is referred to our discussion on the division of tasks (see Section 7.4.2).
9. In the search for solutions, only *promising function structures* should be introduced, which implies that a *selection procedure* (see Section 3.3.1) should be employed, even at this early stage.
10. For the *representation* of function structures it is best to use the *simple and informative symbols* shown in Figure 2.4, supplemented with task-specific verbal clarifications.

11. An *analysis* of the function structure leads to the identification of those subfunctions for which new working principles must be found, and of those for which known solutions can be used. This encourages an efficient approach. The search for solutions (see Section 3.2) then focuses on the subfunctions that are essential for the solution and on which the solutions of other subfunctions depend (see the example in Figure 6.6).

It is sometimes assumed wrongly that auxiliary functions are unimportant. Technical systems do not have functions that are "more important" or "less important". All functions are important because they are needed. Any functions that are not necessary or superfluous functions should be eliminated. It is only in order to reduce effort that designers start their search for solutions with the function that seems most important, i.e. solution-determining. All of the other functions are still necessary and must be fulfilled.

6.4 Developing Working Structures

6.4.1 Searching for Working Principles

Working principles need to be found for the various subfunctions, and these principles must eventually be combined into a working structure. The concretisation of the working structure will lead to the principle solution. A working principle must reflect the physical effect needed for the fulfilment of a given function and also its geometric and material characteristics (see Section 2.1.4). In many cases, however, it is not necessary to look for new physical effects, the form design (geometry and materials) being the sole problem. Moreover, in the search for a solution it is often difficult to make a clear mental distinction between the physical effect and the form design features. Designers therefore usually search for working principles that include the physical process along with the necessary geometric and material characteristics, and combine these into a working structure. Theoretical ideas about the nature and form of function carriers are usually presented by way of diagrams or freehand sketches.

It should be emphasised that the step we are now discussing is intended to lead to several solution variants, that is, a solution field. A solution field can be constructed by varying the physical effects and the form design features. Moreover, in order to satisfy a particular subfunction, several physical effects may be involved in one or several function carriers.

In Section 3.2 we discussed methods and tools for finding solutions. The same methods can be used in the search for working principles. Of particular importance, however, are literature searches, methods for analysing natural and known technical systems, and intuition-based methods (see Section 3.3.2). If preliminary solution ideas are available from product planning or through intuition, systematic analyses of physical processes and the utilization of classification schemes are also helpful (see Section 3.2.3). The last two methods usually provide several solutions.

Other important tools are design catalogues, in particular those proposed by Roth and Koller for physical effects and working principles (see Section 3.2.3)

[6.3, 6.11, 6.14]. When solutions need to be found for *several subfunctions*, it is expedient to select the functions as classifying criteria; that is, the subfunctions become the row headings and the possible working principles are entered in the columns. Figure 6.12 illustrates the structure of such a classification scheme, where the subfunctions are represented by F_i and the solution elements by S_{ij}. Depending on the level of concretisation, these solution elements can be physical effects or even working principles with geometric and material details.

As an example we consider the development of a cylinder–cylinder test rig in which two cylinders run against each other under a pulsating load. The aim was to investigate the friction characteristics for any combination of rolling and sliding speeds [6.9]. Figure 6.13 shows one possible function structure and Figure 6.14 the corresponding classification scheme. The main subfunctions identified are listed in the first column and potential solutions to those subfunctions are entered in the rows.

To sum up: the search for working principles for subfunctions should be based on the following guidelines:

- Preference should be given to the main subfunctions that determine the principle of the overall solution and for which no solution principle has yet been discovered.
- Classifying criteria and associated parameters (characteristics) should be derived from identifiable relationships between the energy, material and signal flows, or from associated systems.
- If the working principle is unknown, it should be derived from the physical effects and, for instance, from the type of energy. If the physical effect has been determined, appropriate form design features (working geometry, working motions and materials) should be chosen and varied. Checklists should be used to stimulate new ideas (see Figures 3.17 and 3.18).
- Designers should also enter solutions found intuitively and analyse which key classifying criteria influence particular working principles. These criteria should then be subdivided, limited or generalised using further headings.
- To prepare for the selection process, the important properties of the working principles should be noted.

Sub-functions \ Solutions		1	2	…	j	…	m
1	F_1	S_{11}	S_{12}		S_{1j}		S_{1m}
2	F_2	S_{21}	S_{22}		S_{2j}		S_{2m}
⋮	⋮	⋮	⋮		⋮		⋮
i	F_i	S_{i1}	S_{i2}		S_{ij}		S_{im}
⋮	⋮	⋮	⋮		⋮		⋮
n	F_n	S_{n1}	S_{n2}		S_{nj}		S_{nm}

Figure 6.12. Basic structure of a classification scheme with the subfunctions of an overall function and associated solutions

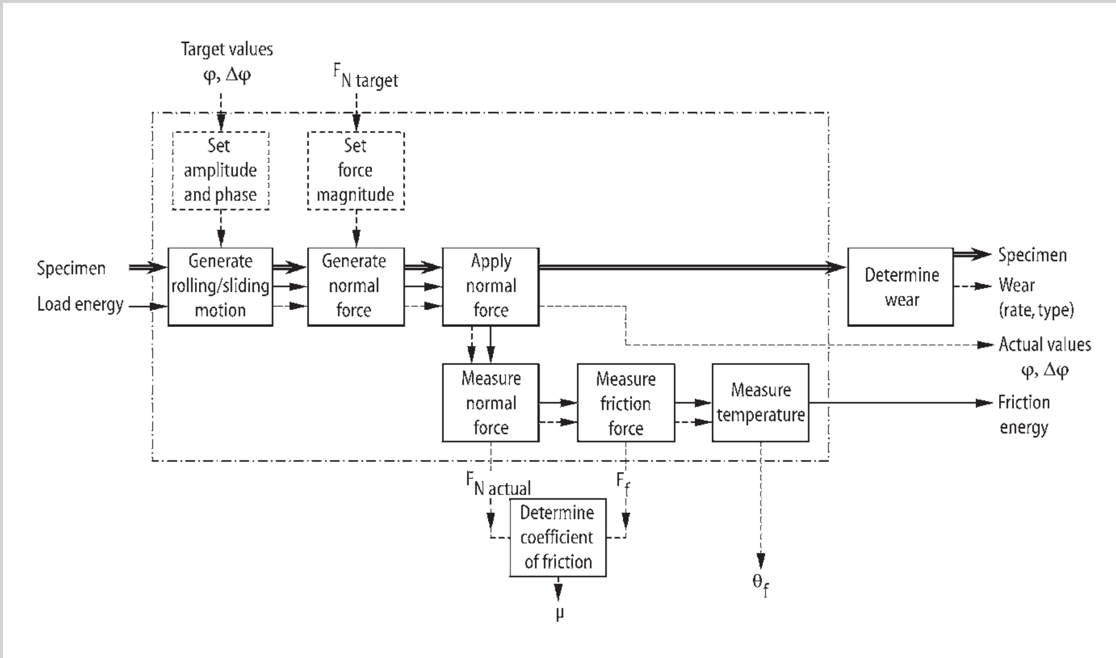


Figure 6.13. Possible function structure for a cylinder–cylinder test rig with a pulsating load for any combination of rolling and sliding motion

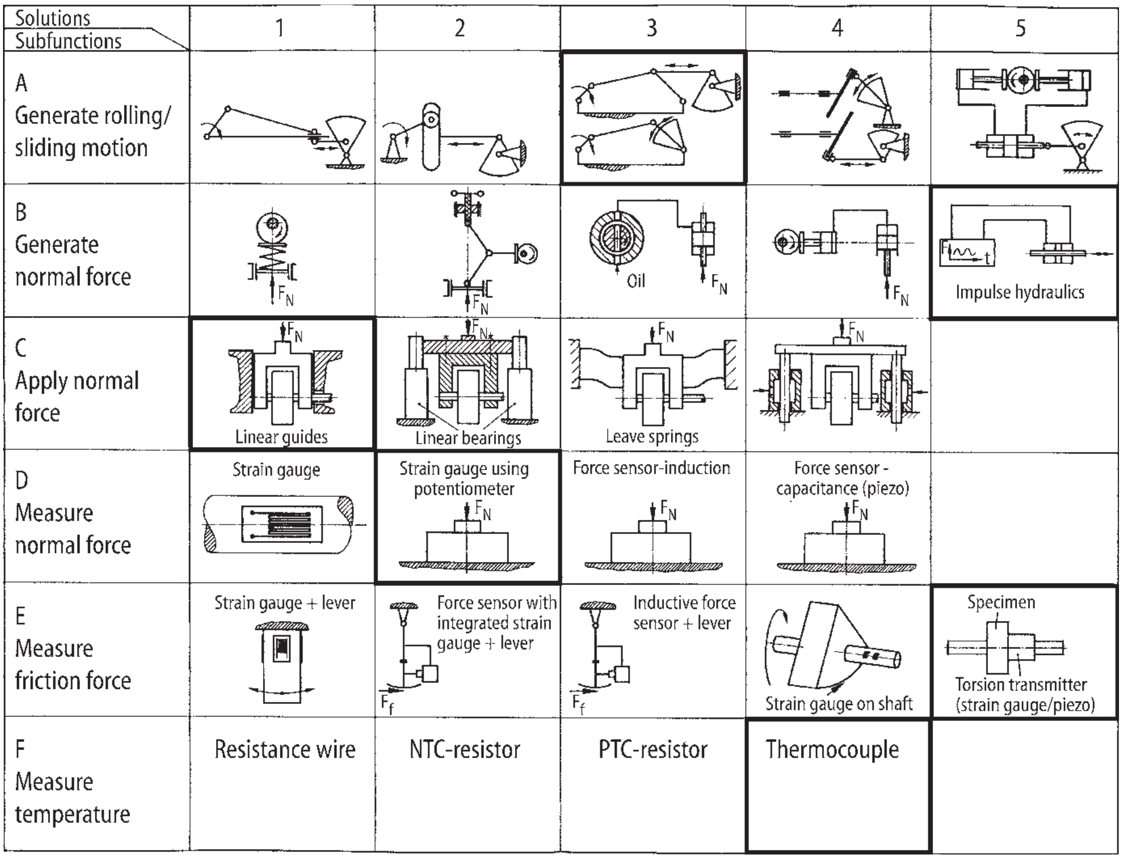


Figure 6.14. Classification scheme with possible solutions for the subfunctions identified in the function structure in Figure 6.13

Section 6.6 provides further examples that illustrate the search for working principles.

6.4.2 Combining Working Principles

To fulfil the overall function, it is then necessary to generate overall solutions by combining the working principles into a working structure, that is, system synthesis. The basis of such a combination is the established function structure, which reflects logically and physically possible or useful associations of the subfunctions.

In Section 3.2.4 the classification scheme of Zwicky (morphological matrix) was proposed as being particularly suitable for systematically combining solutions (see Figure 3.25). In this classification scheme, the subfunctions and the appropriate

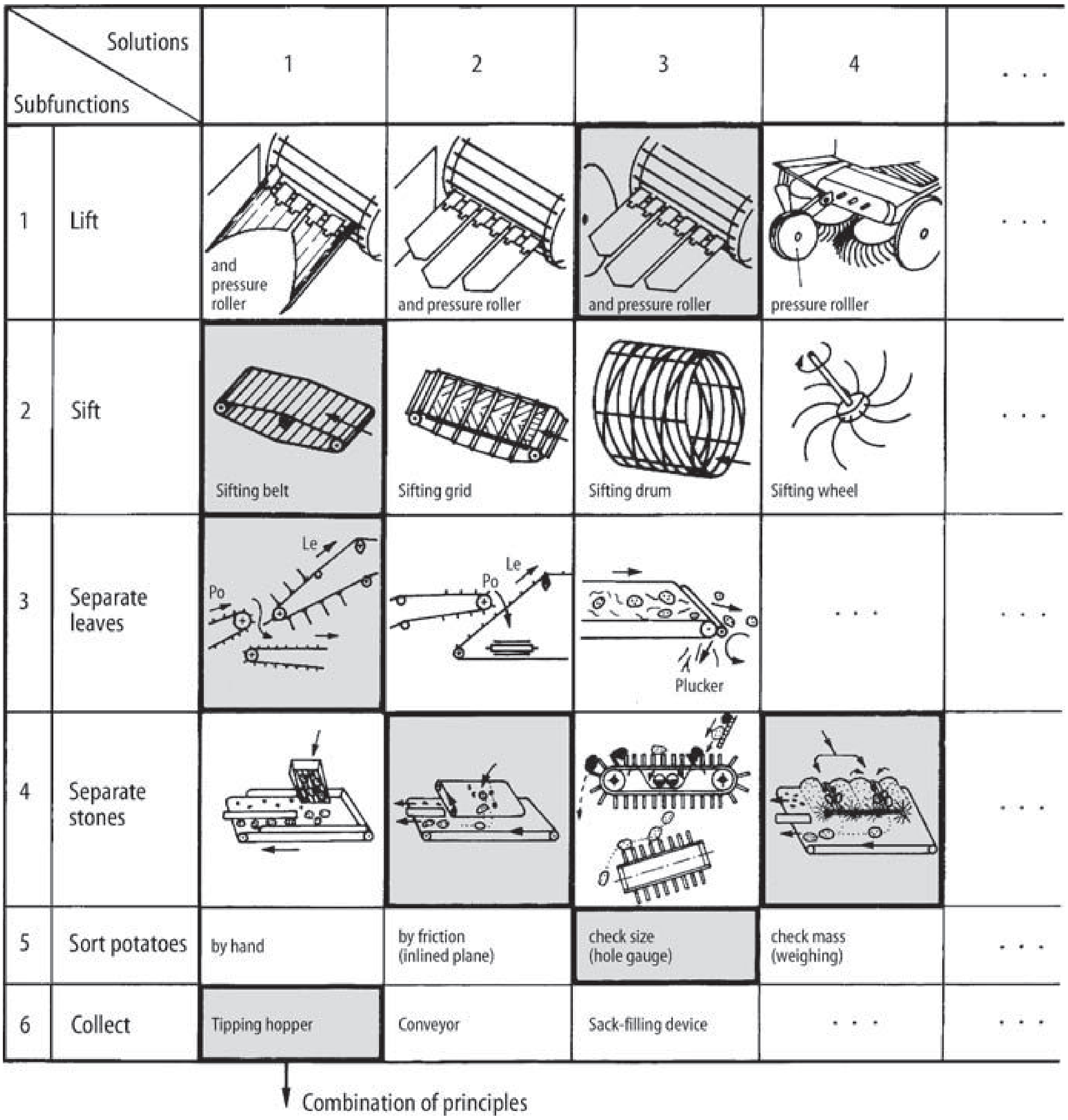


Figure 6.15. Combination of principles used to design a potato harvesting machine in accordance with the overall function structure shown in Figure 6.9. After [6.1]

solutions (working principles) are entered into the rows of the scheme. By systematically combining a working principle fulfilling a specific subfunction with the working principle for a neighbouring subfunction, one obtains an overall solution in the form of a possible working structure. In this process only those working principles that are compatible should be combined.

Figure 6.15 shows a possible combination of working principles for a potato harvesting machine [6.1]. It consists of working principles that are suitable for the subfunctions in the function structure shown in Figure 6.9. These have been made more concrete through rough sketches so that the assessment of their compatibility is facilitated. The principle solution of the harvesting machine based on this working structure is shown in Figure 6.16.

The main problem with combinatorial techniques is ensuring the physical and geometrical compatibility of the working principles to be combined, which in turn ensures the smooth flow of energy, material and signals. A further problem is the selection of technically and economically favourable combinations from the large field of theoretically possible combinations.

Combining solutions using mathematical methods (see Section 3.2.4) is only possible for working principles whose properties can be quantified. However, this is seldom possible at this early stage. Examples where it is possible are variant designs and control system designs, such as those using electronic or hydraulic components.

To sum up:

- Only combine compatible subfunctions (the compatibility matrix shown in Figure 3.26 is a useful tool).

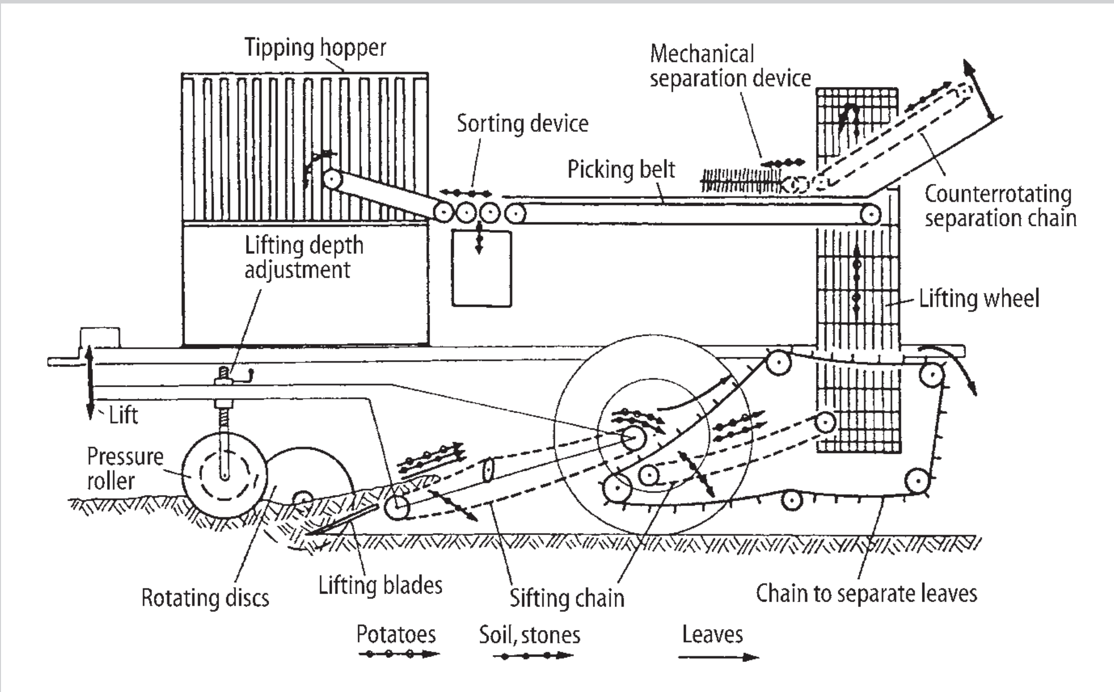


Figure 6.16. Principle solution of a potato harvesting machine, using a combination of principles from Figure 6.15

- Only pursue solutions that meet the demands of the requirements list and look like falling within the proposed budget (see the selection procedures in Sections 3.3.1 and 6.4.3).
- Concentrate on promising combinations and establish why these should be preferred above the rest.

6.4.3 Selecting Working Structures

Because working structures are generally not very concrete and the properties are only known qualitatively, the most suitable selection procedure is the one described in Section 3.3.1. This procedure is characterised by the activities of selecting and indicating preferences, and it makes use of a schematic selection chart that provides a clear overview and can be checked.

The solution field shown in Figure 6.14 for the cylinder–cylinder test rig is now evaluated for each subfunction's solution using a selection procedure. Figure 6.17 shows part of the selection chart indicating the most promising subfunction solutions, i.e. A3, B5, C1, etc. This suggests that combination A3-B5-C1-D2-E5-F4 could be a suitable combination for further concretisation. The working principles for this combination are highlighted in Figure 6.14.

Another way to make a rapid selection is to apply two-dimensional classification schemes, similar to the compatibility matrices shown in Figure 3.26. This will be illustrated using the gear coupling test rig shown in Figure 6.18.

The specification of the test rig demanded an axial displacement in the test coupling so that the axial forces which then appear could be measured. It was therefore necessary to move at least one half of the gear coupling.

The possible position of displacement (classifying criterion of the rows) and the axial force input (classifying criterion of the columns) were combined into the classification scheme shown in Figure 6.19. The various combinations were checked against the requirements list and unsuitable variants were eliminated for a number of immediately obvious reasons. These reasons were documented in the selection chart, but cannot be included because of space restrictions. The result is shown in the legend of Figure 6.19.

Selected working structures (the working combinations) now have to undergo further concretisation.

6.4.4 Practical Application of Working Structures

The development of working structures is the most important stage in the creation of original designs. This stage makes the most demands on the creativity of designers. This creativity is influenced by cognitive psychological processes associated with problem solving, by the use of a general working methodology, and by generally applicable solution finding and evaluation methods. As a consequence, various approaches can be employed at this stage and the one chosen depends on

TH Darmstadt	SELECTION CHART for Cylinder test rig	Part: 1 Page: 1

Solution variants (Sv) evaluated by
SELECTION CRITERIA
(+) Yes
(–) No
(?) Lack of information
(!) Check requirements list

DECISION
Mark solution variants (Sv)
(+) Pursue solution
(–) Eliminate solution
(?) Collect information (re-evaluate solution)
(!) Check requirements list for changes

Enter solution variant (Sv):

A: Compatibility assured
B: Fulfils demands of requirements list
C: Realisable in principle
D: Within permissible costs
E: Incorporates direct safety measures
F: Preferred by designer's company

Sv		A	B	C	D	E	F	G	Remarks (Indications, Reasons)	DECISION
A1	1	+	–						Change of set up, rattling of bearings	–
A2	2	+	–						Change of set up, too much play	–
A3	3	+	+	+	+				Sinusoidal motion not full realised, error < 1%	+
A4	4	+	+	+	–				Production expenditure too high	–
A5	5	+	+	+	–				Total expenditure too high	–
B1	6	+	–						Not adjustable or only with high expenditure	–
B2	7	+	–						Not adjustable	–
B3	8	+	+	–					Too slow	–
B4	9	+	+	–					Too slow, little variability	–
B5	10	+	+	+	+				Flexible allocation, very fast	+
C1	11	+	+	+	+				Simple solution	+
C2	12	+	+	+	?				Expenditure questionable	?
C3	13	+	+	–					Too much space required	–
C4	14	+	+	–					Expenditure too high	–
D1	15	+	+	–					No space, force flow path too flexible	–
D2	16	+	+	+	+				Preferred measuring procedure of institute	+
D3	17	+	+	+	–				Expenditure higher than D2	–
D4	18	–	–	+	–				Expenditure higher than D2	–
E1	19	+	+	–					Too much space required, element not stiff	–
..										
..										
..										

Date: Nov 01 Initials: La

Figure 6.17. Part of the selection chart for the solution space shown in Figure 6.14

the novelty of the task (the number of new problems to be solved), on the mentality, ability and experience of the designers, and on the product ideas from product planning or clients.

The procedure suggested in Sections 6.4.1 to 6.4.3 only provides the basis for an expedient stepwise design process. The actual process can vary considerably.

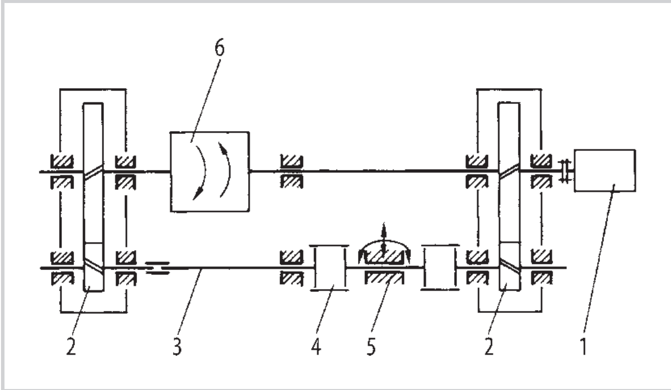


Figure 6.18. Sketch showing the principle of a test rig for gear couplings. *1* drive; *2* gearbox; *3* high-speed shaft; *4* test gear coupling; *5* adjustable bearing block for setting the alignment; *6* device for applying torque

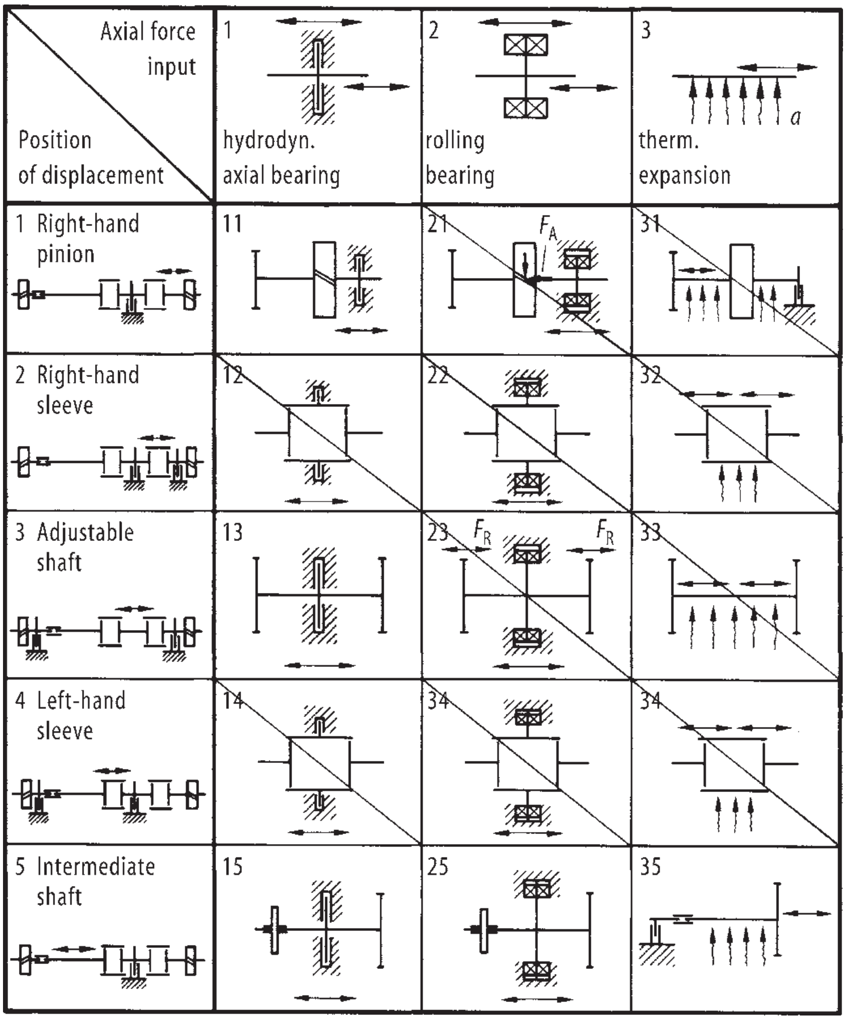


Figure 6.19. Systematic combination and elimination of variants that are unsuitable in principle.
Combinations 12, 14: Disturbance of coupling kinematics
Combination 21: F_A too great (life of rolling bearings too short)
Combination 23: 2 F_R, hence life of rolling bearings too short
Combinations 22, 24: Peripheral speed too great (life of rolling bearings too short)
Combinations 31–34: Thermal length too small

For *original designs without precedents*, the initial search for solutions should focus on the *main function* that appears to be *solution determining* for the overall function (see Figure 6.6). For the solution determining main function, one must first select some preliminary physical effects or working principles using intuition-based methods, literature and patent searches and previous products. The relationship between the functions in these solutions must be analysed to identify other important subfunctions for which physical effects and working principles need to be found. These working principles are selected from those that are compatible with the other working principles selected to fulfil the main functions. A simultaneous, independent search for working principles for all subfunctions will, in general, be too elaborate and will result in several working principles that will have to be eliminated later from the overall combination.

It is recommended that the most promising solution principles (not more than six) should be identified at a relatively low level of concretisation. One of these is then selected for elaboration to a higher level of concretisation. From the variants that then emerge at this level, the most promising is again taken forward to an even higher level of concretisation. Adopting this approach avoids the need to deal with too many variants at the same time, which can result in too much effort being devoted to variants that eventually turn out to be unsuitable.

An important strategy for the creation of solution fields is therefore the systematic variation of the physical effects and form design features that were recognised as being essential in the initial solutions. *Classification schemes* are very useful but usually need several trials, based on variation and correction of the classifying criteria, before an optimum scheme can be arrived at. This requires some experience.

When *concrete solution ideas* are available from product planning or other sources, these have to be analysed to identify their essential solution determining characteristics. These are then systematically varied and combined to arrive at a solution field.

In the case of *evolutionary developments*, the known working principles and working structures should be checked to see if they still meet current technological standards and the latest requirements.

When an approach is strongly based on *intuition*, or when previous experience is applicable, working structures that fulfil the overall function will often be found directly without first searching for solutions for the individual subfunctions.

In particular, the *stepwise* generation of working principles, through the search for physical effects and the subsequent form design features, is often integrated mentally by producing *sketches of solutions*. This is because designers think more in configurations and representation of principles than in physical equations.

The use of intuition-based and discursive–systematic methods can quickly lead quickly to extensive solution fields. To limit subsequent design effort, these should be reduced as soon as *feasible working principles* emerge by checking against the demands in the requirements list.

At this stage it is often not possible to assess the characteristics of a principle solution with quantitative data, particularly with regard to production and cost.

Therefore, the selection of suitable working principles requires an *interdisciplinary team* discussion, similar to a value analysis team (see Section 1.2.3(2)), in order to base the qualitative decision on a broad spectrum of experience.

6.5 Developing Concepts

6.5.1 Firming Up into Principle Solution Variants

The principles elaborated in Section 6.4 are usually not concrete enough to lead to the adoption of a definite concept. This is because the search for a solution is based on the function structure, and so it is aimed, first and foremost, at the fulfilment of a technical function. A concept must, however, also satisfy the conditions laid down in Section 2.1.7—at least in essence—for only then is it possible to evaluate it. Before concept variants can he evaluated they must be firmed up, and experience has shown that this almost invariably involves considerable effort.

The selection process may already have revealed gaps in information about very important properties, sometimes to such an extent that not even a rough and ready decision is possible, let alone a reliable evaluation. The most important properties of the proposed combination of principles must first be given a much more concrete *qualitative*, and often also a rough *quantitative*, definition.

Important characteristics of the working principle (such as performance and susceptibility to faults), of the embodiment (such as space requirements, weight and service life) and finally of important task-specific constraints must all be known, at least approximately. More detailed information need only be gathered for promising combinations. If necessary, a second or third selection process should follow the collection of further information.

The necessary data are essentially obtained with the help of such proven methods as:

- rough calculations based on simplified assumptions
- rough sketches or rough scale-drawings of possible layouts, forms, space requirements, compatibility, etc.
- preliminary experiments or model tests used to determine the main properties or to obtain approximate quantitative statements about the performance and scope for optimisation
- construction of models in order to aid analysis and visualisation (for example, kinematic models)
- analogue modelling and systems simulation, often with the help of computers; for example stability and loss analyses of hydraulic systems using electrical analogies
- further searches of patents and the literature with narrower objectives
- market research of proposed technologies, materials, bought-out parts, etc.

With these fresh data it is possible to firm up the most promising combinations of principles to the point at which they can be evaluated (see Section 6.5.2). The variants must reveal technical as well as economic properties, thus permitting the most accurate evaluation possible. When firming up into principle solutions, it is therefore advisable to keep in mind potential evaluation criteria (see Section 3.3.2), as this encourages purposeful elaboration of the information.

An example will show how it is possible to firm up working principles into principle solutions. To that end, we return once more to our fuel gauge.

Figure 6.20 shows the working principle of the first proposal shown in Figure 3.27 and Table 3.3. It is possible to obtain the total force statically, either by measuring three bearing forces or by measuring just one bearing force in combination with a pivot. The weight of the contents of the fuel tank, to be used as a measure of the quantity of liquid, can be determined by deducting the weight of the empty tank. The measuring devices to be used, however, measure the total force, including those components caused by accelerations. If the force is converted into motion it can be detected via a potentiometer for example.

Estimates of the weights and inertia forces form the basis of the firming up procedure.

Total force of 20 to 160 litres of the liquid (static):

$$F_{\text{tot}} = \varrho \cdot g \cdot V = 0.75 \times 10 \times (20 \ldots 160) = (150 \ldots 1200)\ \text{N (fuel)}\ .$$

Additional forces due to acceleration $\pm 30\ \text{m/s}^2$ (only the liquid is taken into consideration):

$$F_{\text{add}} = m \cdot a = (15 \ldots 120) \times \pm 30 = \pm(450 \ldots 3600)\text{N}\ .$$

The suppression of motions resulting from accelerational forces calls for considerable damping.

Conclusion: develop solution further, provide damping, seek appropriate subsolutions and firm up by means of rough scale drawings. Figure 6.21 shows the result. Once the necessary parts and their arrangements are drawn, the proposal can be evaluated. This confirms the indication in the selection chart (see Figure 3.27) that the effort required to complete solution variant 1 could be too high.

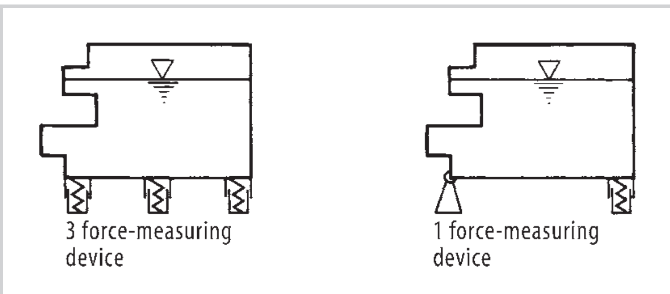


Figure 6.20. Solution principle 1 (Figure 3.27 and Table 3.3): measure weight of liquid (signal = force)

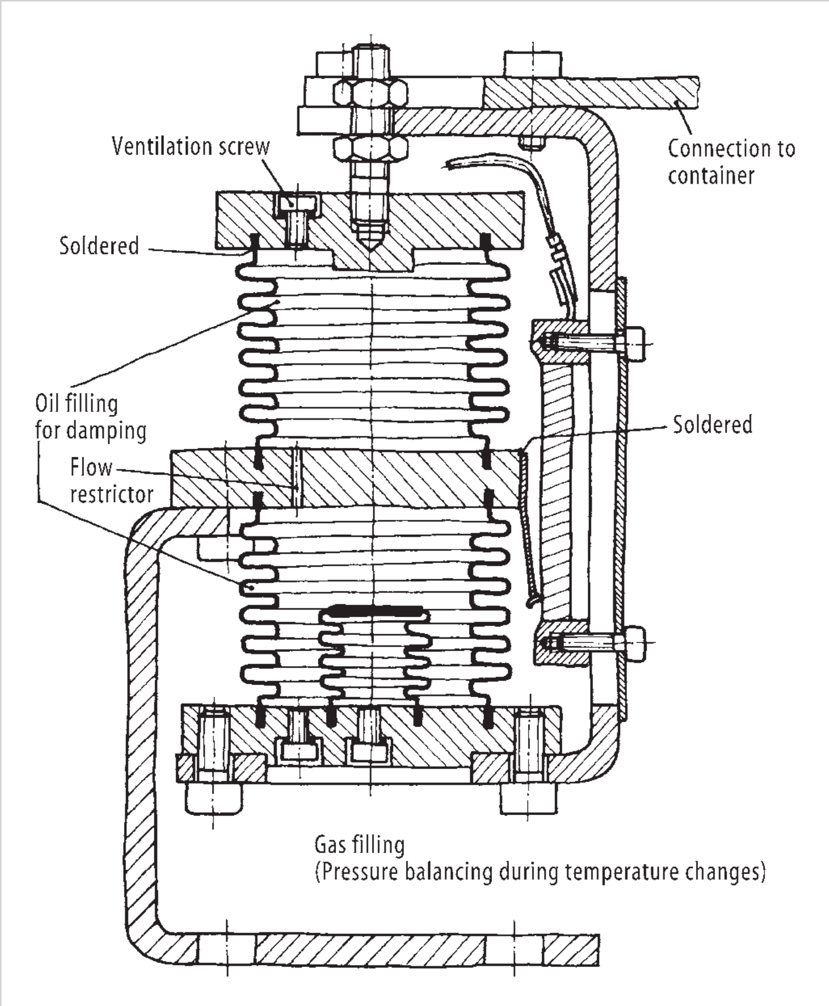


Figure 6.21. Firmed up principle solution shown in Figure 6.20

6.5.2 Evaluating Principle Solution Variants

In Section 3.3.2 we explained generally applicable evaluation methods, in particular Cost–Benefit Analysis and the VDI 2225 procedure [6.15].

When evaluating principle solution variants, the following steps are recommended.

1. Identifying Evaluation Criteria

This step is based, first of all, on the *requirements list*. During a previous selection procedure (see Section 6.4.3) unfulfilled demands may have led to the elimination of variants that were found to be unsuitable in principle. Further information was subsequently gathered during firming up into principle solutions. Hence it is advisable, with all the newly acquired information, to establish whether all of the proposals to be evaluated still satisfy the demands of the requirements list. This can involve new yes/no decisions—a new selection process.

Even though we are at a more concrete stage, we cannot expect this decision to be made with certainty for all of the variants unless much further effort is applied, which the designers may not wish or are not able to provide at this stage. At the current level of information, it may only be possible to decide how likely it is that certain requirements can be fulfilled. In that case, the likelihood of fulfilling particular requirements may become an additional evaluation criterion.

A number of requirements are minimum requirements. It is important to establish whether or not these should be exceeded. If they should, further evaluation criteria may be needed.

For evaluation during the conceptual phase, both the *technical* and the *economic characteristics* should be considered as early as possible [6.4]. At the firming up stage, however, it is not usually possible to give the costs in figures. Nevertheless, the economic aspects must be taken into consideration, at least qualitatively, and so must industrial and environmental safety requirements.

Hence it is necessary to consider technical, economic and safety criteria at the same time. It is suggested that the evaluation criteria are derived from the main headings in Figure 6.22. These are in accordance with the embodiment design checklist (see Section 7.6) and other proposals [6.8].

Main headings	Examples
Function	Characteristics of essential auxiliary function carriers that follow out of necessity from the chosen solution principle or concept variant
Working principles	Characteristics of the selected principle or principles with respect to simple and clear-cut functioning, adequate effect, few disturbing factors
Embodiment	Small number of components, low complexity, low space requirement, no special problems with layout or form design
Safety	Preferential treatment of direct safety techniques (inherently safe), no additional safety measures needed, industrial and enviromental safety guaranteed
Ergonomics	Satisfactory man-machine relationship, no strain or impairment of health, good aesthetics
Production	Few and established production methods, no expensive equipment, small number of simple components
Quality control	Few tests and checks needed, simple and reliable procedures
Assembly	Easy, convenient and quick, no special aids needed
Transport	Normal means of transport, no risks
Operation	Simple operation, long service life, low wear, easy and simple handing
Maintenance	Little and simple upkeep and cleaning, easy inspection, easy repair
Recycling	Easy recovery of parts, safe disposal
Costs	No special running or other associated costs, no scheduling risks

Figure 6.22. Checklist with main headings for design evaluation during the conceptual phase

Every heading in the checklist relevant to the task must be assigned at least one evaluation criterion. The criteria must, moreover, be independent of one another in terms of the overall objective, so as to avoid multiple evaluations. Consumer criteria are essentially contained in the first five and last three headings, while producer criteria are contained in the following headings: embodiment, production, quality control, assembly and costs.

Evaluation criteria are accordingly derived from:

1. The requirements list:
   - Probability of satisfying the demands (how probable, despite which difficulties?)
   - Desirability of exceeding minimum requirements (exceed by how much?)
   - Wishes (satisfied, not satisfied, how well are they satisfied?)
2. General technical and economic characteristics from the checklist, see Figure 6.22 (to what extent are they present, how well are they satisfied?)

During the conceptual phase, the total number of evaluation criteria should not be too high: 15–30 criteria are usually enough (see Figure 6.41).

2. Weighting the Evaluation Criteria

The evaluation criteria adopted may differ markedly in importance. During the conceptual phase, in which the level of information is fairly low because of the relative lack of embodiment, weighting is not generally advisable. It is much more advantageous in the selection of evaluation criteria to strive for an approximate balance, ignoring low-weighted characteristics for the time being. As a result, evaluation will be concentrated on the main characteristics and hence provide a clear picture at a glance. Extremely important requirements, however, which cannot be ignored until later, must be introduced with the help of weighting factors.

3. Compiling Parameters

It has proved useful in the past to list the identified evaluation criteria in the sequence of the checklist headings and to assign the parameters of the variants to them. Whatever quantitative information is available at this stage should also be included. Such quantitative data generally result from the step we have called "firming up into principle solution variants". However, since it is impossible to quantify all the parameters during the conceptual phase, the qualitative aspects should be put into words and correlated with the value scale.

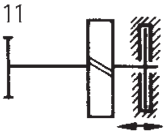
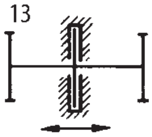
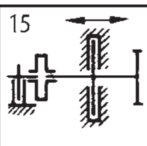
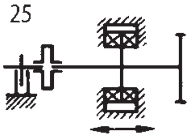
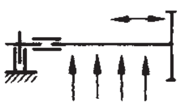
4. Assessing Values

Though the attribution of points raises problems, it is not advisable to evaluate too timidly during the conceptual phase.

Those using the 0–4 scale proposed in VDI Guideline 2225 may feel the need to assign intermediate values, particularly when there are many variants, or when the evaluation team cannot agree on a precise point. It may prove helpful in such cases to attach a tendency sign (↑ or ↓) to the point in question (see Figure 6.41). Identifiable tendencies can then be taken into account when estimating the evaluation uncertainties. The 0–10 scale, again, may suggest a degree of accuracy that does not really exist. Here, arguments about a point are often superfluous. If there is absolute uncertainty in the attribution of points, which happens quite often during the evaluation of concept variants, the point under consideration should be indicated with a question mark (see Figure 6.41).

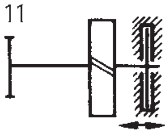
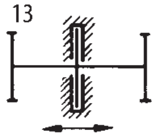
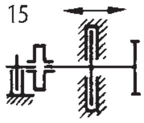
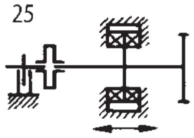
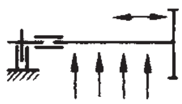
During the conceptual phase it may prove difficult to put actual figures to the costs. It is not therefore generally possible to establish an *economic rating* R_e with respect to the production costs. Nevertheless, the technical and economic aspects can be identified and separated qualitatively, to a greater or lesser extent. The *strength diagram* (see Figure 3.35) can be used to much the same effect (see also Figures 6.23 to 6.25 which are for the test rig shown in Figure 6.18).

In a similar way, a classification based on consumers' and producers' criteria often proves useful. Since the consumers' criteria usually involve *technical ratings* R_t

Variants / Technical criteria	11 	13 	15 	25 	35 
1) Small disturbance of coupling kinematics	(1) 3	4	4	4	3
2) Simple operation	3	4	4	4	3
3) Easy exchange of coupling	4	3	4	4	4
4) Functional safety	2	4	3	3	3
5) Simple construction	(1) 2	2	2	2	3
Total	14	17	17	17	16
$R_t = \frac{\text{Total}}{20}$	0.7	0.85	0.85	0.85	0.80

(1) Torque changes with axial displacement of pinion

Figure 6.23. Technical evaluation of the remaining principle solution variants, see Figure 6.19

Variants / Economic criteria	11 	13 	15 	25 	35 
1) Low material costs	2	3	4	4	(1) 2
2) Low reassembly costs	2	(2) 1	3	3	3
3) Short testing time	2	4	3	3	2
4) Possibility of manufacturing in own workshop	3	3	3	3	2
Total	9	11	13	13	9
$R_e = \frac{\text{Total}}{16}$	0.56	0.69	0.81	0.81	0.56

(1) Austenitic shaft (2) Torque measuring shaft must be moved

Figure 6.24. Economic evaluation of the remaining principle solution variants, see Figure 6.19

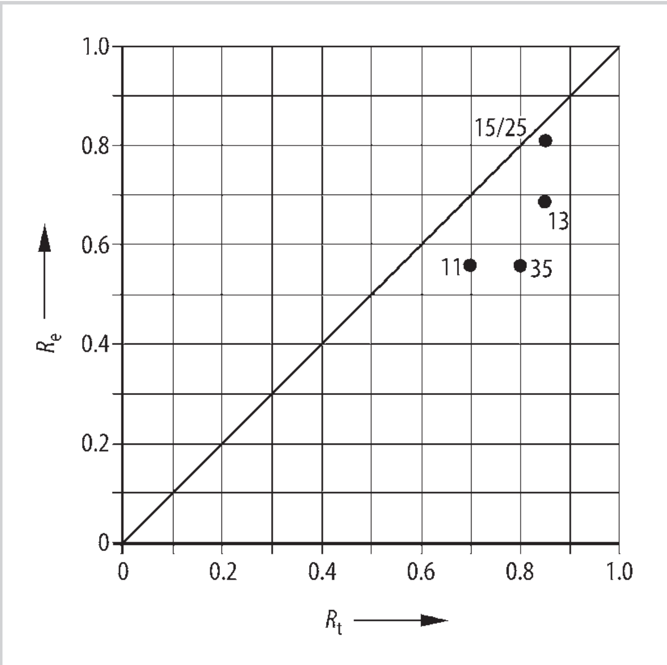


Figure 6.25. Comparison of the technical and economic ratings of the principle solution variants in Figures 6.23 and 6.24

and the producers' criteria involve *economic ratings* R_e, it is possible to proceed to a similar classification to the one mentioned above.

Depending on the problem and the amount of information available, one of the following three possible forms of representation is chosen:

- technical rating with implicit economic aspects (see Figures 6.41 and 6.55)
- separate technical and economic ratings (see Figures 6.23 to 6.25)
- additional comparison of consumers' and producers' criteria.

5. Determining Overall Value

The determination of the overall value is a matter of simple addition, once the points have been assigned to the evaluation criteria and the variants. If, because of the evaluation uncertainty, it is only possible to assign a range of points to individual variants, or if tendency signs are used, one can additionally determine the possible minimum and maximum overall point number and so obtain the probable overall value range (see Figure 6.41).

6. Comparing Concept Variants

An absolute value scale is generally more suitable for the purposes of comparison. In particular, it makes it fairly simple to tell whether particular variants are relatively close to or far from the target (theoretical ideal).

Concept variants that are some 60% below the target are not worth further development. Variants with ratings above 80% and a balanced value profile—those without extremely bad individual characteristics—can generally be moved on to the embodiment design phase without further improvement.

Intermediate variants should only be released for embodiment design after the elimination of weak spots or an improved combination.

It often happens that two or more variants are found to be practically equivalent. It is a very grave mistake, in that case, to base the final decision on such slight differences. Instead, evaluation uncertainties, weak spots and the value profile should be looked at more closely (see Figure 3.38). It may also be necessary to firm up on such variants in a further step. Schedules, trends, company policy and so on must be assessed separately and taken into account [6.4].

7. Estimating Evaluation Uncertainties

This step is very important, especially during the conceptual phase, and must not be omitted. Evaluation methods are mere tools, not automatic decision mechanisms. Uncertainties must be determined as indicated earlier. At this point, however, only the information gaps that impact on the best concept variants (for example, variant B in Figure 6.41) need to be closed.

8. Searching for Weak Spots

During the conceptual phase, the value profile plays an important role. Variants with a high rating but definite weak spots (unbalanced value profile) may prove extremely troublesome during subsequent development. If, because of an unidentified evaluation uncertainty, which is more likely to occur in the conceptual than in the embodiment phase, a weak spot should make itself felt later, then the whole concept may be put in doubt and all the development work may prove to have been in vain.

In such cases it is very much less risky to select a variant with a slightly lower rating but a more balanced value profile (see Figure 3.38).

Weak spots in favourite variants can often be eliminated by the transfer of better subsolutions from other variants. Moreover, with better information, it is possible to search for a replacement for the unsatisfactory subsolution. Thus the criteria we have listed played an essential role in the selection of the best variant in the problem discussed in Section 6.6 (see Figure 6.41). When estimating evaluation uncertainties and also when searching for weak spots it is advisable to assess the probability and magnitude of the possible risk, especially in the case of important decisions.

6.5.3 Practical Application of Developing Concepts

The selection of the concept, or the principle solution, provides the *basis for starting the embodiment design phase* (see Figure 6.1). This often indicates a need for changes in organisation and personnel because the nature of the work alters. Thus, firming up of suitable working structures into principle solution variants and the subsequent evaluation at the end of the conceptual design phase are of major importance for product development. The large number of variants has to be reduced to one concept, or just a few, to be pursued further. This decision incurs a heavy responsibility and can only be made when the principle solutions are in a state suitable for evaluation. In extreme cases this may require rough scale layouts backed up by preliminary calculations and sometimes tests. From research in industry and universities [6.8], it is known that calculating and representation add up to 60% of the total time spent on conceptual design.

The *representation* of working principles and working structures is likely to remain the domain of conventional sketching. Rough layouts, and in particular the more important details of solutions are now commonly represented using CAD. Sketching working structures by hand has the advantage that one does not need to consider the formalities of CAD user interfaces during this highly creative stage. Firming up solution principles using CAD is useful, despite the effort needed to enter the initial product model into the system, because making variations to the layout and individual components becomes very efficient. For dynamic systems it is also possible to do initial simulations using the CAD model.

In any case, it is expedient (for *reasons of efficiency* and to identify essential characteristics) not to firm up the whole working structure to the same level of detail. The aim should be to focus on those working principles, components or parts of the structure that are essential for the evaluation of the concepts and the selection of the one that will be transferred to the embodiment stage. Richter provides proposals for this task [6.10].

At this point it must be emphasised again that *iterations* often occur in the steps mentioned in Sections 6.4 and 6.5. On the one hand, it might be necessary to detail working principles in order to combine and select them, and on the other hand a completely new idea for a working principle might emerge while making a rough layout of a principle solution.

It must be stressed that principle solutions or concepts have to be *unambiguously documented*. It must also be clear which parts of the working structure or function carriers can be realised by existing and standard components, and which ones will need to be specially designed.

6.6 Examples of Conceptual Design

This section provides two examples of how the approach can be applied: the first to a task whose main flow is material, and the second to one whose main flow is energy. The embodiment design phase of the second example is continued in Section 7.7. An example of signal flow has been used throughout the previous sections in this chapter (see Figures 6.4 to 6.6 and 6.20).

6.6.1 One-Handed Household Water Mixing Tap

A one-handed mixing tap is a device for regulating water temperature and throughflow independently with one hand. This task was sent to the design department by the planning department in the form shown in Figure 6.26.

One-handed water mixing tap

Required: one-handed household water mixing tap with the following characteristics:

Throughput	10 l/min
Max. pressure	6 bar
Normal pressure	2 bar
Hot water temperature	60 °C
Connector size	10 mm

Attention to be paid to appearance. The firm's trade mark to be prominently displayed. Finished product to be marketed in two years' time. Manufacturing costs not to exceed DM 30 each at a production rate of 3000 taps per month.

Figure 6.26. One-handed mixing tap. Example of an assignment suggested by the product planning department

Step 1: Clarifying the Task and Setting Up the Requirements List

New data on fittings, standards, safety regulations and ergonomic factors led to the replacement of the original requirements list by the revised version shown in Figure 6.27.

TH Darmstadt		Requirements list for one-handed mixing tap		Page 1
Changes	D W	Requirements		Responsible
	D	1 Throughput (mixed flow) max. 10 l/min at 2 bar		KMW
	D	2 Max. pressure 10 bar (test pressure 15 bar as per DIN 2401)		LTMB
	D	3 Temp. of water standard 60 °C, 100 °C (short-time)		
	D	4 Temperature setting independent of throughput and pressure		
	W	5 Permissible temp fluctuation ± 5 °C at a pressure diff. of ± 5 bar between hot and cold supply		
	D	6 Connection 2 x Cu pipes, 10 x 1 mm, l = 400 mm		
	D	7 Single-hole attachment $\varnothing 35^{+2}_{-1}$ mm, basin thickness 0–18 mm (Observe basin dimension DIN EN 31, DIN EN 32, DIN 1368)		
	D	8 Outflow above upper edge of basin, 50 mm		
	D	9 To fit household basin		
	W	10 Convertible into wall fitting		
	D	11 Light operation (children)		
	D	12 No external energy		
	D	13 Hard water supply (drinking water)		
	D	14 Clear identification of temperature setting		
	D	15 Trade mark prominently displayed		
	D	16 No connection of the two supplies when valve shut		
	W	17 No connection when water drawn off		
	D	18 Handle not heated to above 35 °C		
	W	19 No burns from touching the fittings		
	W	20 Provide scalding protection if extra costs small		
	D	21 Obvious operation, simple and convenient handling		
	D	22 Smooth, easily cleaned contours, no sharp edges		
	D	23 Noiseless operation, (≤ 20 dB as per DIN 52218)		
	W	24 Service life 10 years at about 300 000 operations		
	D	25 Easy maintenance and simple repairs. Use standard spare parts		
	D	26 Max. manuf. costs DM 30 (3000 units per month)		
	D	27 Schedules from inception of development conceptual design / embodiment design / detail design / prototype after 2 / 4 / 6 / 9 months		
		Replaces 1st issue of 12.6.1973		

Figure 6.27. Requirements list for a one-handed mixing tap

Step 2: Abstracting to Identify the Essential Problems

The basis for abstraction is the requirements list, from which it is possible to arrive at Figure 6.28. Simple household solutions for mixing taps suggested that the chosen solution principle must be based on metering out the water through a diaphragm or valve. Alternatives such as heating and cooling by the introduction of external energy through heat exchangers could be dismissed: they were more expensive and involved a time lag. Selecting sound solution principles without further investigation, because they have proved their worth in previous company products, is a common and justified approach in some branches of engineering.

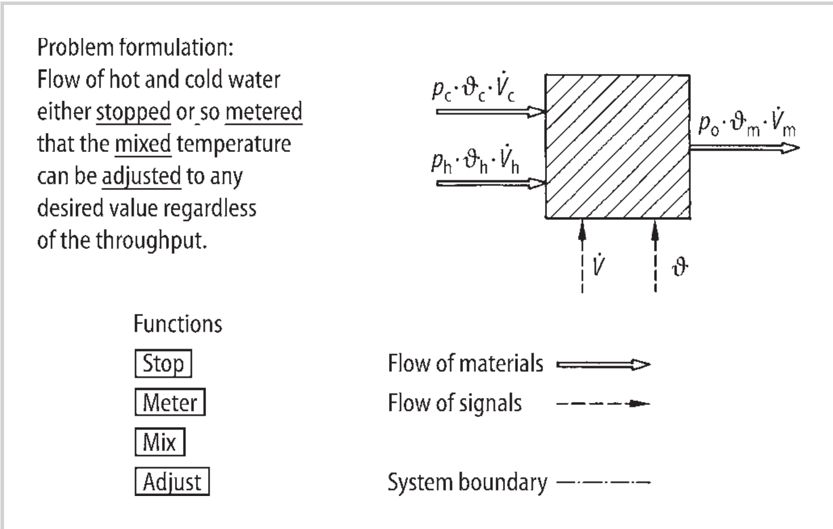


Figure 6.28. Problem formulation and overall function as per the requirements list, see Figure 6.27. $\dot{V}$ = volume rate, p = pressure, ϑ = temperature. Index: c = cold, h = hot, m = mixed, o = atmosphere

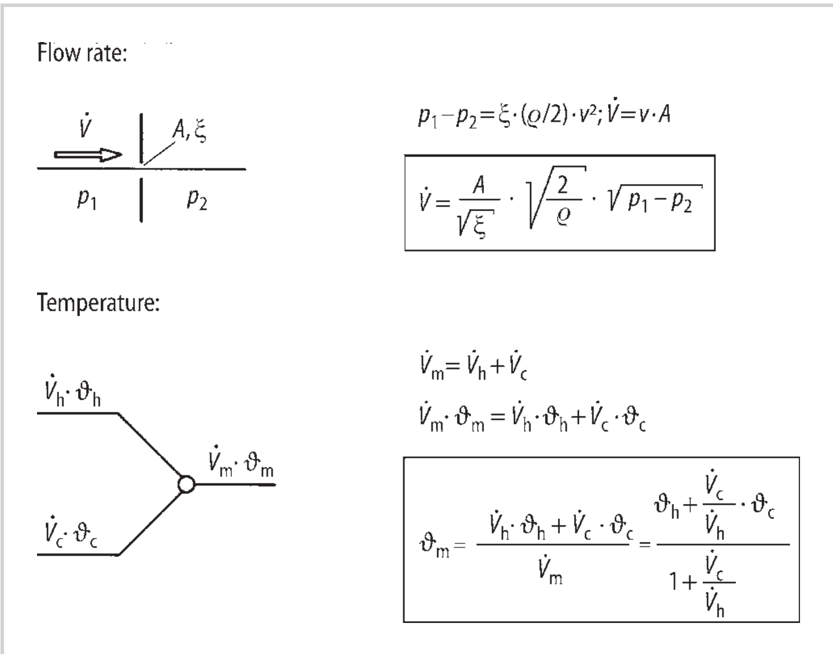


Figure 6.29. Physical relationships for flow rate and temperature of a mixed flow of the same fluid

Next, the physical relationships for the diaphragm (or valve) flow rate and the temperature of a mixed flow of similar fluids were determined (see Figure 6.29).

Temperature and flow rate adjustments are based on the same physical principle—a diaphragm or valve.

Upon *changing the flow rate* $\dot{V}_m$, the flows must be changed linearly and in the same sense as the signal setting $s_{\dot{V}}$. The output temperature ϑ_m, however, must remain unchanged: that is, the relationship $\dot{V}_c/\dot{V}_h$ must remain constant and independent of the signal positions $s_{\dot{V}}$.

Upon *changing the output temperature* ϑ_m, the volume flow rate $\dot{V}_m$ must remain unchanged: that is, the sum of $\dot{V}_c + \dot{V}_h = \dot{V}_m$ must remain constant. To that end the component flows $\dot{V}_c$ and $\dot{V}_h$ must be changed linearly and in the opposite sense to the signal setting for the output temperature s_ϑ.

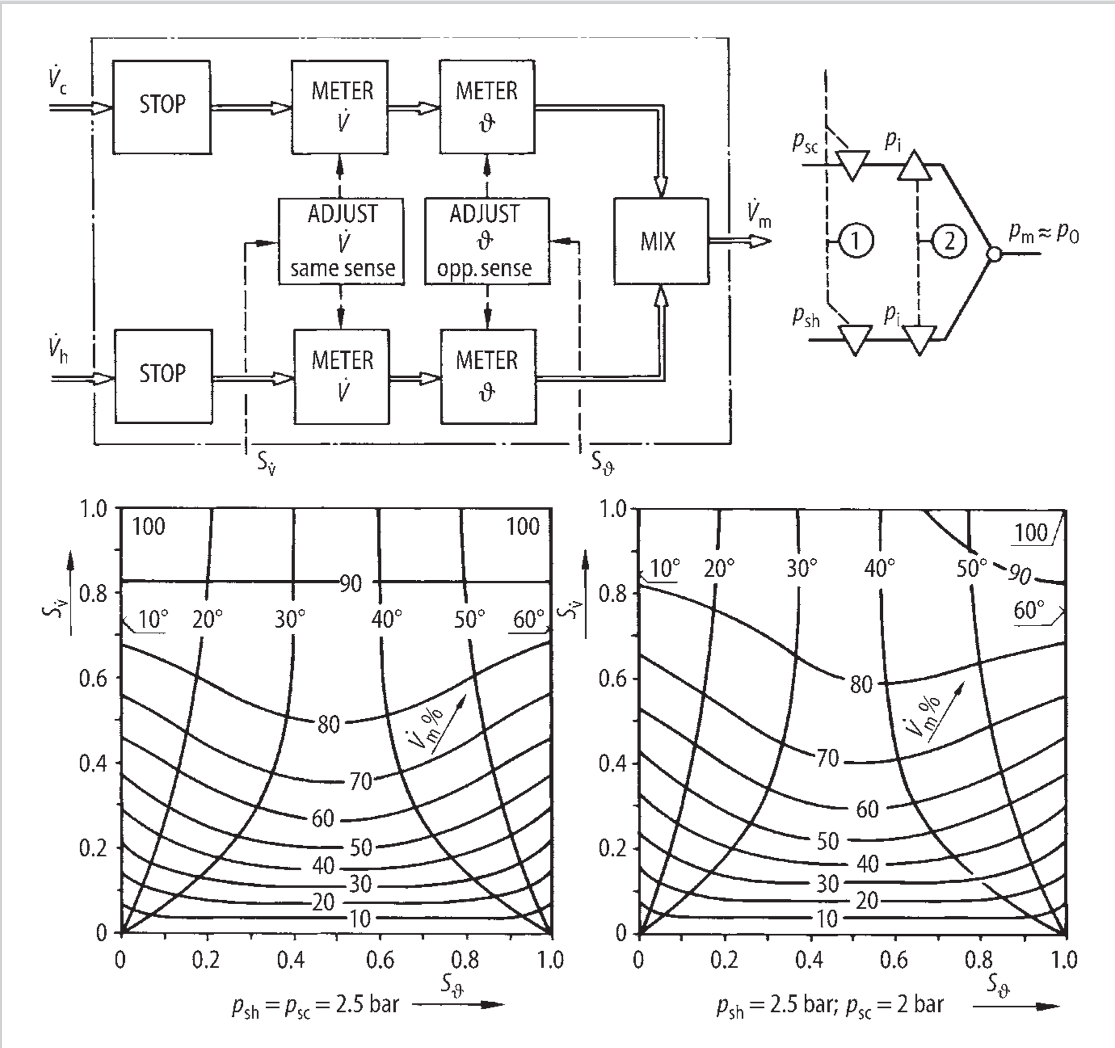


Figure 6.30. Function structure for a one-handed water mixing tap based on Figure 6.28, metering flow ① and adjusting temperature ② separately before mixing. In the graphs, lines of constant temperature and constant percentage flow rate have been plotted for given temperature settings (s_ϑ) and flow rate settings ($s_{\dot{V}}$). Due to the mutual effects of the pressures on the inlets at ① and ②, the temperature and flow characteristics are not linear except for the setting $s_{\dot{V}} = 0.825$, and hence are unsuitable for small flow rates. At a particular pressure difference between the cold and hot water supplies (in this case $p_{sh} - p_{sc} = 0.5$bar) the lines shift. The settings are no longer independent of each other, even for the settings $s_{\dot{V}} = 0.825$ (*diagram on right*)

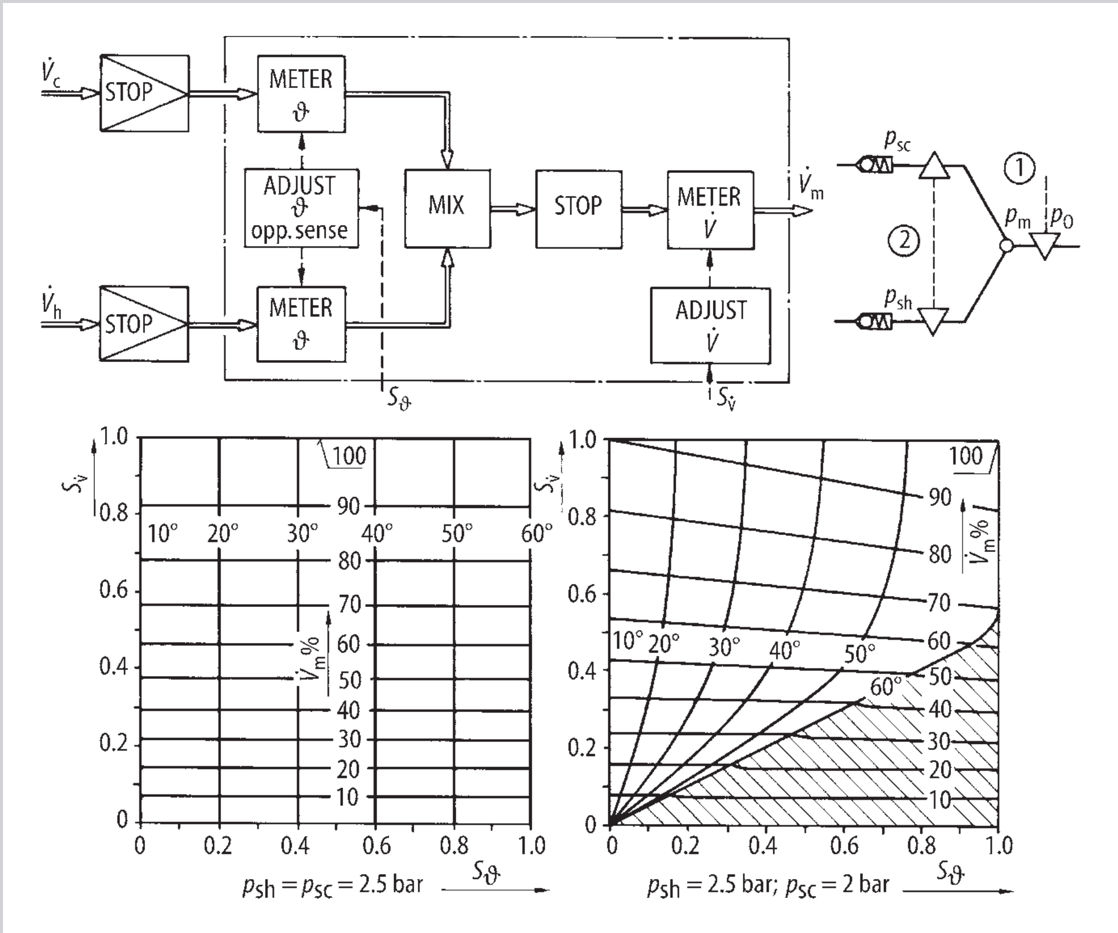


Figure 6.31. Function structure based on Figure 6.28 in which the temperature is set before and the flow metered after mixing. With equal pressures in the supply pipes, the flow and temperate settings are independent of each other due to equal pressure differences across each temperature-flow-metering valve. The behaviour is linear. With different supply pressures, however, the characteristic ceases to be linear and is strongly displaced, especially with small quantities, when the pressure in the mixing chamber approximates the smaller supply pressure. If it is exceeded, then only cold or (here) hot water will run out regardless of the temperature setting

Step 3: Establishing Function Structures

The first function structure was derived from the subfunctions:

- Stop–meter–mix
- Adjust flow rate
- Adjust output temperature.

Since the physical principle was well-known—metering using a valve—the structural layout of the first function structure was varied and developed to determine the best system and its behaviour (see Figures 6.30 to 6.32). From the results, the function structure shown in Figure 6.32 was chosen as being the most satisfactory because of its approximately linear characteristic for the output temperature.

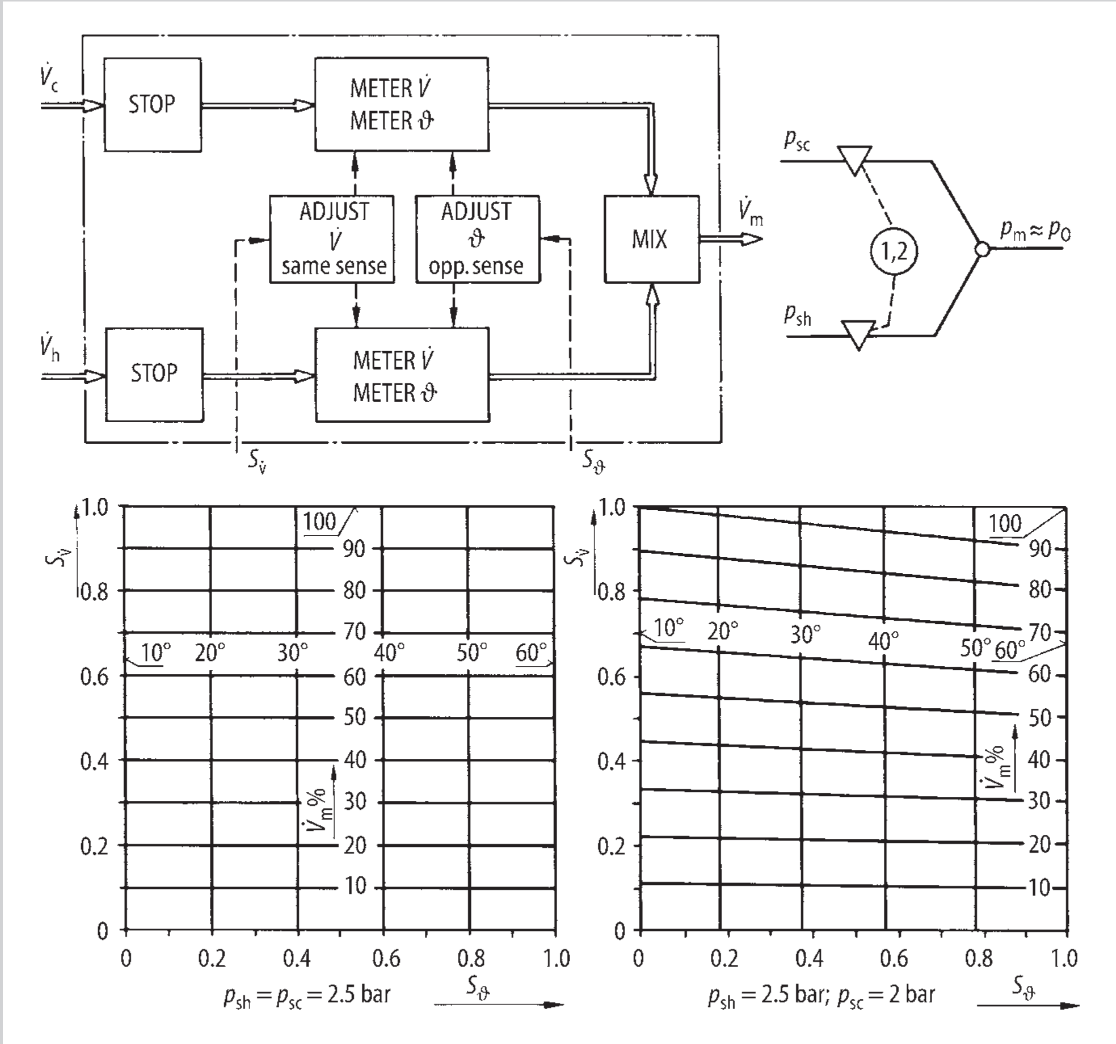


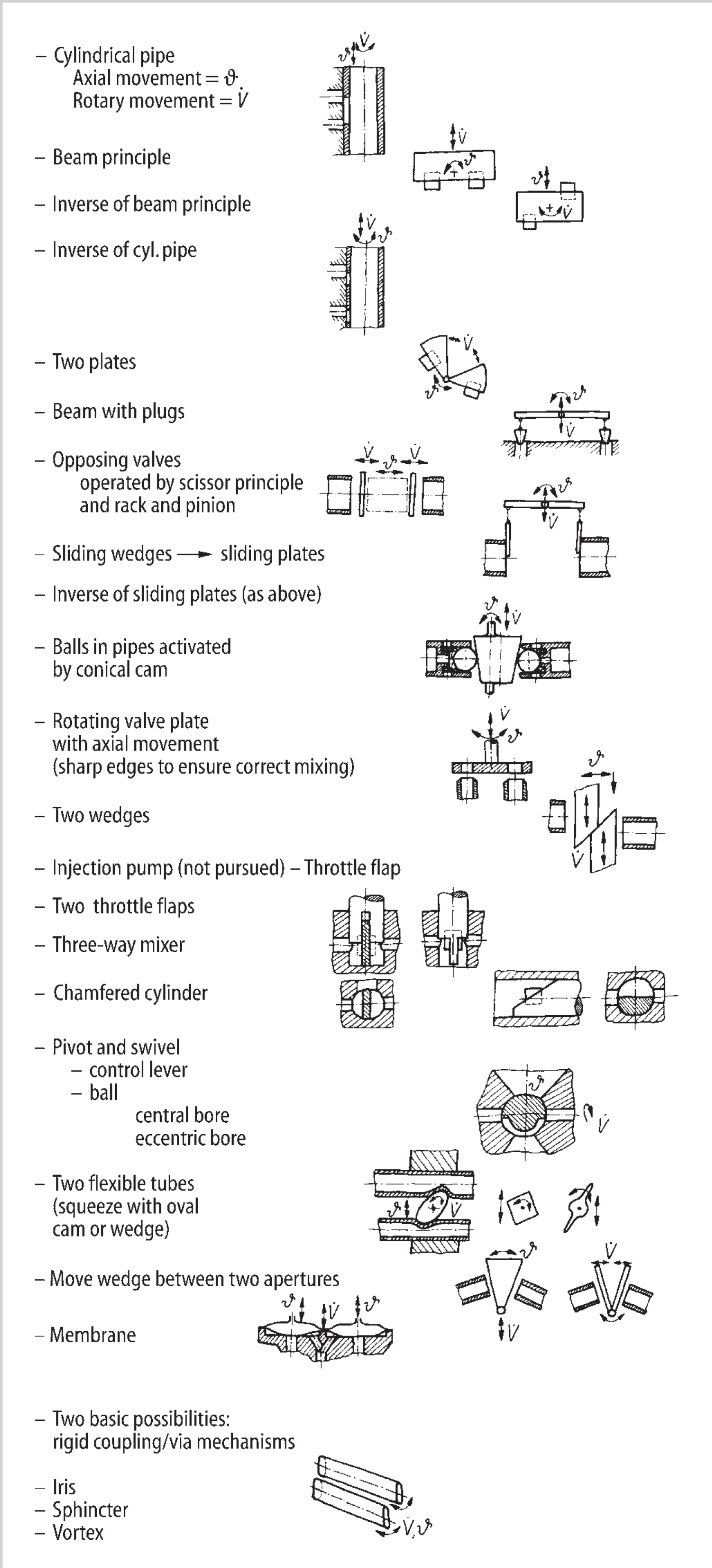
Figure 6.32. Function structure based on Figure 6.28, in which the temperature and flow at each inlet is metered out independently and then mixed. Linear temperature and flow characteristics are obtained. No serious changes are seen, even at different supply pressures

Step 4: Searching for Working Principles

Because the function structure shown in Figure 6.32 exhibited the best behaviour, the task became one of "varying two flow areas, simultaneously or successively, in one sense by one movement and in the opposite sense by a second, independent, movement". Brainstorming was used as a first attempt to find solutions. The results are shown in Figure 6.33.

The solutions suggested during the brainstorming session were checked, in particular, to establish whether the $\dot{V}$ and ϑ settings were independent. An analysis of the combined movements suggested the following characteristics for the working principles that were generated:

Figure 6.33. Result of a brainstorming session to discover solution principles for the assignment "vary two flow areas, simultaneously or successively, in one sense by one movement and in the opposite sense by a second, independent movement"


– Cylindrical pipe
Axial movement = ϑ
Rotary movement = $\dot{V}$
– Beam principle
– Inverse of beam principle
– Inverse of cyl. pipe
– Two plates
– Beam with plugs
– Opposing valves
operated by scissor principle
and rack and pinion
– Sliding wedges ⟶ sliding plates
– Inverse of sliding plates (as above)
– Balls in pipes activated
by conical cam
– Rotating valve plate
with axial movement
(sharp edges to ensure correct mixing)
– Two wedges
– Injection pump (not pursued) – Throttle flap
– Two throttle flaps
– Three-way mixer
– Chamfered cylinder
– Pivot and swivel
– control lever
– ball
central bore
eccentric bore
– Two flexible tubes
(squeeze with oval
cam or wedge)
– Move wedge between two apertures
– Membrane
– Two basic possibilities:
rigid coupling/via mechanisms
– Iris
– Sphincter
– Vortex

1. Solutions with separate movements for $\dot{V}$ and ϑ tangential to the valve seat face

- The independence of the $\dot{V}$ and ϑ settings is only guaranteed if each of the flow areas of the valves are bounded by two edges running parallel to the corresponding movements. This implies that the movements must proceed at an angle to each other and in a straight line. Every valve setting thus has two pairs of straight and parallel bounding edges (see Figure 6.34). This ensures that when one setting is adjusted the other setting is not simultaneously adjusted.
- Distribution of bounding edges: each of the components producing the valve flow areas must have at least two edges that face each other and lie in the direction of the movement.
- When setting $\dot{V}$, both valve areas must approach zero simultaneously.
- When setting ϑ, one area must approach zero as the other approaches its maximum $\dot{V}_{max}$.
- This implies, when setting $\dot{V}$, that the bounding edges on both valve areas must move towards each other or away from each other in the same sense. When setting ϑ, the bounding edges on the two valve areas must move in the opposite sense to each other.
- The seat face may be plane, cylindrical or spherical.
- Solutions of this type can be effected with a single valve element, and seem simple to design.

2. Solutions with separate movements for $\dot{V}$ and ϑ normal to the valve seat face

- This group includes all movements which involve lifting a valve from its seat face. However, only a movement at right angles to the seat face is possible in practice.
- The independent settings of $\dot{V}$ and ϑ can only be achieved with additional control elements (coupling mechanism).
- The design seems to require greater effort.

3. Solutions with one type of movement for $\dot{V}$ and ϑ tangential to the seat face

- To guarantee the independence of the $\dot{V}$ and ϑ settings, additional coupling elements are needed.

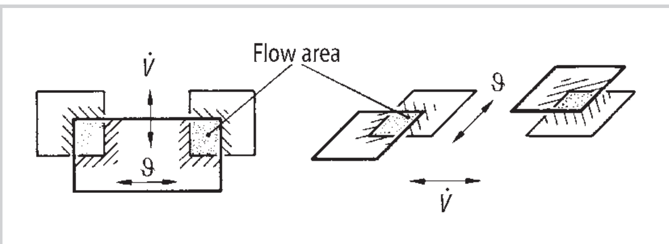


Figure 6.34. Movements and bounding edges of valve positions

	Classifying Criteria	Parameters
Rows	Form of working elements	Flate plate Wedge (–) Cylinder Cone (–) Ball Special elastic body (–)
Columns	Coupling of movements	Direct (one part) Indirect (mechanism) (–)
	Movement	$\left.\begin{array}{l}\dot{V}\\ \vartheta\end{array}\right\}$ One element $\left.\begin{array}{l}\dot{V}\\ \vartheta\end{array}\right\}$ Several elements
	Direction of movement $\dot{V}$ and ϑ Type of movement $\dot{V}$ and ϑ	Normal to seat face (⊥) (–) Tangential to seat face (⇌) Transitional Rotational

Figure 6.35. Classifying criteria and parameters for working principles of one-handed water mixing tap

- The solutions are similar to those listed under 2. They only differ in the shape of the seat face and the resulting movement.

4. Solutions with one movement for $\dot{V}$ normal to, and one movement for ϑ tangential to, the seat face and vice versa

- These solutions do not, even with the help of coupling mechanisms, satisfy the demand for independent $\dot{V}$ and ϑ settings. The overall function is not achieved.

The first group of solutions (movements for $\dot{V}$ and ϑ tangential to the valve seat face) have unambiguous behaviour and seem to be less complex. Therefore they were pursued; a formal selection procedure was not necessary. On the other hand

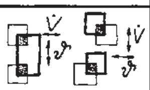
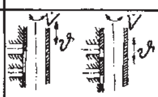
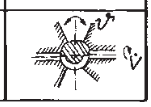
Form of valve \ Type of movement		trans./trans.	trans. /rot.	rot./rot.
		1	2	3
plane plate	A		○	○
cylinder	B	○		○
cone	C	○	○	○
sphere	D	○	○	

Figure 6.36. Classification scheme for solutions to the one-handed mixing tap problem. Movement tangential to the seat face. Two independent movements at an angle for $\dot{V}$ and ϑ

useful working parts and types of movement still had to be analysed. This analysis resulted in the classification criteria shown in Figure 6.35, with the least suitable characteristics indicated with (−). Figure 6.36 shows a classification scheme of possible working principles based on different forms and working movements.

Step 5: Selecting Working Principles

All the working principles shown in Figure 6.36 fulfil the demands of the requirements list and appear to be economic. Hence all three were firmed up into principle solutions.

Step 6: Firming Up into Principle Solution Variants

With the help of further research into possible setting or operating elements that we have not discussed here, the working principles could then be firmed up into principle solution variants and evaluated (see Figures 6.37 to 6.40).

Step 7: Evaluating Principle Solution Variants

In accordance with VDI 2225, this step was taken with the help of an evaluation chart. In addition, evaluation uncertainties and weak spots were examined (see Figure 6.41).

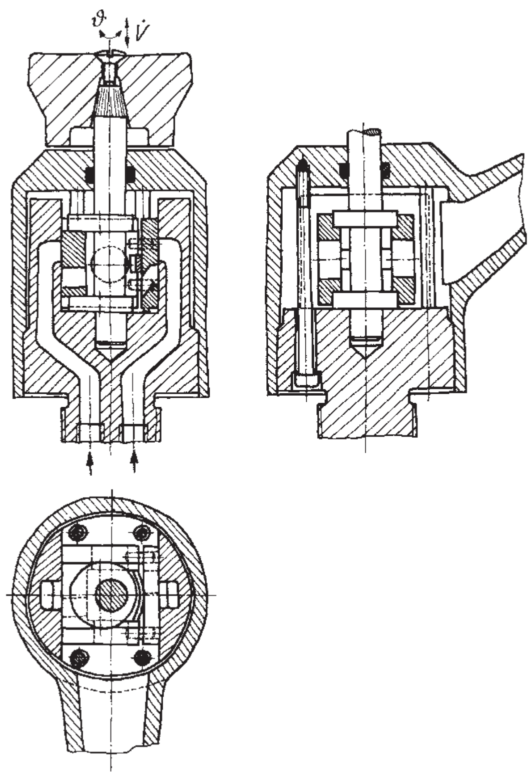


Figure 6.37. One-handed mixing tap, solution variant A: "plate solution with eccentric and pull-and-turn grip"

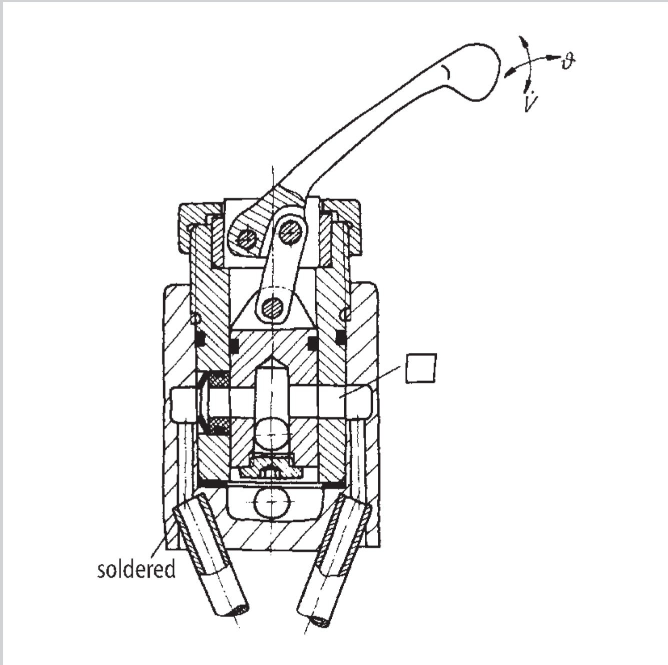


Figure 6.38. One-handed mixing tap, solution variant B: “cylinder solution with lever”

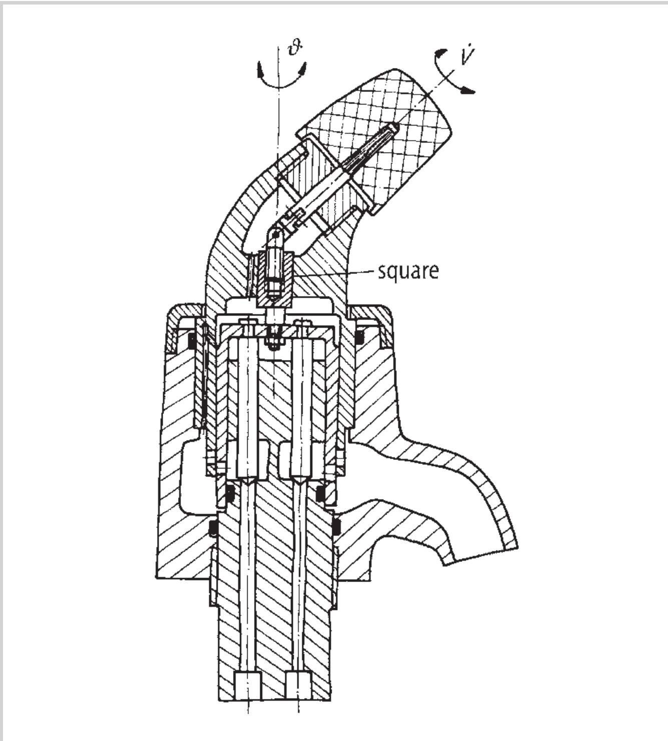


Figure 6.39. One-handed mixing tap, solution variant C: “cylinder solution with end valves and additional sealing”

Thanks to the balanced profile and the discernible improvement possibilities, Solution B (see Figure 6.38) was found to be preferable to all the others. The ball solution D (see Figure 6.40) would only have been considered if further studies into production and assembly problems had been undertaken and had led to positive results.

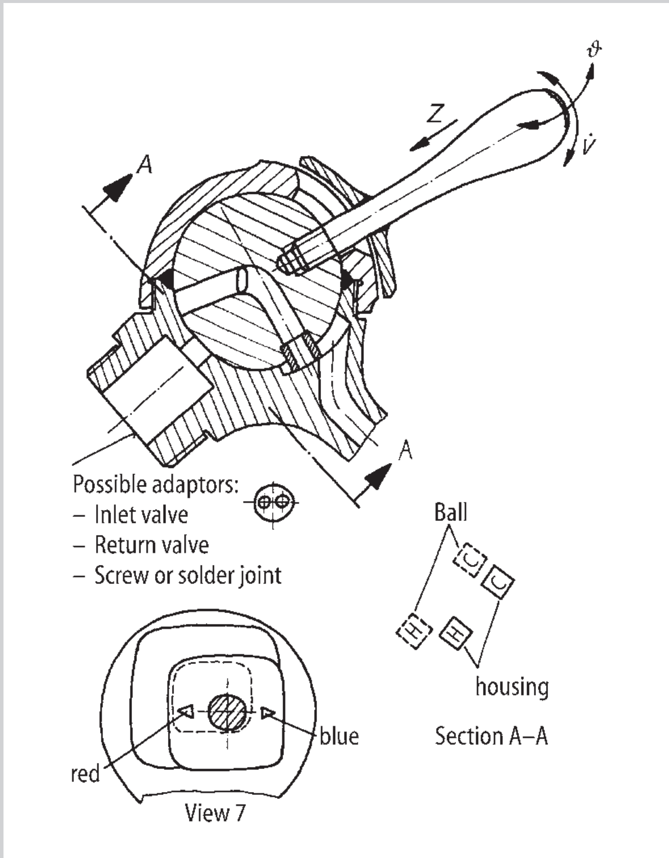


Figure 6.40. One-handed mixing tap, solution variant D: "ball solution"

Step 8: Determining the Next Steps

It was decided to produce dimensional layout drawings of Solution B with improvements to the operating lever with respect to space requirements, easier cleaning and number of parts, and also to improve the level of information for Solution D with a view to reexamining it for final evaluation.

6.6.2 Impulse-Loading Test Rig

Step 1: Clarifying the Task and Setting Up the Requirements List

The second example describes the development of a test rig [6.12]. This test rig was used to investigate the durability of shaft–hub connections subjected to impulsive loads with predefined torques, applied both singly and continuously. Prior to setting up the requirements list, the following questions had to be answered:

- What is meant by impulsive loading?
- Which impulsive torques occur in rotating machines in practice?
- Which stress measurements are possible and useful for keyed connections?

To answer the first two questions, the characteristics of torque–time variations for milling machines, crane drives, agricultural machines and rolling presses were

TH Darmstadt	EVALUATION CHART for: One-handed mixing tap	Page 1

In the order of the checklist headings		P: present variant (P): possible after improvement		A		B		C		D		E		F	
	No.	Evaluation criterion	W	P	(P)	P	(P)	P	(P)	P	(P)	P	(P)	P	(P)
Funct.	1	Reliability of stopping flow without drips	1	1		3		3 I	4	1					
Work Princ.	2	Reliable, reproducible setting (calcium-resistant, few wearing parts)	1	2		3		2 I	3	3					
Embod	3	Low space requirement	1	3↑		2		2		4					
Prod.	4	Few parts	1	1		2 I		1 W		4					
	5	Simple manufacture	1	1		3		2		1?	4				
Assy.	6	Easy assembly	1	2		3		2		2↑ J	3				
Operation	7	Convenient operation, sensitive setting	1	1		3		4		2					
	8	Easy upkeep (easy to clean)	1	4↓		2 I		3		2					
Maint.	9	Simple maintenance (with standard tools, fittings need not be dismantled)	1	1		3		2 W		1? J	3				
	10														
	11														
	12														
	13														
	14														
? Evaluation uncertain ↑ Tendency: better ↓ Tendency: worse		$P_{max} = 4$	Σ	16		24	(26)	21	(23)	(20)	(26)				
		R_t		0,45		0,67		0,58		0,56					
		Ranking		4		1	(1)	2	(3)	3	(2)				

Justification (J), Weak spot (W), Improvement (I) of variant/criterion			
C1	Provide rubber seal		
B4	Simplify lever mechanism		
D6 D9	Indeterminate position of ball during assembly		
B8	Improve with B4		
D9	Attachment of lever		

Decision	Develop solution B with improvement of control elements Solution D: Examine production possibilities, present result in 2 months

Date: 11.10.73	Initials: Dhz	

Figure 6.41. One-handed mixing tap: evaluation of principle solution variants A, B, C, D

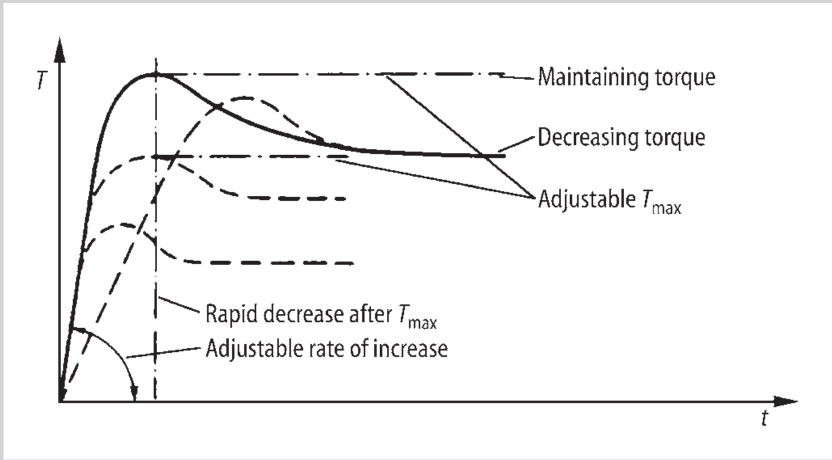


Figure 6.42. Setting magnitudes for an impulsive torque: rate of increase, magnitude and duration

TU Berlin		REQUIREMENT LIST for: Impulse-loading test ring	Part: 1 Page: 1
Changes	D W	Requirements	Responsible
		Geometry:	
	D	Test connection held in position	
	D	Diameter of shaft to be tested ≤ 100 mm (dimensions of the key according to DIN 6885)	
	D	Hubside load take off variable in axial direction	
		Kinematics:	
	D	Loading applied to stationary shaft	
	D	Oscillating load applied in one sense only	
	W	Sense of loading selectable	
	W	Torque input variable (from hub to shaft or from shaft to hub)	
		Forces:	
	D	Shaft-hub connection subjected to torque only (i.e. free of shear forces and bending moments)	
	D	Maximum torque maintained for at least 3 seconds	
	D	Loading frequency low (reason: chosen measuring principle)	
	W	As little vibration as possible in the shaft-hub-key system	
	D	Torque adjustable up to 15 000 Nm in accordance with the load capacity of a shaft of 100 mm diameter	
	D	Rapid decrease of torque after reaching the maximum should be possible	
	D	Adjustable torque increase dT/dt of up to $125 \cdot 10^3$ Nm/s	
	D	Reproducible torque profile	
	W	Plastic deformation, or even destruction, of the connection should be possible	
		Energy:	
	D	Power consumption ≤ 5 kW/380 Volt	
		Material:	
	W	Shaft and hub: C45	
		Replaces issue of	

Figure 6.43. Requirements list for impulse-loading test rig. After [6.12]

TU Berlin	REQUIREMENT LIST for: Impulse-loading test ring		Part: 1	Page: 2
Changes	**D W**	**Requirements**		**Responsible**
		<u>Signal:</u>		
	D	Quantities to be measured: torque in front of and behind the test connection, surface pressure over the length of the connection and the key		
	D	Quantities to be measured should be recordable		
	W	Accessible measuring locations		
		<u>Safety and Ergonomics:</u>		
	W	Easy operation of the test rig (i.e. quick and easy resetting of the test rig)		
	W	Environmentally-friendly operating principle (little noise, dirt, vibration ...)		
		<u>Production and Control:</u>		
	D	One-off production of all parts		
	D	Quality of shaft-hub-connection according to DIN 6885 (as far as described in this standard), otherwise according to standards for shaft ends on drives, electric motors etc: DIN 748, Parts 2 and 3		
	W	Production of the test rig in own workshops		
	W	Use bought-out and standard parts wherever possible		
		<u>Assembly and Transport:</u>		
	W	Test rig with small dimensions and low weight		
	W	No special foundation		
		<u>Operation and Maintenance:</u>		
	W	Parts subjected to wear should be few and simple		
	W	Preferably free of maintenance		
		<u>Costs:</u>		
		Manufacturing costs ~ 20 000 DM (see research proposal)		
		<u>Schedule:</u>		
	D	Concept phase finished by July 1973 Concept phase finished by 20 July 1973		Mr. Militz
		Replaces issue of		

Figure 6.43. (continued)

obtained from the literature. A maximum rate of torque increase of $\mathrm{d}T/\mathrm{d}t = 125 \times 10^3$ Nm/s was selected. The torque–time graph shown in Figure 6.42 was used to establish the necessary parameters to vary.

These requirements, along with others, were documented in the requirements list shown in Figure 6.43. They were classified according to the checklist in Figure 6.22.

Step 2: Abstracting to Identify the Essential Problems

Following the recommendations in Section 6.2.3, the requirements list was abstracted. The results are shown in Table 6.2.

Table 6.2. Abstraction and problem definition on the basis of the requirements list shown in Figure 6.43

Results from Steps 1 and 2

- Diameter of shaft to be tested $\leq 100\,\text{mm}$
- Hubside load take off variable in axial direction
- Loading applied to stationary shaft
- Pure torque loading: adjustable up to 15 000 Nm
- Maximum torque maintained for at least 3 seconds
- Rapid decrease of torque possible
- Maximum torque increase $\mathrm{d}T/\mathrm{d}t$ of 125×10^3 Nm/s
- Reproducible torque profile
- Quantities $T_{\text{in front}}$, T_{behind} and p measurable

Results from Step 3

- Loading of the shaft-hub-key connection adjustable regarding torque magnitude, torque increase, torque holding time and torque decrease
- Check torque and loading with shaft stationary

Results from Step 4

- Adjustable dynamic torque to be applied when testing the specimen
- Measurements of input load levels and of stresses and strains should be possible

Result from Step 5

- "Apply dynamically changing torque while at the same time measuring load levels, stresses and strains"

Step 3: Establishing Function Structures

Establishing the function structure initially involved formulating the overall function, which was extracted directly from the problem statement, see Figure 6.44.

In this example, the essential subfunctions result from the energy flow and, for the measurements, from the signal flow:

- Transform input energy into load (torque)
- Transform input energy into auxiliary energy for the control functions
- Store energy for the impulsive action
- Control load energy and magnitude
- Change load magnitude
- Guide load energy
- Apply load to specimen, i.e. its working surface
- Measure load
- Measure specimen stresses

Setting up the function structure in a stepwise manner resulted in different arrangements and, by adding and removing individual subfunctions, several func-

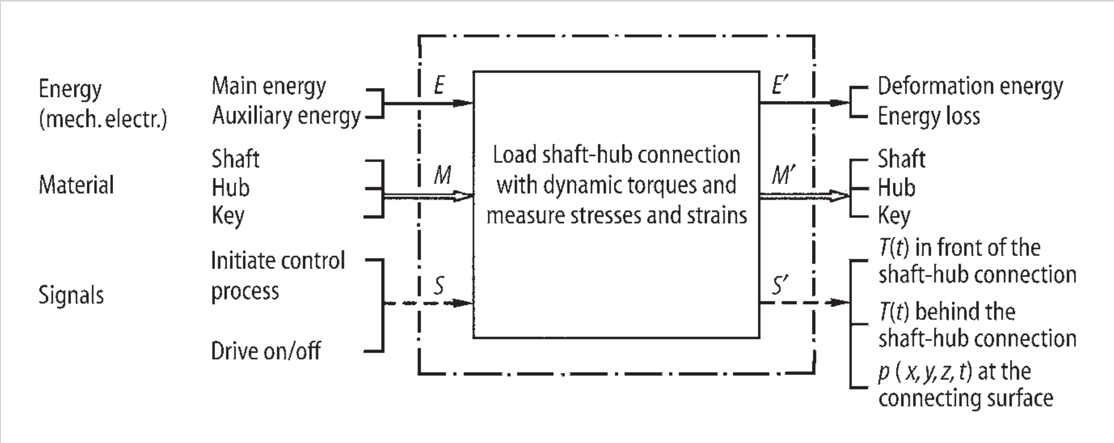


Figure 6.44. Overall function of the impulse-loading test rig

tion structure variants were produced. Figure 6.45 shows these variants in the order in which they appeared. At this stage, the measuring functions do not appear to determine the concept. Variant 4 was chosen to search for solutions because it contained all of the subfunctions of the equally promising Variant 5.

Step 4: Searching for Working Principles

To find working principles, the following methods discussed in Section 3.2 were applied:

- Conventional methods: literature search and analysis of an existing test rig
- Intuitive methods: brainstorming
- Discursive methods: systematic search with the help of classification schemes using types of energy, working movements and working surfaces, as well as the use of a catalogue on varying forces.

To combine the working principles that were found, a classification scheme was produced (see Figure 6.46). For reasons of space, only the most important subfunctions and working principles are shown. Those principles that were clearly unsuitable were either rejected early on or crossed out in the classification scheme. Timely rejection is important in order to minimise subsequent effort.

Step 5: Combining Working Principles

The working principles were combined based on the classification scheme shown in Figure 6.46. Figure 6.47 shows the seven possible combinations (variants) in accordance with the selected function structure variants 4 and 5. The sequences of the subfunctions differ in parts from those of the function structure variants.

Step 6: Selecting Suitable Combinations

A preselection is recommended when a large number of combinations (working structures) have been generated before firming up is attempted (see Section 6.4.3).

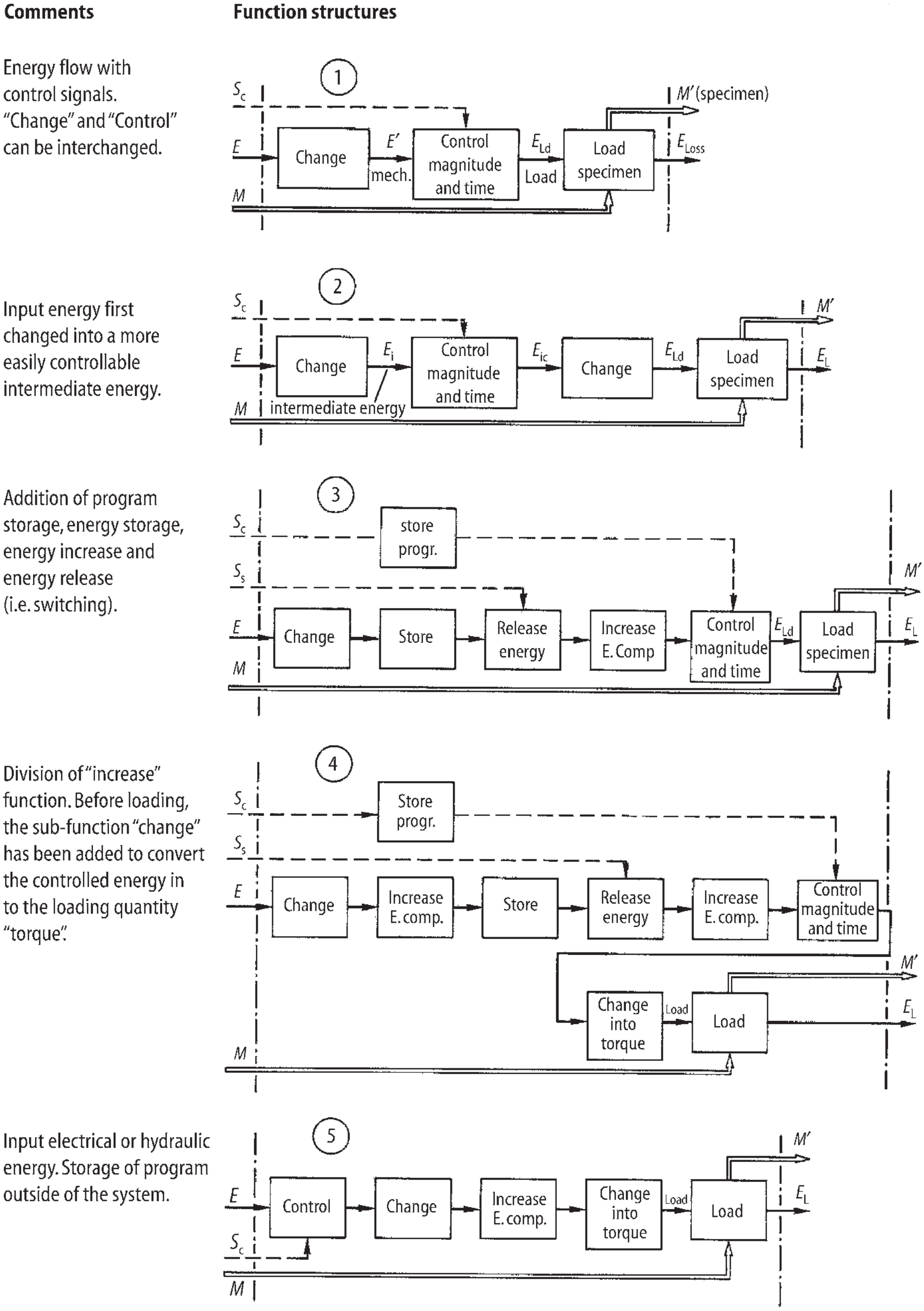


Figure 6.45. Stepwise development of function structure variants

This reduces effort by rejecting less suitable combinations as early as possible. After using the procedure presented in Section 3.3.1, four out of seven combinations appeared promising (see Figure 6.48), but had to be firmed up further to allow for more precise evaluation.

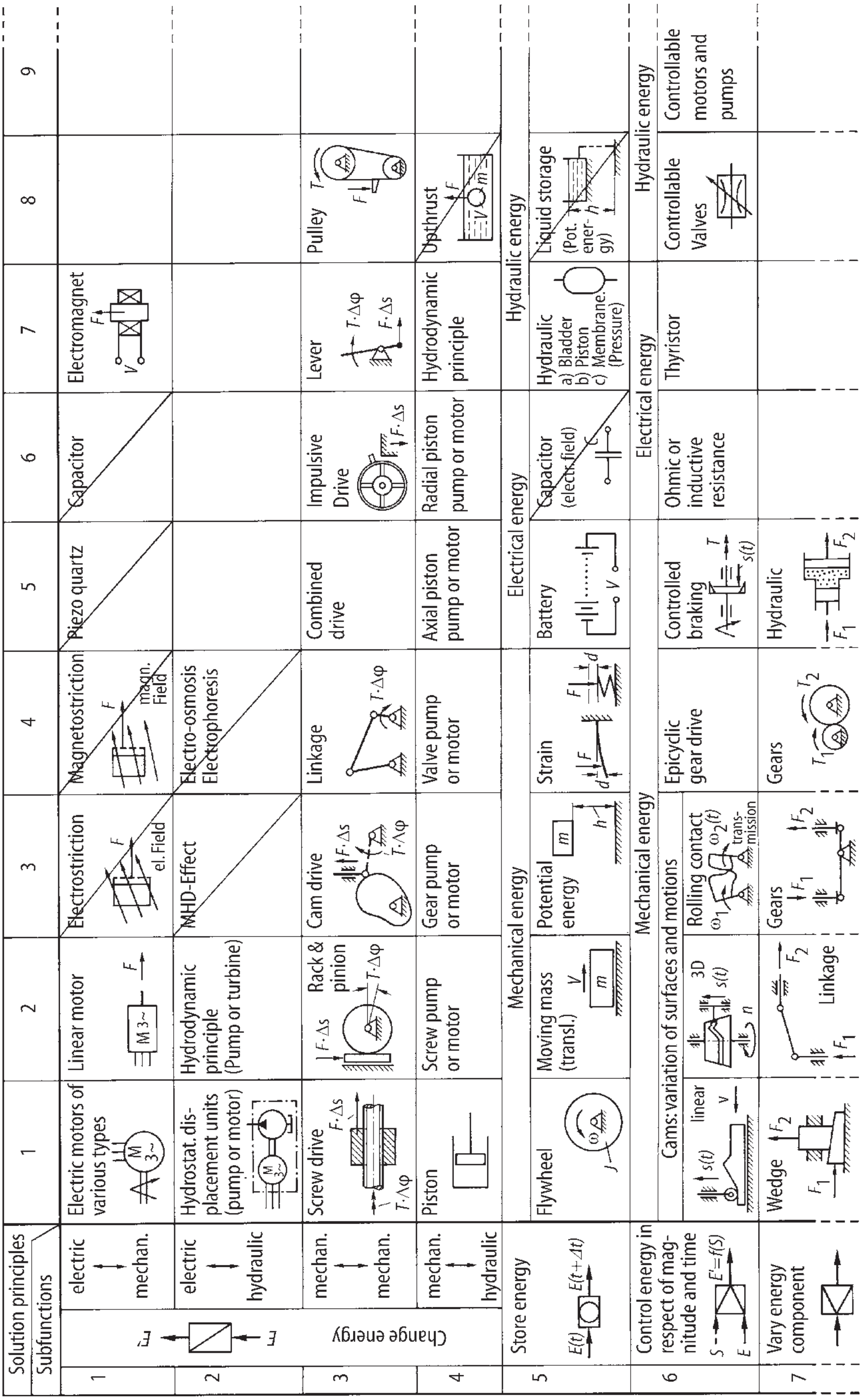


Figure 6.46. Extract from a classification scheme for an impulse-loading test rig

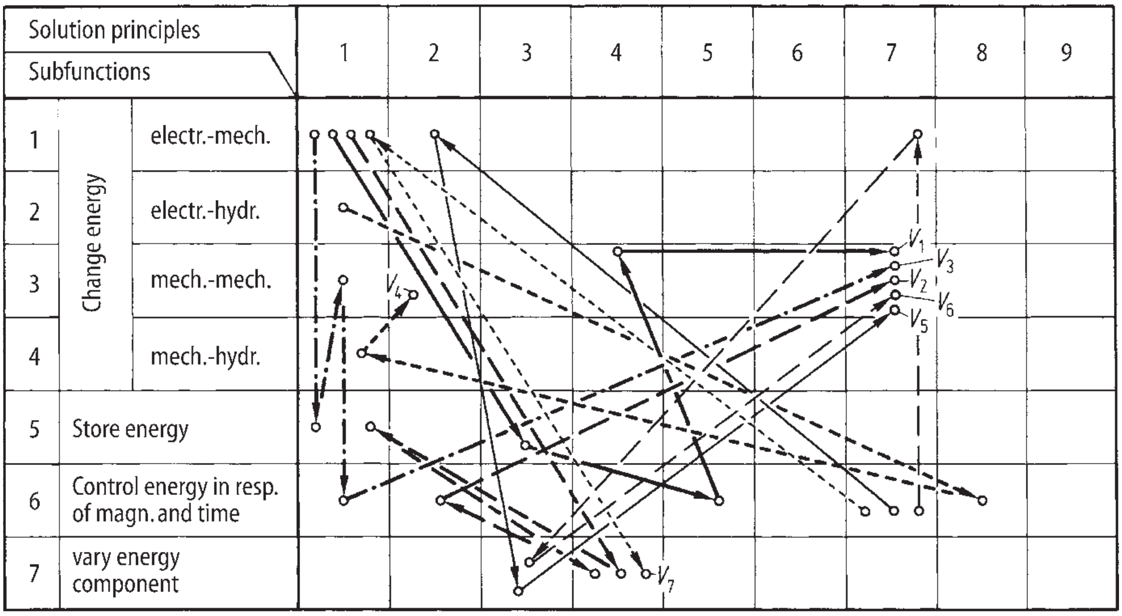


Figure 6.47. Combination scheme showing seven combinations of solution principles in accordance with Figure 6.46.
Variant 1: 1.1 – 5.3 – 6.5 – 3.4 – 3.7;
Variant 2: 1.1 – 7.4 – 5.1 – 7.4 – 6.2 – 3.7;
Variant 3: 1.1 – 5.1 – 3.1 – 6.1 – 3.7;
Variant 4: 2.1 – 6.8 – 4.1 – 3.2;
Variant 5: 6.7 – 1.2 – 7.3 – 3.7;
Variant 6: 6.7 – 1.7 – 7.3 – 3.7;
Variant 7: 6.7 – 1.1 – 7.4

Step 7: Firming Up into Principle Solution Variants

To allow a confident decision to be made about the most suitable principle solution (concept) variant, the selected working structures have to be developed to a state that allows evaluation. This requires that suitable concept drawings such as those shown in Figures 6.49 to 6.52 are produced. Rough sketches often do not provide sufficient detail to assess how well proposals fulfil their functions.

Rough calculations or model tests can be useful at this stage. As an example, calculations will now be made for the cylindrical cam drive used to control the impulsive torque and also the required moment of inertia of the flywheel (energy store) for concept variant V_2.

Can the cylindrical cam shown in Figure 6.53 produce the required torque increase of $\mathrm{d}T/\mathrm{d}t = 125 \times 10^3$ Nm/s and the maximum torque of $T_{\max} = 15 \times 10^3$ Nm?

Calculation steps:

- Time needed to reach the maximum torque at the required rate:

$$\Delta t = \frac{15 \times 10^3}{125 \times 10^3} = 0.12 \text{ s}$$

- Force at the end of the loading lever:

$$F_{\max} = \frac{T_{\max}}{l} = \frac{15 \times 10^3}{0.85} = 17.6 \times 10^3$$

TU Berlin

SELECTION CHART

for Impulse-loading test rig

Part: 1 Page: 1

Solution variants (Sv) evaluated by

SELECTION CRITERIA

- (+) Yes
- (–) No
- (?) Lack of information
- (!) Check requirements list

DECISION

Mark solution variants (Sv)

- (+) Pursue solution
- (–) Eliminate solution
- (?) Collect information (re-evaluate solution)
- (!) Check requirements list for changes

Enter solution variant (Sv):

- A: Compatibility assured
- B: Fulfils demands of requirements list
- C: Realisable in principle
- D: Within permissible costs
- E: Incorporates direct safety measures
- F: Preferred by designer's company
- G: Adequate information

Sv		A	B	C	D	E	F	G	Remarks (Indications, Reasons)	DECISION
V1	1	+	?	+	+	?	–	–	Layout of controlable brakes problematic	+
V2	2	+	+	+	+	+	+	+		+
V3	3	+	+	+	+	+	+	+		+
V4	4	+	+	+	?	+	–	+	Hydraulics not yet applied	+
V5	5	+	?	+	–	+	–	–	No experience with linear motors	–
V6	6	+	?	+	?	+	–	?	Power demand of magnet too great	–
V7	7	+	?	+	–	+	–	?	No experience with thyristor control	–
	8									
	9									
	10									
	11									
	12									
	13									
	14									
	15									
	16									
	17									
	18									
	19									
	20									

Date: 15. 7. 73 Initials:

Figure 6.48. Selection chart for the seven combinations in Figure 6.47

The loading lever is treated as a weak cantilever spring with the end moving through a distance of $h = 30\,\text{mm}$ with a force of F_{max} in such a way that the permissible bending stress is not exceeded.

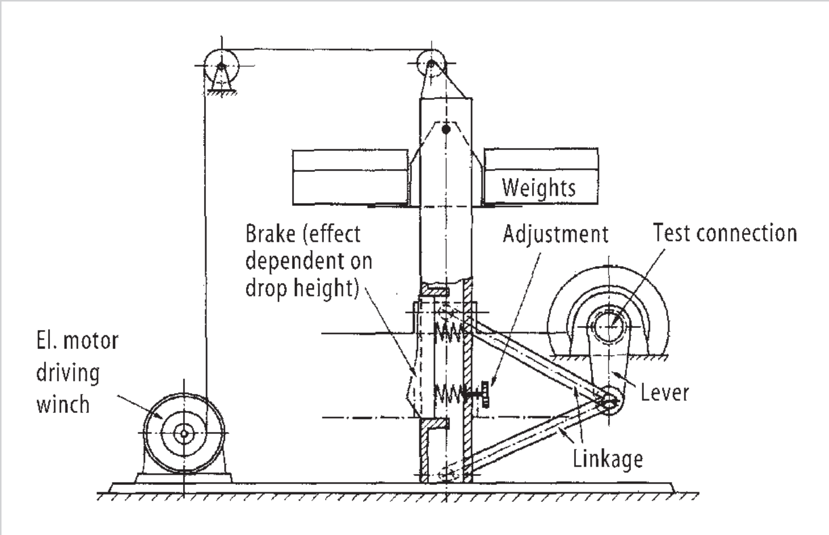


Figure 6.49. Concept variant V_1

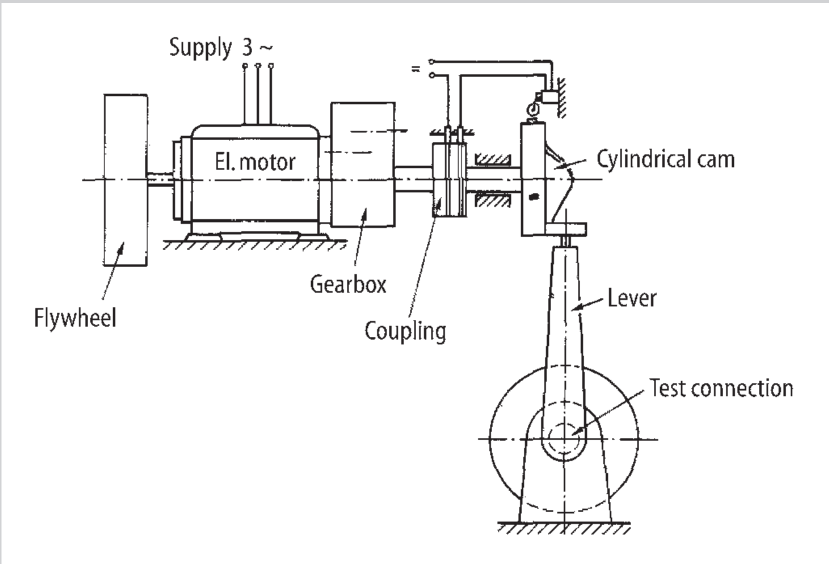


Figure 6.50. Concept variant V_2

- Tangential velocity of the cylindrical cam:

$$v_x = v_y = \frac{h}{\Delta t} = \frac{30}{0.12} = 250 \text{ mm/s}$$

- Angular velocity and rpm of cylindrical cam:

$$\omega = \frac{0.25}{0.125} = 2.0 \text{ rad/s}; \; n = \frac{60\omega}{2\pi} = 19 \text{ rev/min}$$

- Period of revolution:

$$t_\mathrm{r} = \frac{2\pi}{\omega} = 3.14 \text{ s}$$

Since the switching times of the electromagnetically operated clutches used to connect and disconnect the cam drive are in the region of a few tenths of a second, there

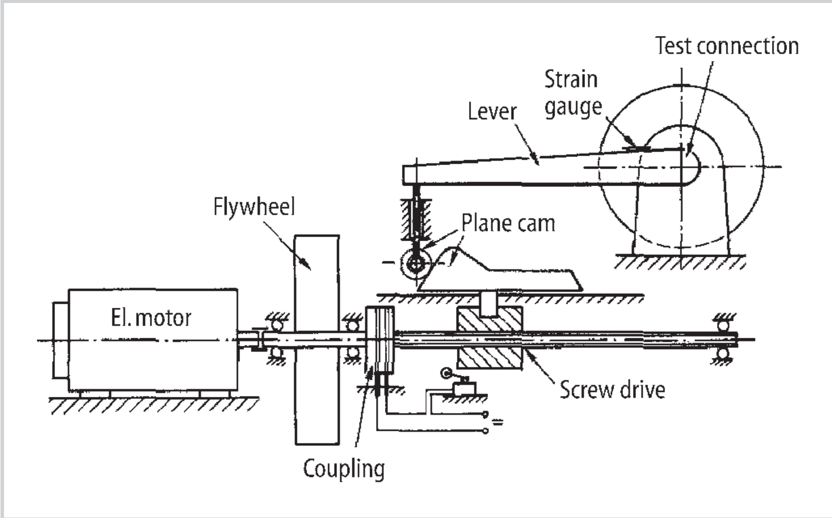


Figure 6.51. Concept variant V_3

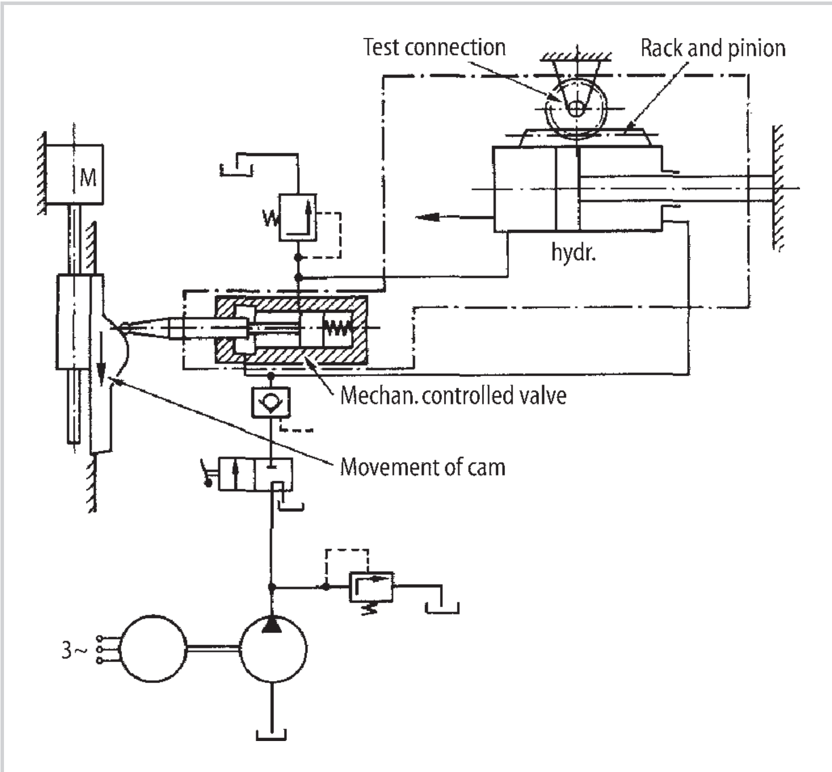


Figure 6.52. Concept variant V_4

should be no problem with applying this principle. The magnitude of, and rate of increase in, the impulse torque loading can be altered by means of interchangeable cams and also by varying the period of revolution.

Steps for estimating the flywheel's moment of inertia:

- The estimate of the energy needed for the impulse (and hence of the energy to be stored) is based on the assumption that all load-carrying parts are elastically deformed.

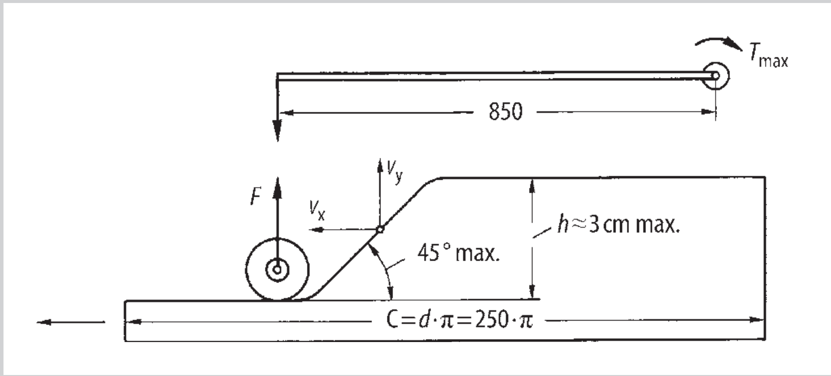


Figure 6.53. Development of cylindrical cam

Stored energy at maximum impulse torque loading:

$$E_{max} = \frac{1}{2} F_{max} \cdot h = 260 \text{ J}$$

This amount of energy is needed in the time interval $\Delta t = 0.12$ s.

- Flywheel dimensions:
  Selected maximum rpm, $n_{max} = 1200$ rev/min; $\omega \mathrel{\hat{=}} 126$ rad/s.
  For flywheel dimensions $r = 0.2$ m and $w = 0.1$ m, the flywheel mass $m_f = 100$ kg, and moment of inertia $J_f = \frac{1}{2} m_f \cdot r^2 = 2 \text{ kg m}^2$.
  Stored energy of flywheel:

$$E_f = \frac{1}{2} J_f \cdot \omega^2 = 159 \times 10^2 \text{ J}$$

- Rotational speed after the impulse:

$$E_{after} = E_f - E_{max} = 15\,640 \text{ J}$$

$$\omega_{after} = \sqrt{\frac{2E_{after}}{J_f}} = 125 \text{ rad/s}; \; n_{after} = 1190 \text{ rev/min}$$

The drop in rpm is therefore very low, and so a motor with a small output is all that is needed.

Step 8: Evaluating Principle Solution Variants

The four variants that were selected in Step 6 and firmed up in Step 7 are evaluated using Cost–Benefit Analysis (see Section 3.3.2).

Important wishes in the requirements list provide a series of evaluation criteria of varying complexity. These are assessed and elaborated with the help of the checklist shown in Figure 6.22. Next, a hierarchical classification (objectives tree) is drawn up to facilitate closer identification and better assignment of the weighting factors and the parameters of the variants. Figure 6.54 shows an objectives tree for the test rig. Its lowest objective level provides the evaluation criteria entered into the table shown in Figure 6.55.

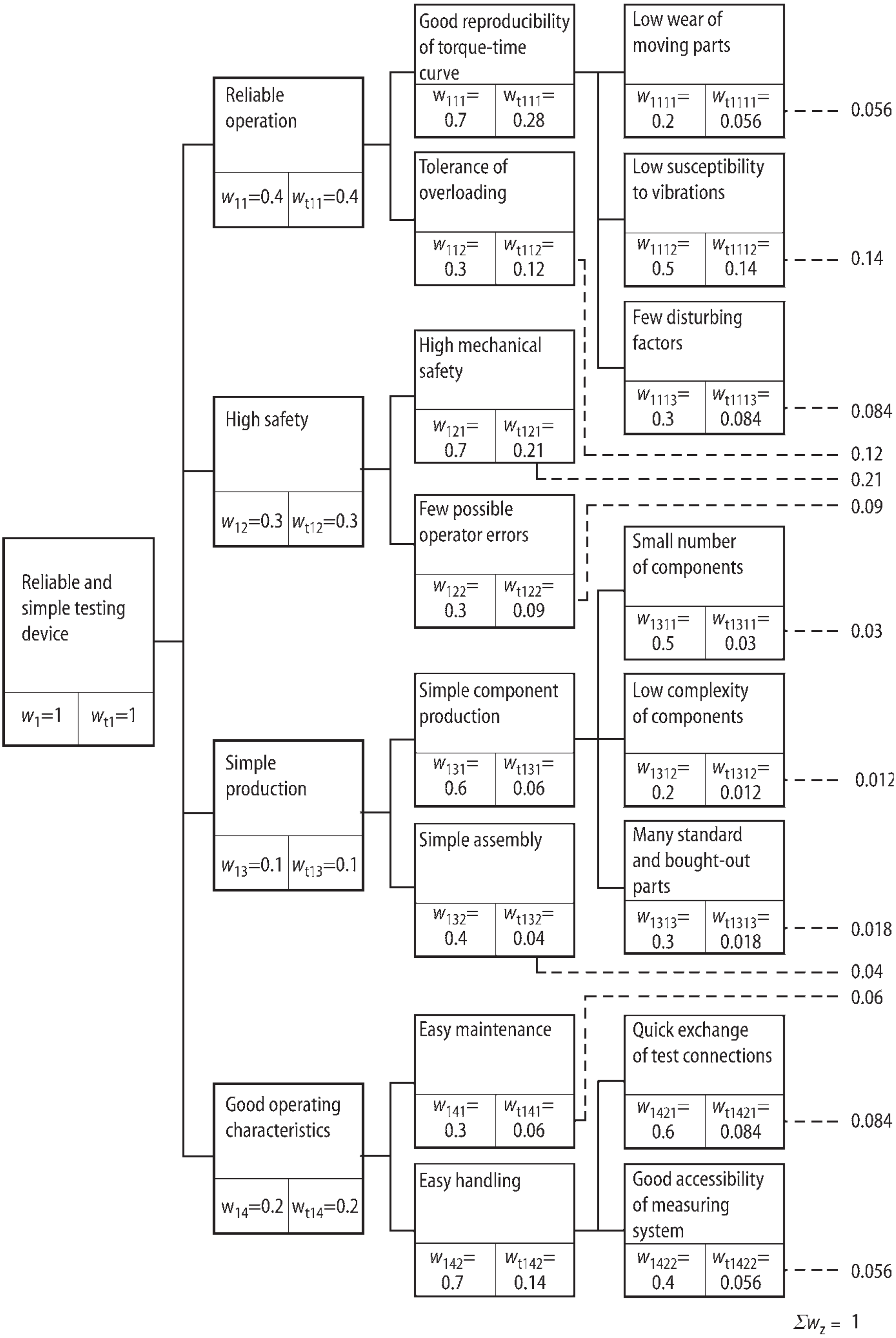


Figure 6.54. Objectives tree for impulse-loading test rig

Evaluation criteria			Parameters		Variant V_1			Variant V_2			Variant V_3			Variant V_4		
No.		Wt.		Unit	Magn. m_{i1}	Value V_{i1}	Weighted value WV_{i1}	Magn. m_{i2}	Value V_{i2}	Weighted value WV_{i2}	Magn. m_{i3}	Value V_{i3}	Weighted value WV_{i3}	Magn. m_{i4}	Value V_{i4}	Weighted value WV_{i4}
1	Low wear of moving parts	0.056	Amount of wear	–	high	3	0.168	low	6	0.336	average	4	0.224	low	6	0.336
2	Low susceptibility to vibrations	0.14	Natural frequencey	$\frac{1}{s}$	410	3	0.420	2370	7	0.980	2370	7	0.980	< 410	2	0.280
3	Few disturbing factors	0.084	Disturbing factors	–	high	2	0.168	low	7	0.588	low	6	0.504	(average)	4	0.336
4	Tolerance of overloading	0.12	Overload reserve	%	5	5	0.600	10	7	0.840	10	7	0.840	20	8	0.960
5	High mechanical safety	0.21	Expected mechanical safety	–	average	4	0.840	high	7	1.470	high	7	1.470	very high	8	1.680
6	Few possible operator errors	0.09	Possibilities of operator errors	–	high	3	0.270	low	7	0.630	low	6	0.540	average	4	0.360
7	Small number of components	0.03	No. of components	–	average	5	0.150	average	4	0.120	average	4	0.120	low	6	0.180
8	Low complexity of components	0.012	Complexity of components	–	low	6	0.072	low	7	0.084	average	5	0.060	high	3	0.036
9	Many standard and bought-out parts	0.018	Proportion of standard and bought-out components	–	low	2	0.036	average	6	0.108	average	6	0.108	high	8	0.144
10	Simple assembly	0.04	Simplicity of assembly	–	low	3	0.120	average	5	0.200	average	5	0.200	high	7	0.280
11	Easy maintenance	0.06	Time and cost of maintenance	–	average	4	0.240	low	8	0.480	low	7	0.420	high	3	0.180
12	Quick exchange of test connections	0.084	Estimated time needed to exchange test connections	min	180	4	0.336	120	7	0.588	120	7	0.588	180	4	0.336
13	Good accessibility of measuring systems	0.056	Accessibility of measuring system	–	good	7	0.392	good	7	0.392	good	7	0.392	average	5	0.280
		$\sum w_i = 1.0$				$OV_1=51$ $R_1=0.39$	$OWV_1=3.812$ $WR_1=0.38$		$OV_2=85$ $R_2=0.65$	$OWV_2=6.816$ $WR_2=0.68$		$OV_3=78$ $R_3=0.60$	$OWV_3=6.446$ $WR_3=0.64$		$OV_4=68$ $R_4=0.52$	$OWV_4=5.388$ $WR_4=0.54$

Figure 6.55. Evaluation of the four principle solution variants for the impulse-loading test rig

It appears that variant V_2 has the highest overall value and the best overall rating. However, variant V_3 follows close behind. For the detection of weak spots, a value profile was drawn (see Figure 6.56). The profile shows that variant V_2 is well-balanced with respect to all of the important evaluation criteria. With a weighted rating of 68%, variant V_2 thus represents a good principle solution (concept) with which to start the embodiment design phase, during which the identified weak spots have to be addressed (see Section 7.7).

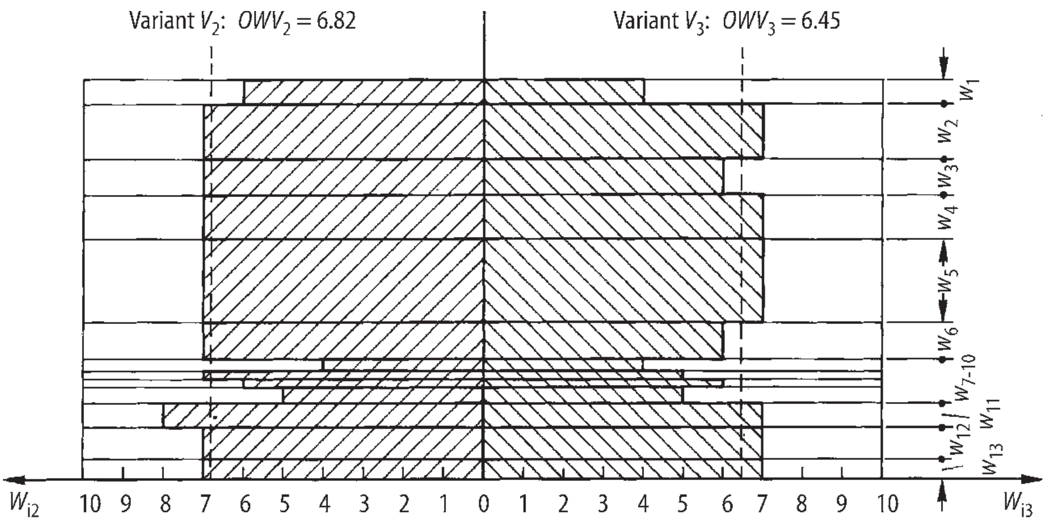


Figure 6.56. Value profile for detection of weak spots

7 Embodiment Design

Embodiment design is the part of the design process in which, starting from the principle solution or concept of a technical product, the design is developed in accordance with technical and economic criteria and in the light of further information, to the point where subsequent detail design can lead directly to production (see Section 4.2).

The draft guideline VDI 2223: *Systematic Embodiment of Technical Products* [7.295] builds on recommendations from the fourth German edition of this book along with other sources. In doing so, it presents a generally established systematic procedure for embodiment design.

7.1 Steps of Embodiment Design

Having elaborated the principle solution during the conceptual phase, the underlying ideas can now be firmed up. During the embodiment phase at the latest, designers must determine the overall layout design (general arrangement and spatial compatibility), the preliminary form designs (component shapes and materials) and the production processes, and provide solutions for any auxiliary functions. During all of this, technological and economic considerations are of paramount importance. The design is developed with the help of scale drawings, critically reviewed, and subjected to a technical and economic evaluation.

In many cases several embodiment designs are needed before a definitive design appropriate to the desired solution can emerge.

In other words, the *definitive layout* must be developed to the point where a clear check of function, durability, production, assembly, operation and costs can be carried out. Only when this has been done is it possible to prepare the final production documents.

Unlike conceptual design, embodiment design involves a large number of corrective steps in which analysis and synthesis constantly alternate and complement each other. This explains why the familiar methods underlying the search for solutions and evaluation must be complemented with methods facilitating the identification of errors (design faults) and optimisation. The collection of information on materials, production processes, repeat parts and standards involves considerable effort.

The embodiment process is complex in that:

- many actions must be performed simultaneously
- several steps must be repeated at a higher level of information
- additions and alterations in one area have repercussions on the existing design in other areas.

Because of this, it is not always possible to draw up a strict plan for the embodiment design phase. However, it is possible to suggest a general approach with main working steps, see Figure 7.1. Particular problems may demand deviations and subsidiary steps, which can rarely be predicted precisely. The approach has to be planned to match the problem at hand, realising that further modifications will have to be made. Basically, the process will proceed from the qualitative to the quantitative, from the abstract to the concrete, and from rough to detailed designs. It is important to make provision for checks and, if necessary, for corrections.

1. Starting with the principle solution, and using the requirements list, the first step is to identify those *requirements that have a crucial bearing* on the embodiment design:
   - size-determining requirements, such as output, throughput, size of connectors, etc.
   - arrangement-determining requirements, such as direction of flow, motion, position, etc.
   - material-determining requirements, such as resistance to corrosion, service life, specified materials, etc.

   Requirements such as those based on safety, ergonomics, production, assembly and recycling involve special embodiment considerations, which may affect the size, arrangement, and selection of materials (see Sections 7.2 to 7.5).
2. Next, the *spatial constraints* determining or restricting the embodiment design must be identified (for instance clearances, axle positions, installation requirements, etc.).
3. Once the embodiment-determining requirements and spatial constraints have been established, a rough layout, derived from the concept, is produced with the emphasis on the overall embodiment-determining *main function carriers*, that is, the assemblies and components fulfilling the main functions. The following subsidiary questions must be settled, with due regard paid to the principles of embodiment design (see Section 7.4):
   - Which main functions and function carriers determine the size, arrangement and component shapes of the overall layout (for instance, the blade profiles in turbomachines or the flow area of valves)?
   - Which main functions must be fulfilled by which function carriers jointly or separately (for instance, transmitting torque and allowing for radial movement by means of a flexible shaft or by means of a stiff shaft plus a special coupling)? This step is similar to division into realisable modules, as shown in Figure 1.9.

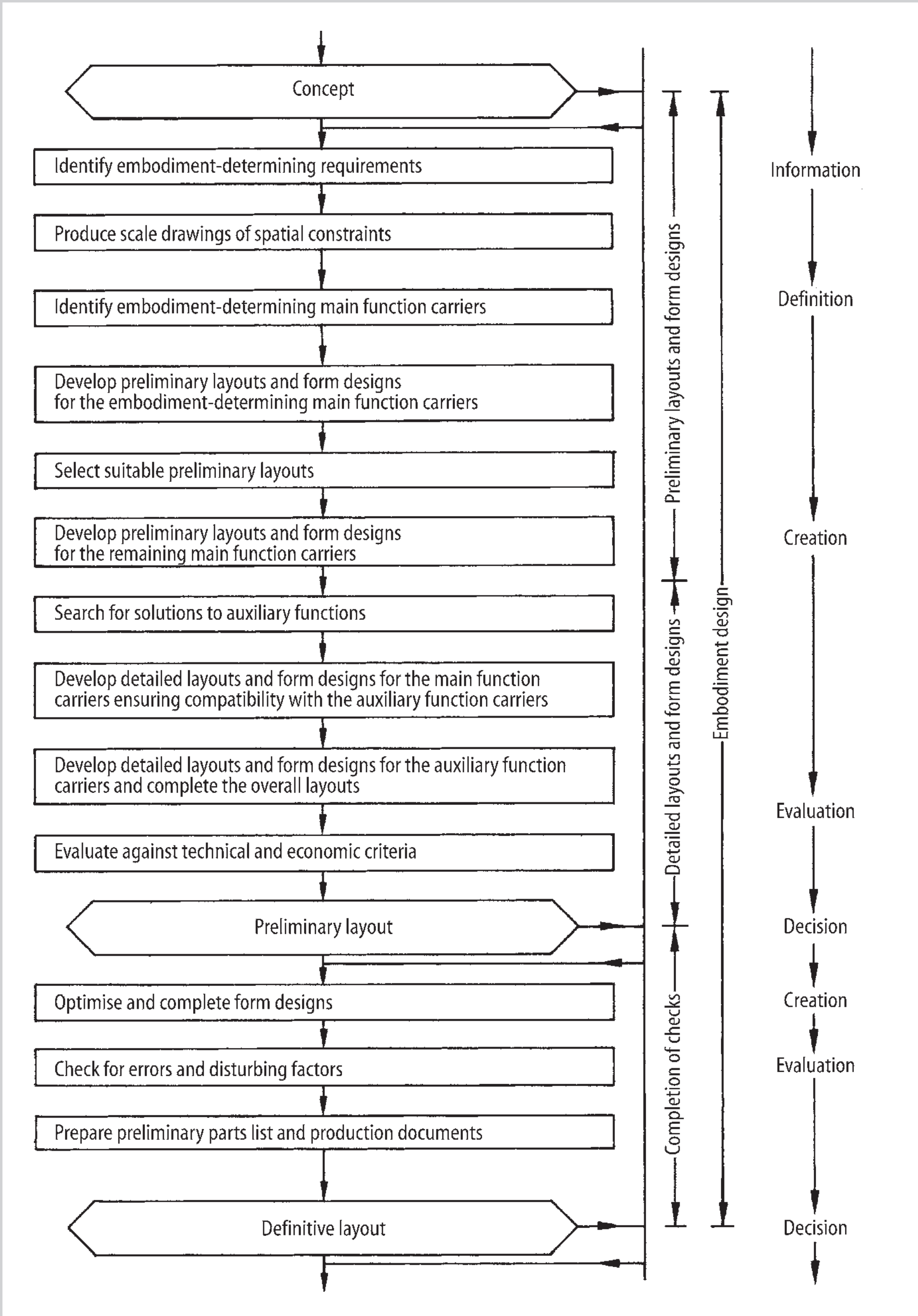


Figure 7.1. Steps of embodiment design

4. Preliminary scale layouts and form designs for the embodiment-determining main function carriers must be developed; that is, the general arrangement, component shapes and materials must be determined provisionally. To that end, it is advisable to work systematically through the items under the heading "layout" in the checklist shown in Figure 7.3. The result must meet the overall spatial constraints and then be completed so that all of the relevant main functions are fulfilled (for instance by specifying the minimum diameters of drive shafts, provisional gear ratios, minimum wall thicknesses, etc.). Known solutions or existing components (repeat parts, standard parts, etc.) must be shown in simplified form. It may be useful to start working on selected areas only, combining these into preliminary layouts later.
5. One or more suitable *preliminary layouts* must be selected in accordance with the procedure described in Section 3.3.1 (modified if necessary) by considering the relevant items in the checklist shown in Figure 7.3.
6. Preliminary layouts and form designs must now be developed for the remaining main function carriers that have not yet been considered because known solutions exist for them or they are not embodiment-determining until this stage.
7. Next, determine which essential *auxiliary functions* (such as support, retention, sealing and cooling) are needed and, where possible, *exploit known solutions* (such as repeat parts, standard parts, catalogue solutions). If this proves impossible, *search for special solutions*, using the procedures already described in Section 3.2 and Chapter 6.
8. *Detailed layouts and form designs for the main function carriers* must now be developed in accordance with the embodiment design rules and guidelines (see Sections 7.3 to 7.5), paying due attention to standards, regulations, detailed calculations and experimental findings, and also to the problem of compatibility with those auxiliary functions that have been realised. If necessary, divide into assemblies or areas that can be elaborated individually.
9. Proceed to develop the *detailed layouts and form designs for the auxiliary function carriers*, adding standard and bought-out parts. If necessary, refine the design of the main function carriers and combine all function carriers into overall layouts.
10. *Evaluate* the layouts against technical and economic criteria (see Section 3.2.2). If a particular project requires several concepts to be put in more concrete form prior to evaluation, then the embodiment process must not, of course, be pursued beyond what the evaluation of the variants demands. Depending on the circumstances, it is thus possible, in some cases, to take a decision just as soon as the main function carriers have reached the preliminary layout stage, while in other cases the decision will have to be deferred until after a great deal of detail design. In either event, all of the designs to be compared must be at the same level of embodiment, since no reliable evaluation is possible otherwise.

11. Fix the preliminary overall layout. The overall layout describes the complete construction structure (see Figure 2.13) of the system or product being designed.
12. Optimise and complete the form designs for the selected layout by *eliminating the weak spots* that have been identified during the course of the evaluation. If it should prove advantageous, repeat the previous steps and adopt suitable subsolutions from less favoured variants.
13. *Check* this layout design for *errors* (design faults) in function, spatial compatibility, etc. (see Figure 7.3), and for the effects of *disturbing factors*. Make what improvements may be needed. The achievement of the objectives with respect to cost (see Chapter 11) and quality (see Chapter 10) must be established at this point at the latest.
14. Conclude the embodiment design phase by preparing a preliminary *parts list* as well as a preliminary *production and assembly documents*.
15. Fix the *definitive layout* and pass on to the detail design phase.

The representation of the spatial constraints and the embodiment is now generally obtained by creating a full 3-D digital model. Irrespective of whether a 2-D or 3-D representation is used [7.213]:

- the function and type of the objects must be shown
- the positions of and the necessary space for the objects must be recognisable through characteristic dimensions, e.g. the overall dimensions, which can be used to check the overall spatial compatibility and assembly operations.

When 2-D CAD systems or drawing boards are still used simplifications, such as those proposed by Lüpertz [7.174], could be used.

In the embodiment phase, unlike the conceptual phase, it is not necessary to lay down special methods for every individual step, however the following recommendations might prove useful.

The *search for solutions* for auxiliary functions and other subsidiary problems should be based either on the procedure described in Chapter 3, but simplified as far as possible, or else directly on catalogues. Requirements, functions and solutions with appropriate classifying criteria have already been elaborated.

The *embodiment* (layout and form designs) of the function carriers should be based on the checklist (see Figure 7.3) and involves reference to the principles of mechanics and structures, and to materials technology. It calls for calculations ranging from the simplest through to complex differential equations and finite element analyses. For these calculations, the reader is referred to the literature listed in Section 7.5.1, and for even more complex calculations to the domain specific literature. In some cases it might be necessary to build prototypes or to undertake specific tests.

In the elaboration of embodiment designs, many details have to be clarified, confirmed and optimised. The more closely they are examined, the more obvious it becomes as to whether the right solution concept has been chosen. It may appear that this or that requirement cannot be met, or that certain characteristics of the chosen concept are unsuitable. If this is discovered during the embodiment phase, it is advisable to re-examine the procedure adopted in the conceptual phase, for no embodiment design, however perfect, can hope to correct a poor concept. This is equally true of the working principles applicable to the various subfunctions. However, even the most promising concept can cause difficulties in embodiment and detail design. This often happens because various features were originally treated as subordinate or as not in need of further clarification. Attempts to solve these subproblems compel designers to reiterate the appropriate steps while retaining the selected working structure and overall arrangement.

Experience with the proposed approach for embodiment design has confirmed its basic validity, but has also revealed the following important points [7.211]:

- If prior research has been undertaken or embodiment variants already exist, the step of producing preliminary embodiments can often be left out.
- Preliminary embodiments can always be left out when only detailed improvements are required.
- The solutions for auxiliary functions usually influence the preliminary embodiment of the main function carriers, so working on these solutions must not be left until too late in the process.
- A characteristic of successful designers is that they continuously check and monitor their actions to identify direct and indirect effects.

Many products are not developed from scratch, but are developments or improvements of existing ones that take into account new requirements, new knowledge and experiences. Experience has shown that it is useful to start by analysing the failures and disturbing factors for an existing solution (see Sections 10.2 and 10.3) and, based on that analysis, to develop a new requirements list (see Figure 7.2). The result of the clarified task will show whether a new working structure—a new principle solution—is required, or whether it is sufficient to modify the existing embodiment. It is possible to start at many different places within the overall approach. In some cases a new product can be produced by making improvements to the details. In other cases, tests of the existing or modified modules may be necessary. The required steps in the overall approach must be selected appropriately.

To sum up, embodiment design involves a flexible approach with many iterations and changes of focus. The individual steps have to be selected and adapted to the particular situation. The ability to organise one's own approach while paying due regard to the fundamental links between the steps and the recommendations we provide is important (see Section 2.2.1).

In embodiment design, the rules and principles elaborated in Sections 7.2 to 7.5 should be followed. Because of the fundamental importance of the identification of errors (design faults) in several of the steps, the reader is referred to Chapter 10 in particular.

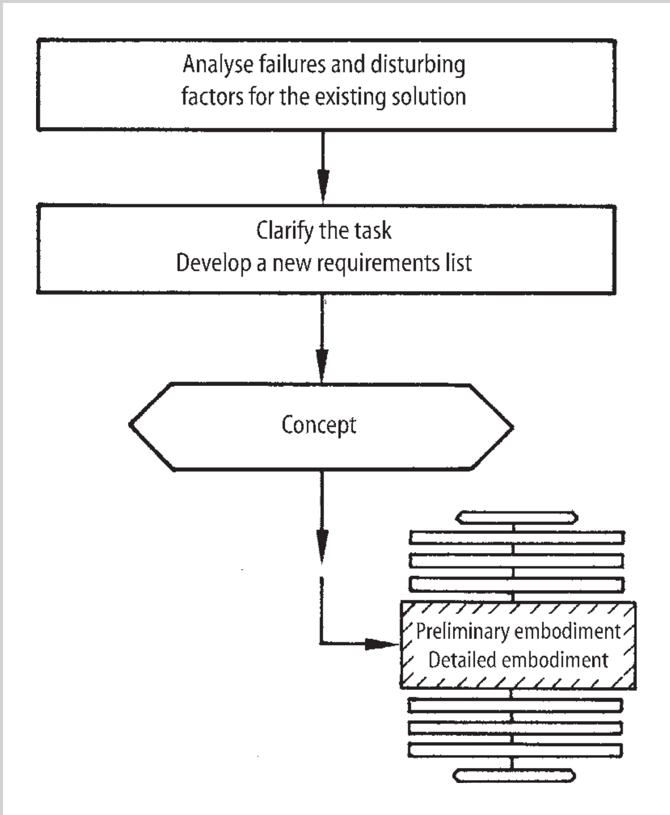


Figure 7.2. Embodiment design phase based on the development of an existing solution. Which of the steps shown in Figure 7.1 needs to be completed follows from an analysis of failures and disturbing factors

7.2 Checklist for Embodiment Design

Embodiment design is characterised by *repeated deliberation and verification* (see Section 7.1). Every embodiment design is an attempt to fulfil a given function with appropriate layout, component shapes and materials. The process starts with preliminary scale layouts based on a rough analysis of spatial requirements, and proceeds to consider safety, ergonomics, production, assembly, operation, maintenance, recycling, costs and schedules.

In dealing with these factors, designers will discover a large number of interrelationships, so that their approach must be progressive as well as iterative (verification and correction). Notwithstanding this double character, however, the approach must always be such as to allow the speedy identification of those problems that must be solved first.

The checklist shown in Figure 7.3 has been derived from the general objectives and constraints discussed in Section 2.1.7. Although the factors are interrelated, this checklist presents them in a useful procedural order and provides designers with a systematic check on each one. The checklist thus not only provides a strong mental impetus, but also ensures that nothing essential is forgotten.

All in all, continuous reference to the headings will help designers to develop and test their progress in a systematic and time-saving way. Each heading should be examined in turn, regardless of its interrelationship with the rest.

Headings	Examples
Function	Is the stipulated function fulfilled? What auxiliary functions are needed?
Working principle	Do the chosen working principles produce the desired effects and advantages? What disturbing factors may be expected?
Layout	Do the chosen overall layout, component shapes, materials and dimensions provide: adequate durability (strength) permissible deformation (stiffness) adequate stability freedom from resonance unimpeded expansion acceptable corrosion and wear with the stipulated service life and loads?
Safety	Have all the factors affecting then safety of the components, of the function, of the operation and of the environment been taken into account?
Ergonomics	Have the human–machine relationships been taken into account? Have unnecessary human stress or injurious factors been avoided? Has attention been paid to aesthetics?
Production	Has there been a technological and economic analysis of the production processes?
Quality control	Can the necessary checks be applied during and after production or at any other required time, and have they been specified?
Assembly	Can all the internal and external assembly processes be performed simply and in the correct order?
Transportt	Have the internal and external transport conditions and risks been examined and taken into account?
Operation	Have all the factors influencing the operation, such as noise, vibration, handling, etc. been considered?
Maintenance	Can maintenance, inspection and overhaul be easily performed and checked?
Recycling	Can the product be reused or recycled?
Costs	Have the stipulated cost limits been observed? Will additional operational or subsidiary costs arise?
Schedules	Can the delivery dates be met? Are there design modifications that might improve the delivery situation?

Figure 7.3. Checklist for embodiment design

The actual sequence is no indication of the relative importance of the various headings, but ensures a systematic approach. For instance, it would be futile to deal with assembly problems before ascertaining if the required performance or minimum durability is ensured. The checklist thus provides a consistent scrutiny of embodiment design and one that is easily memorised.

7.3 Basic Rules of Embodiment Design

The following basic rules apply to all embodiment designs. If they are ignored problems are introduced and breakdowns or accidents may occur. They underlie nearly all of the steps listed in Section 7.1. When used in conjunction with the checklist (see Figure 7.3) and with the design fault identification methods (see Chapter 10), they also provide essential assistance with selection and evaluation.

The *basic rules* of clarity, simplicity and safety are derived from the general objectives set out in Section 2.1.7, that is:

- fulfilment of the technical function
- economic feasibility
- individual and environmental safety.

The literature contains numerous rules of, and guidelines for, embodiment design [7.168, 7.180, 7.198, 7.205]. On closer analysis it appears that clarity, simplicity and safety are fundamental to all of them and are important prerequisites for a successful solution.

Clarity—that is, clarity of function or lack of ambiguity of a design—facilitates reliable prediction of the performance of the final product and in many cases saves time and costly analyses.

Simplicity generally guarantees economic feasibility. A smaller number of components and simple shapes are produced more quickly and easily.

Safety imposes a consistent approach to the problems of strength, reliability, accident prevention and protection of the environment.

In short, by observing these three basic rules, designers can increase their chances of success because they focus attention on, and help to combine, functional efficiency, economy and safety. Without this combination no satisfactory solution is likely to emerge.

7.3.1 Clarity

In what follows we shall be applying the basic rule of clarity to the various headings of the checklist in Figure 7.3.

Function

Within a given function structure, an unambiguous interrelationship between the various subfunctions and the appropriate inputs and outputs must be guaranteed.

Working Principle

The chosen working principle, in terms of the physical effects, must reveal a clear relationship between cause and effect, thus ensuring an appropriate and economical layout.

The chosen working structure, comprising several individual working principles, must guarantee an orderly flow of energy, material and signals. If it does not, undesirable and unpredictable effects such as excessive forces, deformations and wear may ensue.

By paying attention to the deformations associated with a given loading, and also to thermal expansion, designers must make the necessary allowances for possible expansion in a given direction.

The widely used bearing pairs, with a locating and a nonlocating bearing (see Figure 7.4a) have a clearly defined behaviour. The stepped bearing pair (see Figure 7.4b), on the other hand, should be specified only when the expected changes

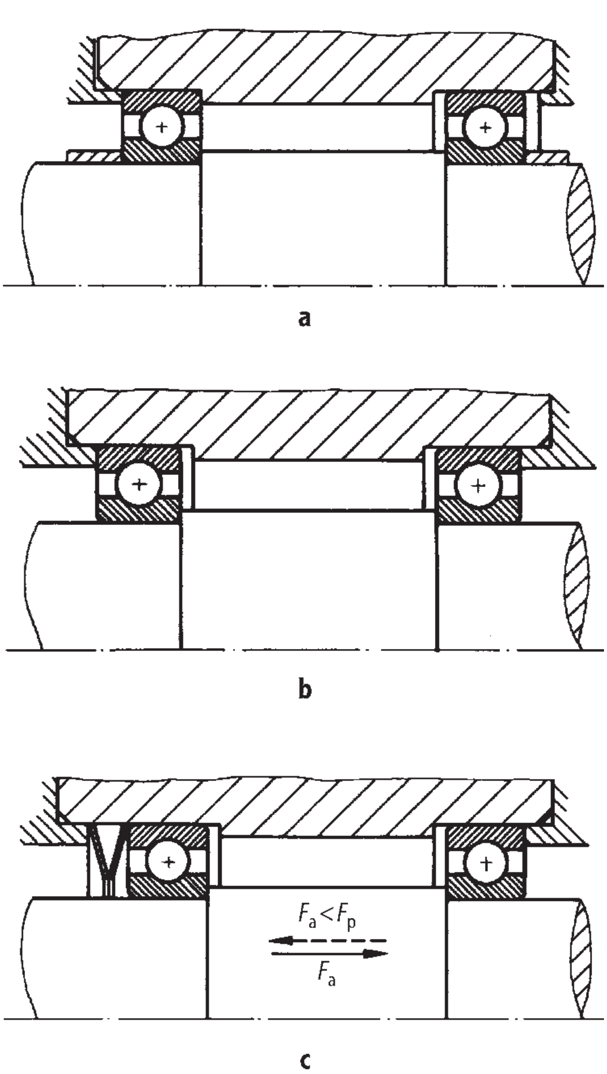


Figure 7.4. Basic bearing arrangements: **a** Locating and nonlocating arrangement: left-hand locating bearing takes up all the axial forces, right-hand sliding bearings permit unimpeded axial movement due to thermal expansion; accurate calculations are possible. **b** Stepped bearing arrangement: the axial loading of the bearings depends on the preload and thermal expansion and cannot be clearly determined; a modification is the "floating arrangement" in which the bearings are provided with axial clearance; in that case, thermal expansion is possible to a limited extent but there is no precise shaft location. **c** Spring-loaded bearing arrangement: here the disadvantages of the stepped bearing arrangement are largely eliminated, though the constantly applied axial load may reduce the bearing life; forces resulting from thermal expansion can be determined by spring force deflection diagrams; the shaft is located precisely provided the axial force F_a acts only towards the right or does not exceed the preloading F_p

in length are negligible or when the resulting play in the bearings is permissible. By contrast, a spring-loaded arrangement, in which the operating axial force F_a does not exceed the pre-load F_p, will permit a clear definition of the force transmission path (see Figure 7.4c).

Combined bearing arrangements often present problems. The combination shown in Figure 7.5a consists of a needle roller bearing which is intended to transmit the radial forces and a ball bearing which is meant to transmit the axial forces. However, this particular arrangement does not clearly define the transmission path for the radial forces, because the inner and outer races of both bearings

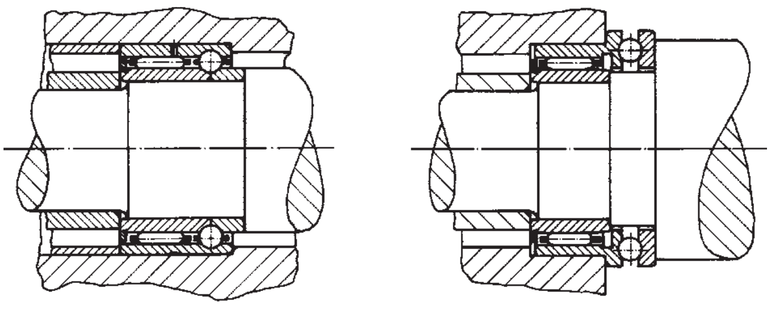


Figure 7.5. Combined rolling-element bearing. **a** Transmission path of radial forces not clear; **b** combined rolling bearing with the same elements as in **a**, but clear identification of the transmission paths of the radial and axial forces

are restrained radially. As a result, the service life cannot be predicted accurately. The arrangement shown in Figure 7.5b, on the other hand, satisfies the clarity rule with similar elements, provided the designer ensures during assembly that the right-hand race has enough radial play, thus making certain that the ball bearing transmits axial forces only.

Double fits conflict with the basic rule of clarity. These occur when a component is supported or guided by two surfaces at the same time, and these surfaces are either on different planes or on different cylindrical sections. In such cases, the surfaces have to be machined separately and will therefore have different dimensions caused by the tolerances. As a consequence, the force flow cannot be predicted clearly and assembly is made more difficult. Even though modern production machines have reduced the problems with tolerances, the lack of clarity will still affect function fulfilment and ease of assembly unless double fits are avoided. Double fits appear in various forms. Figure 7.6 shows examples and corrective measures.

Layout

The layout (general arrangement) and form design (shapes and materials) require a clear definition of the magnitude, type, frequency and duration of loads. If these data are not available, the implementation must be based on reasonable assumptions and the expected service life specified accordingly.

In any case, the embodiment must be such that the loads can be defined and calculated under all operating conditions. No impairment of the function or the durability of a component must be allowed to arise.

Similarly, following the checklist in Figure 7.3, behaviour with respect to stability, resonance, wear and corrosion must be clearly established.

Very often one comes across *double arrangements*, i.e. doubling up working principles for safety's sake, which conflict with the rule of clarity. Thus a shaft–hub connection designed as a interference fit will not have a better load-carrying capacity if it is also provided with a key, as in Figure 7.7. The extra element merely ensures correct positioning in the circumferential sense, but because of the reduction in the area at A, the resulting stress concentration at B and the presence

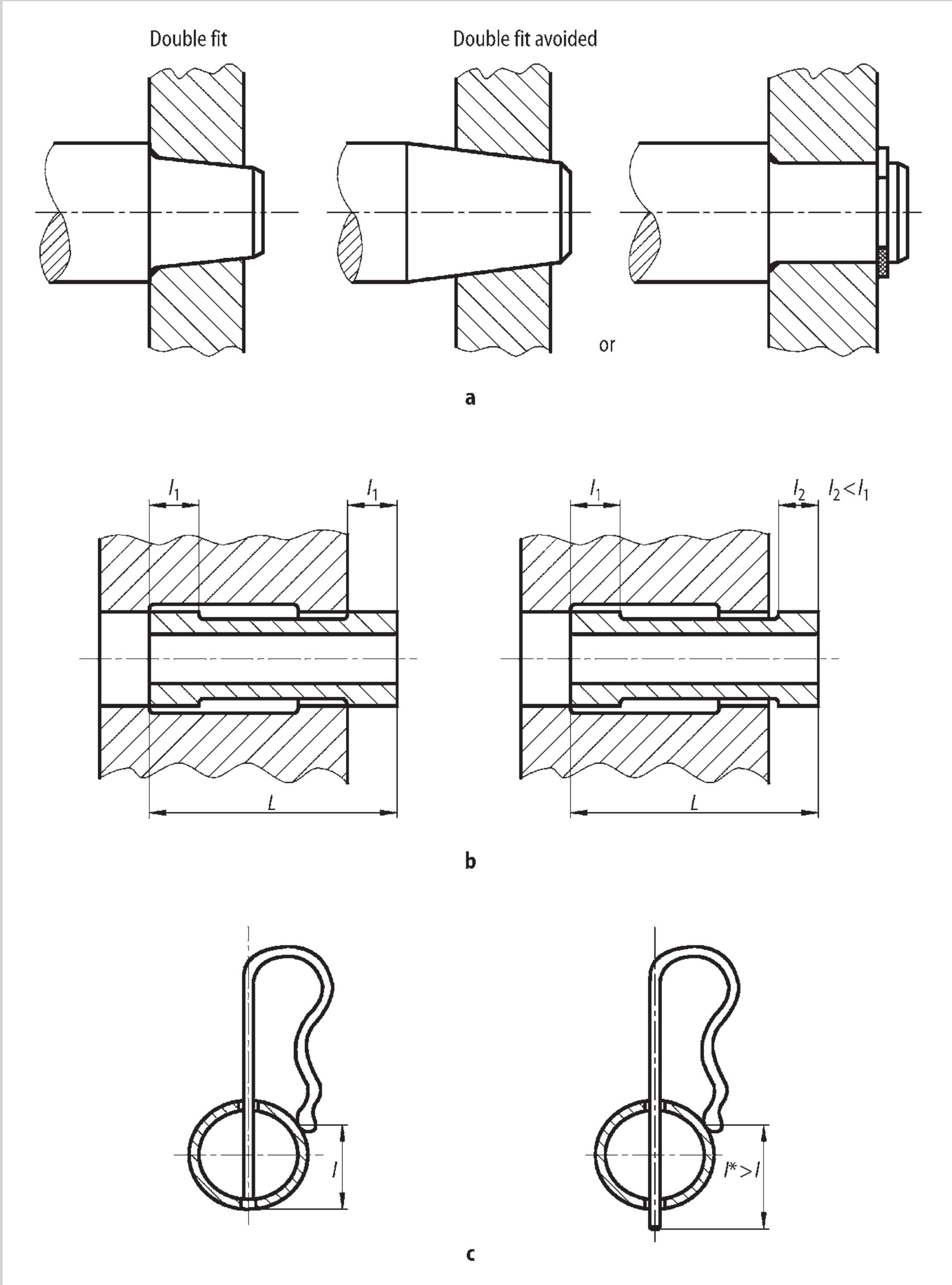


Figure 7.6. Avoiding double fits: **a** Tapered shaft–hub connection with interference (shrink) fit. The simultaneous axial location against the shaft collar and the taper seat creates a double fit: the radial pressure due to the interference fit cannot be determined. The right solution would be to use either a taper without a shaft collar or to use a cylindrical seat with a shaft collar. **b** Supported linear slide using a guiding sleeve in a housing. The simultaneous location of the housing at two points complicates the assembly process. A possible solution is shown in the figure on the right. **c** Spring clip of such a length that the lower end touches the tube at the same time as the pressure point touches the tube. The user will not be able to determine whether the clip is blocked by the tube or whether the spring force has to be overcome. The correct solution is shown in the figure on the right

of complicated and almost incalculable stresses at *C*, it decreases the strength in a drastic and fairly unpredictable manner.

Schmid [7.242] has shown that an axially preloaded taper joint for the transmission of torque requires a spiralling motion when the hub is assembled on the shaft in order to ensure a reliable interference fit, and the use of a key prevents this.

The employment of an interference fit to achieve the maximum torque capacity is only possible by leaving out the key. The solution shown in Figure 7.7 is only acceptable when the correct positioning of the hub relative to the shaft is the crux of the task, in which case a sliding fit is more appropriate.

Figure 7.8 shows a housing adapter for a centrifugal pump which can be used to provide various annulus profiles to fit different blade shapes so that new housings need not be constructed for each case. Unless the intermediate pressure in the gap between the adapter and the housing can be clearly regulated, or some other means of attachment is used, the adapter might travel upwards and damage the blades by rubbing against them.

This is particularly true when similar fits (H7-j6) are chosen for the two locating diameters which are approximately the same size. This is because, depending on production tolerances and working temperatures, gaps may appear, the relative sizes of which are unpredictable and which produce unknown intermediate pressures in the space between the adapter and the housing. The solution shown in Figure 7.8 (detail) ensures, by means of the specially designed connecting passage *A* (which must have a flow area roughly four to five times greater than the maximum gap area that might appear at the upper locating diameter), a clearly definable intermediate pressure, corresponding to the lower inlet pressure of the pump. As a result, the housing adapter is always pressed

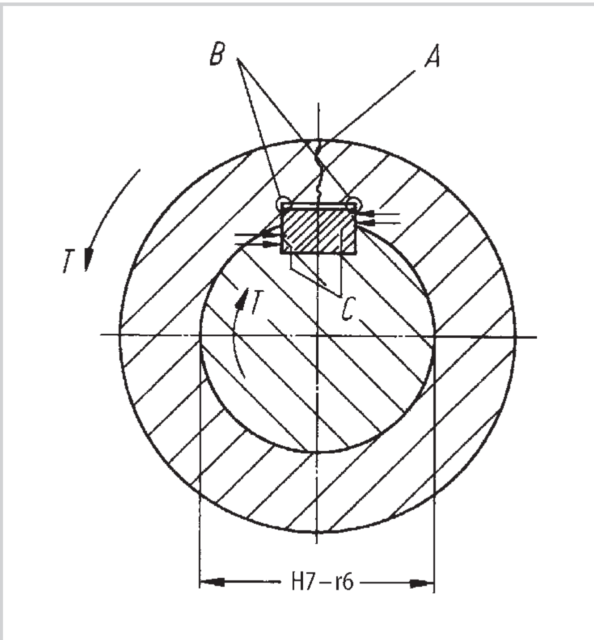


Figure 7.7. Combined shaft–hub connection achieved by means of shrink fit and key: an example of not applying the principle of clarity

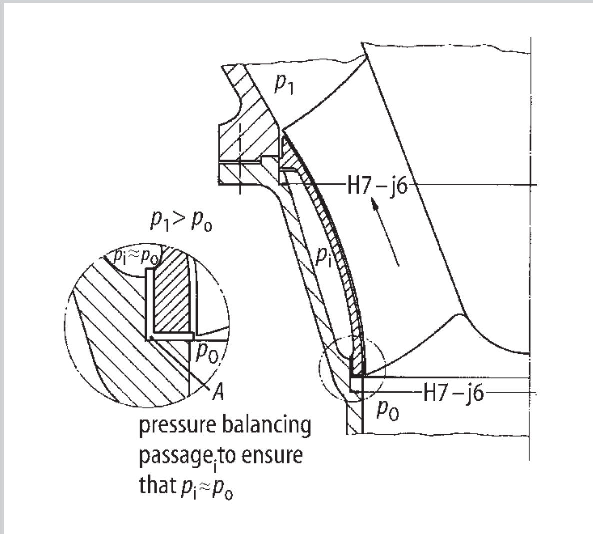


Figure 7.8. Housing adapter in a cooling-water pump

downwards when the pump is in operation, and attachments are only needed as locating aids for assembly and to prevent any tendency of the adapter to rotate.

Serious damage has been reported in gate valves whose operational or loading conditions were not clearly defined [7.130, 7.131]. When closed, gate valves separate, say, two steam pipes and at the same time close off the inside of the valve housing. The result is a self-contained pressure chamber, as shown in Figure 7.9. If condensate has collected in the lower part of the valve housing, and steam appears on the inlet side with the valve closed so that the valve is heated, then the enclosed condensate may evaporate and produce an unpredictable increase in pressure inside the valve housing. The result is either a ruptured housing or serious damage to the housing cover connection. If the latter is self-sealing, serious accidents may ensue since, in contrast to what happens with overloaded bolted flange connections, there is no preliminary leakage and hence no warning. The danger lies in the failure to specify clear operational and loading conditions. Possible remedies are as follows:

- Connect the inner chamber of the gate valve housing to an appropriate steam pipe, operational conditions permitting ($p_{\text{valve}} = p_{\text{pipe}}$)
- Protect the valve housing against excess pressure (p_{valve} restricted)
- Drain the valve housing, thus avoiding collection of condensate ($p_{\text{valve}} \approx p_{\text{external}}$)
- Design valves in such a way as to minimise the housing volume (collection of condensate kept low).

Similar phenomena in welded membrane seals are discussed in [7.206].

Safety

See basic rule in Section 7.3.3.

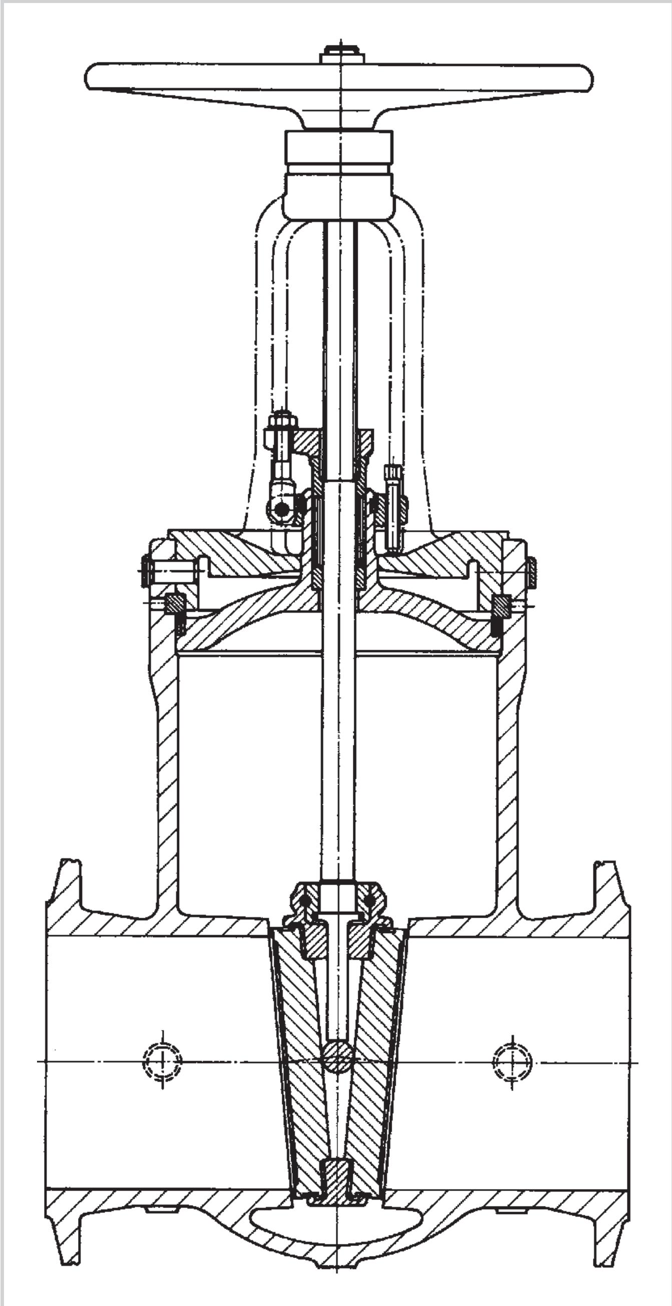


Figure 7.9. Gate valve with relatively large lower collecting area

Ergonomics

In human–machine relationships, correct operation must be ensured via the logical layout of equipment and controls.

Production and Quality Control

These must be facilitated by clear and comprehensive data in the form of product models as well as drawings, parts lists and instructions; and adherence to the prescribed production and quality control procedures.

Assembly and Transport

Much the same is true of assembly and transport. A clear assembly sequence preventing mistakes should be incorporated into the design (see Section 7.5.8).

Operation and Maintenance

Clear installation instructions and the appropriate embodiment design must ensure that:

- the performance is easily checked
- inspection and maintenance involves the smallest possible variety of tools and equipment
- the scope and schedules of inspection and maintenance are defined
- inspection and maintenance can be checked after they have been carried out (see Section 7.5.10).

Recycling

Designers should provide (see Section 7.5.11):

- clear separation of materials that are incompatible with regard to recycling
- clear sequences of assembly and disassembly.

7.3.2 Simplicity

For technical applications, the word "simple" means "not complex", "easily understood" and "easily done".

A solution seems simpler if it can be effected with fewer components, because, for example, the probability of lower production costs, less wear and lower maintenance is then greater. However, this is only true if the arrangement and shapes of the components are kept simple. Hence designers should always aim at the minimum number of components with the simplest shapes [7.168, 7.198, 7.206].

As a rule, however, a compromise has to be made. The fulfilment of a function always demands a certain minimum number of components. Cost efficiency often necessitates a decision between numerous components with simple shapes but with greater overall production effort, and, for example, a single cheaper cast component with the greater uncertainty it may entail in delivery. Simplicity must always be assessed from a holistic perspective—what constitutes "simpler" in individual cases depends on the problem and the constraints.

In what follows we shall be applying the basic rule of simplicity to the various headings of the checklist shown in Figure 7.3.

Function

In principle, only a minimum number and a clear and consistent combination of subfunctions should be pursued when considering the function structure.

Working Principle

In selecting working principles, only those involving a small number of processes and components, that have obvious validity and involve low costs should be taken into consideration.

In the development of the one-handed mixing tap (see Section 6.6.1), several solution principles were proposed. One group (see Figure 6.36) involved the use of only one component to realise two independent adjustments in directions tangential to the valve seat face (types of motion: translation and rotation). The other group (see Figure 6.33), though involving only movements in one direction (normal or tangential to the seat face), required an additional coupling mechanism to convert the two single adjustments into one direction of movement. Quite apart from the fact that, in the second group, the preset temperature is often lost when the tap is shut off, all solutions represented in Figure 6.33 involve a greater design effort than those in the first group. Hence, designers should always begin with a group like that depicted in Figure 6.36.

Layout

Here the simplicity rule requires:

- geometrical shapes which can be analysed simply for strength and stiffness
- symmetrical shapes which provide clearer identification of deformations during production and under mechanical or thermal loads.

In many cases, designers can reduce the work of calculation and experimentation significantly if they try, by means of a simple design, to facilitate the application of basic mathematical principles.

Safety

See under Section 7.3.3.

Ergonomics

The human–machine relationship should also be simple (see Section 7.5.5) and can be significantly improved by means of:

- obvious operating procedures
- clear physical layout
- easily comprehensible signals.

Production and Quality Control

Production and quality control can be simplified, and at the same time made faster and more accurate, if:

- geometrical shapes permit the use of well-established, time-saving methods
- production operations are minimised and have short setting-up and waiting times
- shapes are chosen to facilitate the inspection process.

Leyer, when discussing changes in production methods [7.166], uses the example of a sliding control valve approximately 100 mm long to demonstrate how the replacement of a complicated casting by a brazed product made of geometrically simple turned parts helped to overcome difficulties and paved the way for more economical production. Even though modern casting techniques now allow more intricate shapes to be produced relatively easily, further simplifications might still be expedient (see Figure 7.10). Step 3 helps to simplify the geometrical shape of the central, tubular part. Step 4 (fewer parts) can be taken when the surface areas at right angles to the valve axis need not be retained.

A further example is provided by the one-handed mixing tap discussed earlier. The design of the lever arrangement shown in Figure 7.11 is expensive to make, difficult to clean (slits, open recesses) and not aesthetically pleasing. The one shown in Figure 7.12 is much simpler and also more suitable for longer production runs. The lever, whose end can slide and rotate in a circumferential groove, requires a smaller number of parts and avoids wear in areas that are difficult to readjust. All in all, therefore, this solution is by far the better because it is more economic, easier to clean and looks nicer.

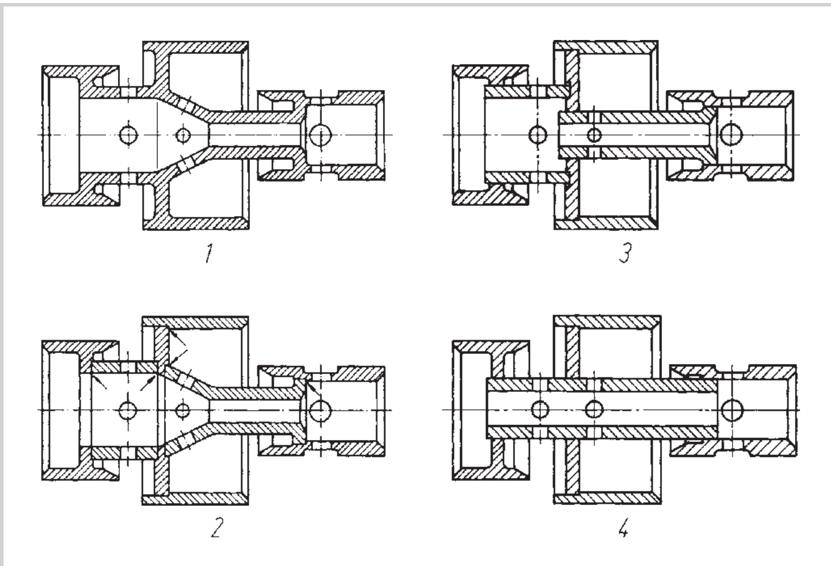


Figure 7.10. Simplification of a sliding control valve: 1 Casting is difficult and expensive; 2 Improvement by splitting into simple, brazed parts; 3 Simplification of central tubular part; 4 Further simplification possibility (1 and 2 after [7.166])

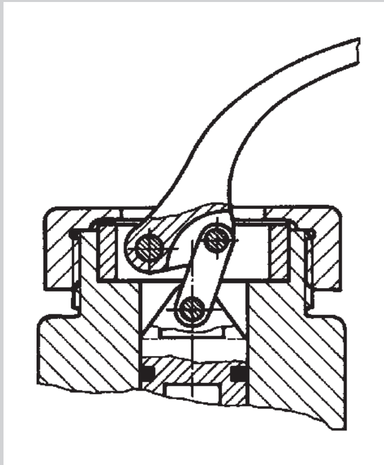


Figure 7.11. Proposed lever arrangement for a one-handed mixing tap with translational and rotational movements

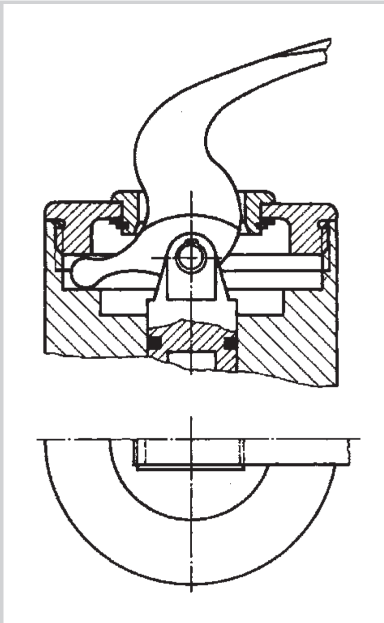


Figure 7.12. Simpler solution with improved embodiment (based on Schulte)

Assembly and Transport

Assembly is simplified—that is, facilitated, speeded-up and rendered more reliable—if:

- the components to be assembled can be identified easily
- the assembly instructions can be followed easily and quickly
- no adjustment has to be repeated
- reassembly of previously assembled components is avoided (see Section 7.5.9).

During assembly, the adjustment ring of a small steam turbine has to be moved vertically and horizontally with the turbine shaft already assembled, in order

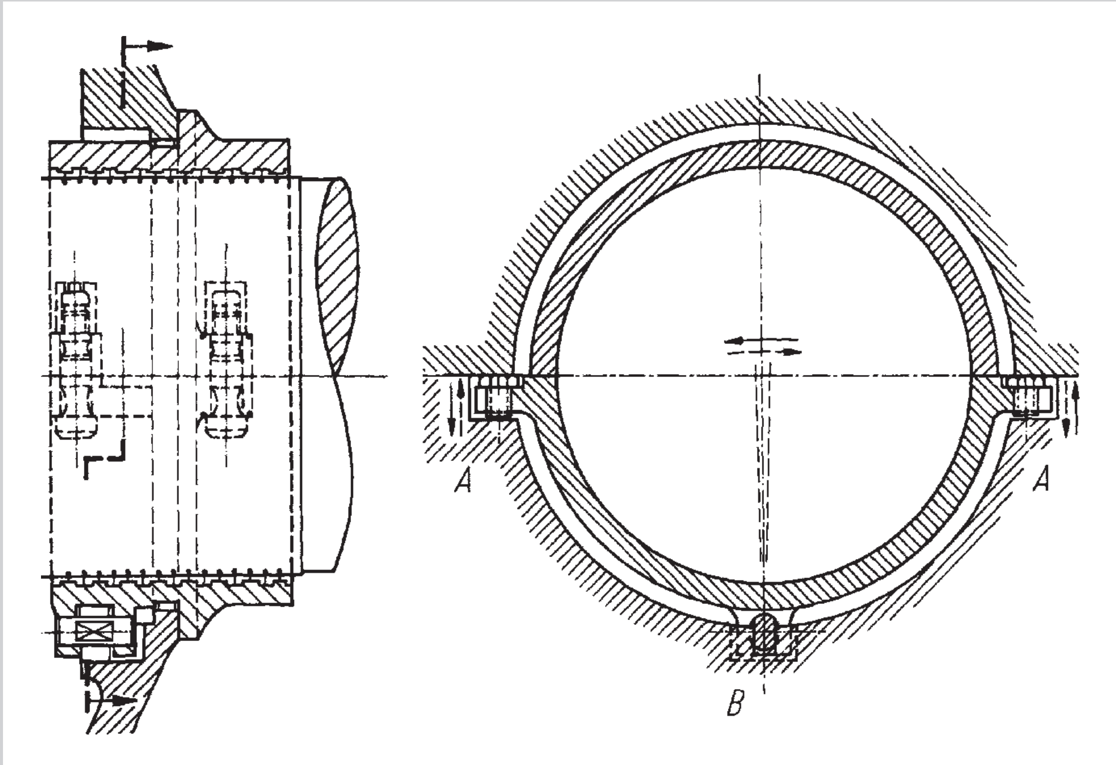


Figure 7.13. Adjustable sealing ring of an industrial steam turbine; adjustments at *A* in the same sense produce vertical movement, adjustments at *A* in the opposite sense produce a rotation about *B* that approximates to a horizontal movement

to ensure uniform clearance around the labyrinth seal. Doing this without having to remove the shaft several times for adjustment poses a problem that can be solved by the design shown in Figure 7.13. The adjustment can be made at the joint by rotating the adjustment screws *A* in the same sense, producing vertical movement only, and by rotation in the opposite sense, producing a tilting movement about pivot *B* that approximates to horizontal movement. The pivot itself must, however, allow for vertical movement during the adjustment and also for radial heat expansion when the turbine is operating. This is achieved with a few easily produced elements with simple shapes. A suitable arrangement of the surfaces, moreover, obviates the need to secure the pivot pin with additional locking elements: it is located in such a way that it can not fall out.

Operation and Maintenance

With respect to operation and maintenance, the simplicity rule means:

- operation must be possible without special or complicated instructions
- the sequence of operations must be clear and simple, and any deviations or faults easily identified
- maintenance must not be awkward, laborious and time-consuming.

Recycling

Simplicity for recycling can be realised by:

- use of recyclable materials
- simple assembly and disassembly processes
- simplicity of the parts themselves (see Section 7.5.11).

7.3.3 Safety

1. Nature and Scope of Safety Measures

Safety considerations affect both the reliable fulfilment of technical functions and also the protection of humans and the environment. Designers have recourse to a safety methodology that, following the German industry standard DIN 31 000 [7.57], includes the following three levels:

- direct safety
- indirect safety
- warnings.

In general, designers should try to guarantee safety by using *direct safety*, that is, by choosing a solution that precludes danger from the outset. Only when this proves impossible should they have recourse to *indirect safety*, in other words, constructing special protective systems [7.58 to 7.60]. *Warnings*, which merely point out dangers and indicate danger areas, can be used to support direct and indirect safety measures by, for example, pointing out special features, obstructions and disturbances. Only as a last resort should warnings be used on their own, and never as an easily implemented safety measure.

In the solution of technical problems, designers are faced with several constraints, not all of which they can hope to overcome simultaneously. They must nevertheless strive to provide a solution that comes nearest to satisfying all the requirements. The strength of an unavoidable safety requirement may, under certain circumstances, put the realisation of the whole project in doubt. A high demand for safety can greatly complicate a design and, by reducing clarity, may even lower the inherent safety of the product. Moreover, safety provisions may also render a product uneconomic and lead to its abandonment.

Such cases, however, are exceptional, because safety and economy generally go hand-in-hand in the long term. This is particularly true of expensive and complex plant and machinery. Only smooth, accident-free and safe operation can ensure long-term economic success. Protection against accidents or damage, moreover, goes hand-in-hand with reliability [7.75, 7.312]. Reliability makes it possible to operate a machine to full capacity, even though poor reliability may not necessarily lead to accidents or damage. All in all, it is therefore advisable to achieve safety by treating direct and indirect safety measures as an integral part of system design.

There are many different ways of applying safety measures in mechanical engineering. Therefore, we consider it necessary to provide some definitions before discussing the measures in detail. The withdrawn German industry standard DIN 31 004 (1979) defined safety as “being free from danger”, a “danger” being a threat for which the type, size and action is known. A dangerous situation is one that can cause damage to persons or things. This DIN standard was replaced in November 1982 by DIN 31 004 Part 1 [7.61]. The basic terms are defined as follows:

Safety	is a state in which the risk is smaller than the risk limit.
Risk limit	is the largest but still acceptable system-specific risk relating to a particular technical process or situation.
Risk	is described by the frequency (probability) and the expected extent of the damage (scope).

Whereas the initial DIN standard defined protection as the limitation of danger in order to prevent damage, the 1982 standard uses the following definition:

Protection	is the reduction of risk by suitable means in order to reduce the frequency of occurrence and/or the extent of damage.

The DIN EN 292 standard [7.57] now uses these terms in a more general way. This development of the standard demonstrates that there is no absolute safety in the sense of complete freedom from danger. In common with many aspects of life, the use of technical systems always involves a certain risk. Safety measures aim to reduce risks to an acceptable level. However, what is acceptable (the risk limit) can only be quantified in a few cases. Now and in the future this limit will be determined by technical knowledge and social standards, and in no small measure by the experience and responsibilities of design engineers.

In the context of safety, it is very important to ensure reliability:

Reliability	is the ability of a technical system to satisfy its operational requirements within the specified limits and for the required life (definition based on [7.75, 7.76]).

It is clear that the reliability of individual components of a machine or the machine itself, as well as the reliability of any protective systems and devices, are important requirements for safety. Without state-of-the-art quality that ensures reliability, protective measures are of doubtful value.

One measure of reliability is the operational availability of a technical system.

Availability	is the percentage of time the system is available for operation compared to the maximum possible time or compared to a particular target time.

Safety concerns the following areas (see Figure 7.14):

Operational safety	is the limitation of danger (reducing risk) during the operation of technical systems in order to prevent damage to the systems themselves and their immediate environment, such as the workplace, neighbouring systems, etc.

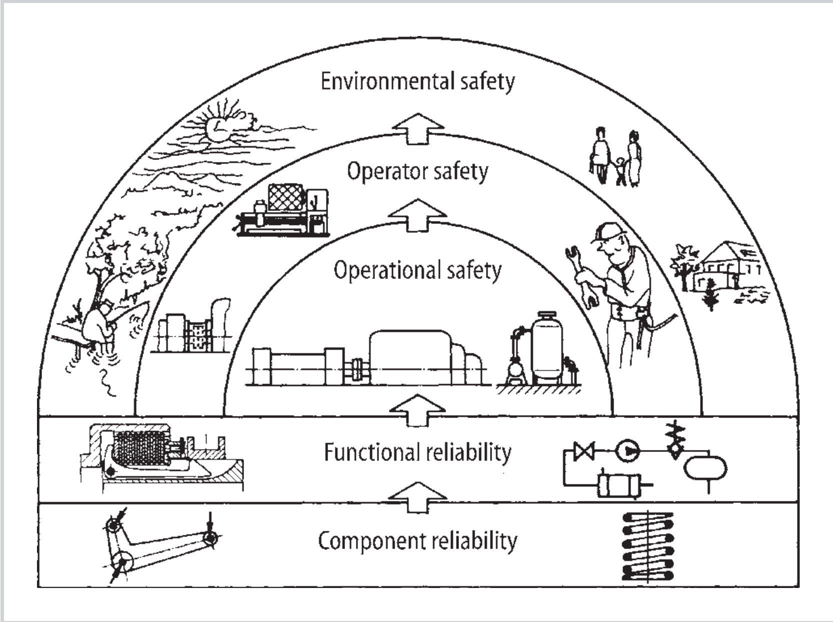


Figure 7.14. Relationship between component and functional reliability on the one hand and operational, operator and environmental safety on the other

Operator safety	is the limitation of danger to persons using technical systems either at their workplace or outside, for example for sport or leisure.
Environmental safety	is the limitation of damage to the environment in which technical systems are used.
Protective measure	is the use of protective systems or devices to limit existing dangers and reduce risks to acceptable levels where these cannot be achieved through direct safety measures.

The reliability of assemblies and of their interaction—that is, the functional reliability of a machine or a protective system—is crucial for operational, operator and environmental safety [7.179]. For designers, all these areas of safety are closely connected when developing a concept and its embodiment. A safety methodology should therefore give equal weight to each of the areas [7.210].

2. Direct Safety

Direct safety measures achieve safety through systems or components actively involved in the performance of a particular task. To ensure and evaluate the safe functioning and durability of components, designers can adopt one of several safety principles [7.210]. There are three basic principles, namely:

- safe-life principle
- fail-safe principle
- redundancy principle.

The *safe-life principle* demands that all components and their connections be constructed in such a way as to allow them to operate without breakdown or malfunction throughout their anticipated lives. This is ensured by:

- clear specification of the operating conditions and environmental factors, such as the anticipated loads, service life, operating conditions, etc.
- adequately safe embodiment based on proven principles and calculations
- numerous and thorough inspections during production and assembly
- analysis of components or systems to determine their durability when they are overloaded (load levels and/or running time) or subjected to adverse environmental influences
- determination of the limits of safe operation, with due regard being paid to possible breakdowns.

It is characteristic of this principle that it bases safety exclusively on accurate qualitative and quantitative knowledge of all of the influences at work or on the determination of the limits of failure-free operation. The application of this principle calls for a great deal of experience, or for costly and time-consuming preliminary investigations, and for continuous monitoring of the state of components. If a failure should nevertheless occur, and if a safe-life is essential, then as a rule there will be a serious accident, for instance the fracture of an aeroplane wing or the collapse of a bridge.

The *fail-safe principle* allows for the failure of a system function or for a component fracture during the service life by ensuring that grave consequences do not ensue. To that end:

- a function or capacity, however small, must be preserved to prevent dangerous conditions
- a restricted function must be fulfilled by the failing component or by some other component until such time as the plant or machine can be removed from operation without danger
- the failure or breakdown must be identifiable
- the effect of the failing component on the overall safety of the system must be assessable.

In essence, the impairment of a main function must be signalled. The signal can take various forms (increasing vibrations, loss of sealing, loss of power, slowing down), each without causing immediate danger. In addition, special monitoring systems may be provided to indicate the incipient failure to the operator. Their layout should be governed by the general principles of protective systems. The fail-safe principle presupposes knowledge of the progress of a failure and provides a means for taking over or maintaining the impaired function.

By way of example, let us consider a spherical rubber element in an elastic coupling (see Figure 7.15). The first visible crack appears on the outer layer, but the function is not yet impaired (State 1). Only when the number of revolutions under load is increased does the stiffness begin to decrease with a consequent change in the behaviour of the coupling, which manifests itself, for instance, by a lowering of the critical speed (State 2). With further operation, the crack grows larger and causes the stiffness to decrease still further (State 3), but even if the

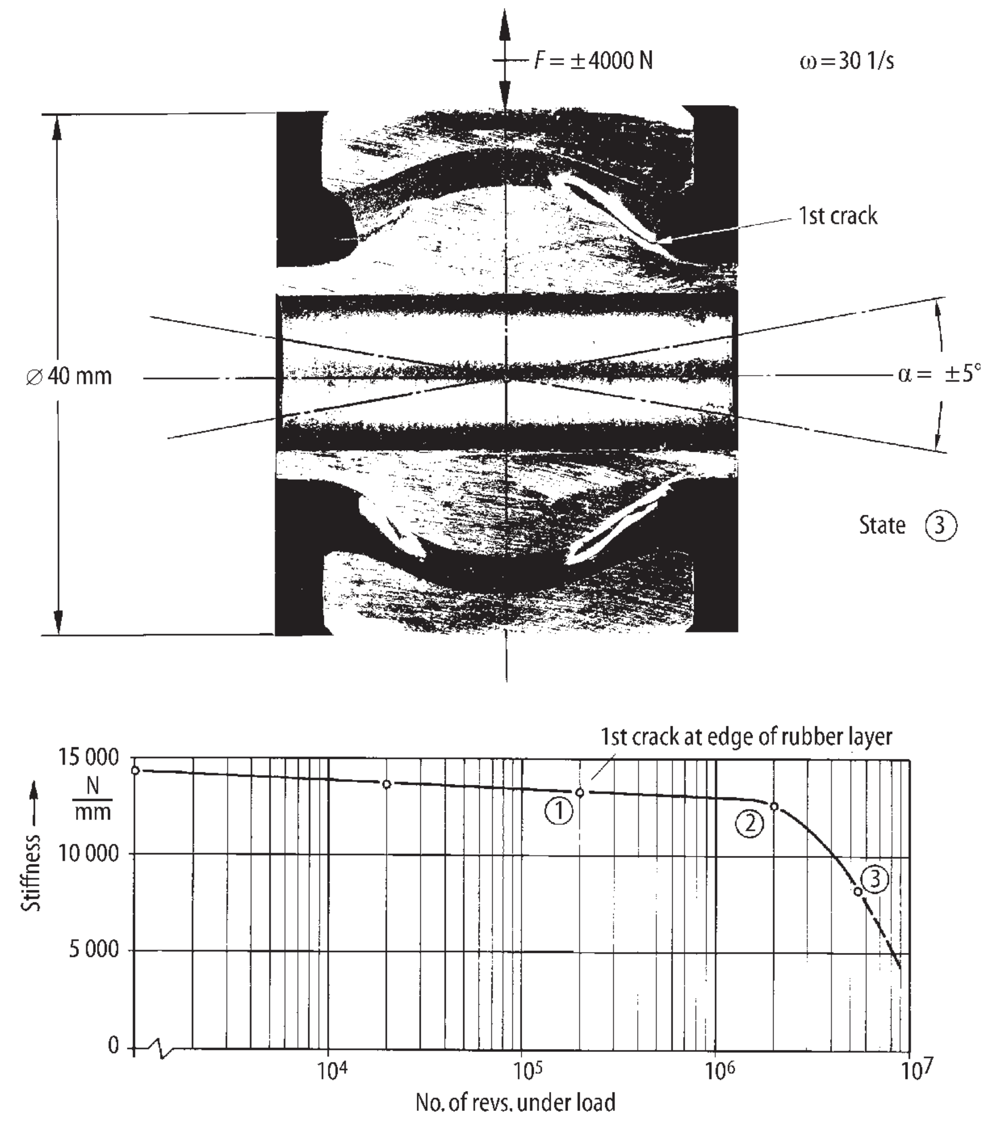


Figure 7.15. Fail-safe behaviour of an elastic coupling: crack-state and stiffness against number of revolutions

crack went right through, there would not be a complete failure of the coupling. Therefore, no sudden effect with serious consequences need be feared.

Another example is the behaviour of flange bolts made of a tough material which, on overloading, exceed their yield strength and deform plastically, resulting in a reduction of preload and, hence, a reduction of the clamping force. Their impaired function is indicated by the resulting loss in flange sealing but does not give rise to sudden failure.

Figure 7.16 illustrates two safe methods of fastening components. The means of attachment should be designed such that, even if the bolts begin to fail, the mountings remain in place, no broken parts can migrate, and the equipment continues to function to some extent [7.206].

The *redundancy principle* provides another means of increasing both the safety and the reliability of systems.

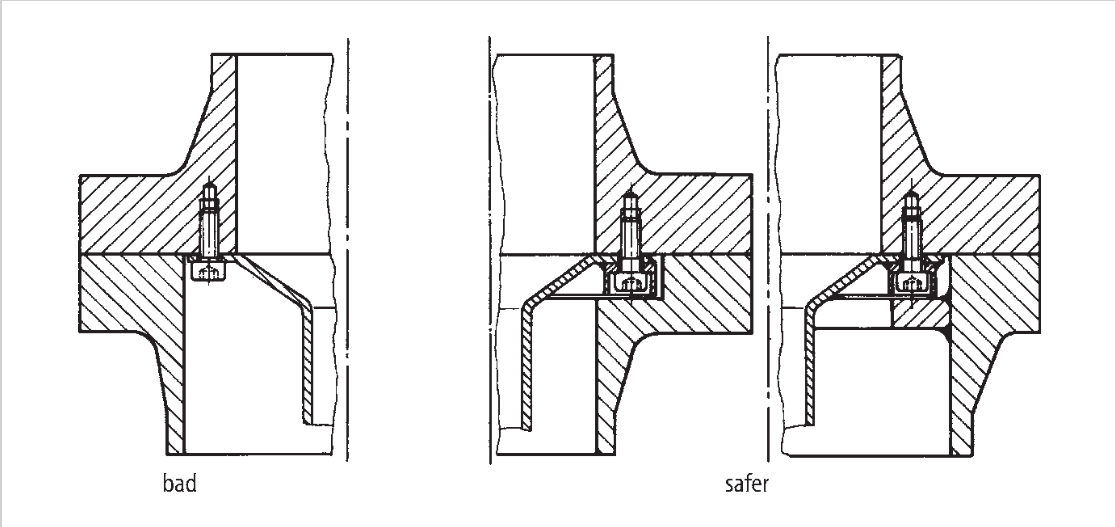


Figure 7.16. Fastening of components: the covering of the bolted connection maintains function and prevents broken parts migrating in the event of bolt failure

In common usage, redundancy means superfluity or excess. In information theory, redundancy refers to that fraction of a message that may be eliminated without loss of essential information. Redundancy is often used deliberately to allow for transmission losses, and hence to safeguard the system. The fact that this safety principle is common in electronics and information technology is useful when integrating these technologies with mechanical engineering systems.

Redundant safety arrangements lead to an increase in safety, provided that the breakdown of a particular element of the system is not dangerous in itself, and that other elements, arranged in parallel or in series, can take over its function fully or at least in part.

The provision of several engines in aircraft, of multistrand cable for a high-voltage transmission line, and of parallel supply lines or generators, all ensure that, should a particular element break down, the function is not completely impaired. In that case, we speak of *active redundancy*, because all the components are actively involved. Partial breakdowns lead to a corresponding reduction in energy or performance.

If reserve elements (for instance alternative boiler feed pumps)—usually of the same type and size—are provided and put into operation during breakdowns, then we speak of *passive redundancy*.

If a multiple arrangement is to be equal in function but different in working principle, then we have *principle redundancy*.

Depending on the situation, safety-enhancing elements can be arranged in parallel, for instance emergency oil pumps, or in series, for instance filter installations. In many cases, layouts in parallel or series will not suffice and crossover links will have to be introduced to guarantee transmission, despite the breakdown of several elements (see Figure 7.17).

In a number of monitoring systems, signals are collected in parallel and compared with one another. *Selective redundancy* (two out of three) and *comparative redundancy* arrangements are shown in Figure 7.17.

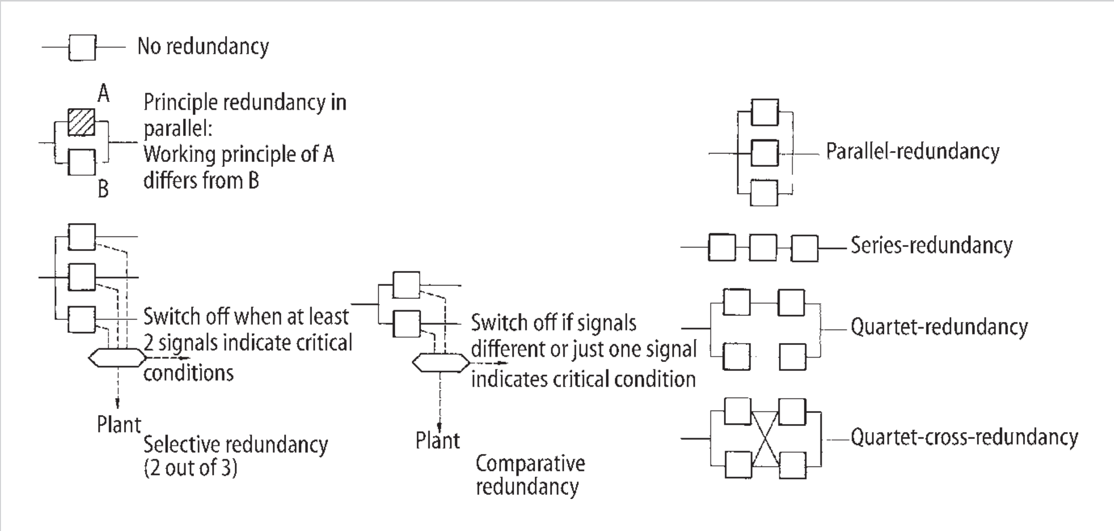


Figure 7.17. Redundant arrangements

Redundancy layouts cannot, however, replace the safe-life or fail-safe principles. Two cable cars operating in parallel will, admittedly, increase the reliability of passenger transport, but this will contribute nothing to the safety of the individual cars. The redundant layout of aircraft engines will not increase safety if any of the engines might explode and hence to endanger the system. In short, an increase in safety can only be guaranteed if the redundant element satisfies the safe-life or the fail-safe principle.

Adherence to all the principles we have mentioned—that is, the attainment of safety in general—is greatly facilitated by the principle of the division of tasks (see Section 7.4.2) and by the two basic rules of clarity and simplicity, as we shall now try to show with the help of an example.

The principle of the division of tasks and the clarity rule have been applied with great consistency to the construction of a helicopter rotor head (see Figure 7.18), and helped the designers to come up with a particularly safe construction based on the safe-life principle. Each of the four rotor blades exerts a radial force on the rotor head due to the centrifugal inertia force, and a bending moment due to the aerodynamic loading. The rotor blades must also be able to swivel so that their angles of incidence can be changed. A high safety level is achieved by the following measures:

- A completely symmetrical layout so that the external bending moments and the radial forces at the rotor head cancel out.
- The radial forces are transmitted exclusively by the torsionally flexible member *Z* to the main central component where they cancel each other out.
- The bending moment is only transmitted through part *B* and is taken up by the roller bearings in the rotor head.

As a result, every component can be optimally designed in accordance with its task. Complicated joints and shapes are avoided and the necessary high level of safety is attained.

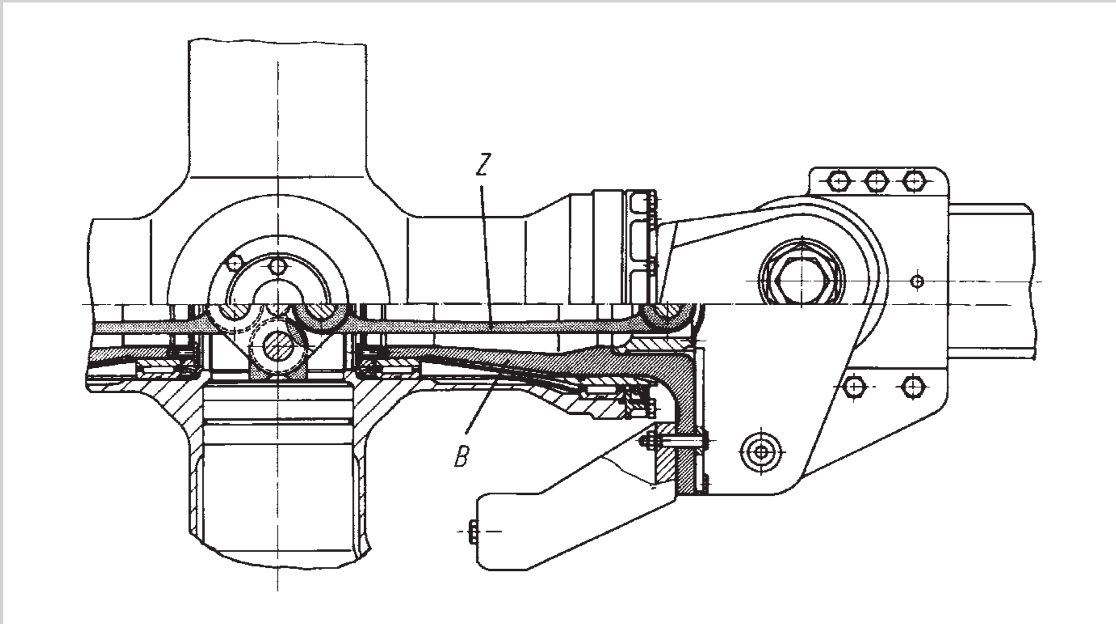


Figure 7.18. Rotor blade attachment of a helicopter based on the principle of the division of tasks (Messerschmitt–Bölkar system)

3. Indirect Safety

Indirect safety measures involve the use of special protective systems and protective devices. They are applied whenever direct safety measures prove inadequate. A detailed discussion of indirect safety measures for technical systems can be found in [7.215]. In what follows, the most important elements of these measures are described.

Protective systems react when danger occurs. To that end, their function structure includes a signal transformation with an input that captures the danger and an output that removes it.

The working structure of a protective system is based on a function structure with the following main functions: capture–process–act. Examples are the multiple redundant monitoring of temperatures in a nuclear reactor; the monitoring of robots in inaccessible workplaces; the sealing of areas when they are subject to X-rays; and the automatic checking of the locking of centrifuge covers prior to operation. The required actions can involve removing, limiting or separating.

Protective devices fulfil protective functions without transforming signals.

Examples are a pressure safety valve (see Figure 7.22); a shaft coupling that slips with torque overload; a pin that shears to limit excessive forces; and safety belts in cars. Their main action is removing or limiting. They can form part of a protective system.

Protective barriers fulfil protective functions without acting.

These barriers are passive, and not able to act on their own. They do not transform signals and therefore do not require a function structure that involves this transformation. They protect by separating; that is, by keeping persons and equipment at a distance from danger using physical barriers, covers, fences, etc. They are described in DIN 31 001, Parts 1 and 2 [7.58, 7.59]. Locking devices, according to Part 5 of this standard [7.60], are regarded as protective systems.

Basic Requirements

Indirect safety measures have to fulfil the following basic requirements:

- operate reliably
- function when danger occurs
- resist tampering.

Operate Reliably

Reliable operation means that: the working principle and the embodiment allow unambiguous operation; the layout follows the established rules; production and assembly are quality-controlled; and the protective systems and devices are rigorously tested. The safety modules and their functional links should be based on direct safety principles and demonstrate safe-life or fail-safe behaviour.

Function When Danger Occurs

This requirement means that:

- the protective function has to be available from the start of the dangerous situation and must last throughout the period of danger
- the protective function should not cease or the protective device should not be removed before the dangerous situation has completely ended.

Figure 7.19 shows example layouts for safety fence contacts for a machine guard. Closed contacts signal that the safety fence is in position. Layout *a* has severe deficiencies because the contact movement relies upon the spring force alone and is not bi-stable (see Section 7.4.4). If the spring breaks or the contacts stick together, the contact will not be broken, that is, the machine can be started with the safety

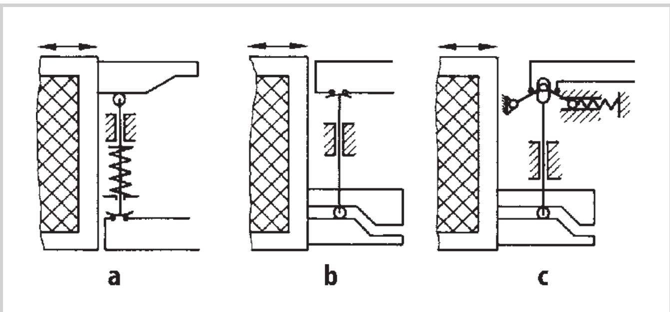


Figure 7.19. Layouts for safety fence contacts for a machine guard. **a** Protection not guaranteed because contact movement relies on a spring force alone. **b** Protection guaranteed because activation relies on form fit. **c** Bi-stable behaviour added to form fit activation in **b**

fence open. Layout *b* will always function when danger occurs. Sticking contacts will be opened because the effect relies on form rather than spring force, and if parts break they will not fall onto the contacts. Layout *c* also makes use of form for activation, but adds spring force and bi-stable behaviour. Further examples can be found in [7.215].

Resist Tampering

Resistance to tampering means that the protection cannot be reduced or removed by unintended or intended actions. If we consider the safety fence contact in Figure 7.19, it should be designed such that actions that prevent correct operation are not possible. The best way to achieve this is to use a cover that cannot be opened without tools or without stopping the machine.

The requirements of protective systems and devices are listed in the following paragraphs followed by those of protective barriers.

Protective Systems and Devices

Protective systems and devices render endangered plant or machinery safe automatically, with the aim of preventing danger to persons and machinery. In principle, the following approaches are available:

- When danger occurs, prevent the consequences by disabling the plant or machinery or preventing any plant or machinery in a dangerous state from being put into operation.
- When there is a continuous danger, avoid its effects by introducing protective measures.

The basic requirements "operate reliably", "function when danger occurs", and "resist tampering" are supported by fulfilling the following requirements.

Warning

When a protective system notes changes in the working conditions, a warning must be provided that indicates the change and the cause of the warning. Examples are "oil level too low", "temperature too high", and "safety fence open". Recommended acoustic and optical signals are given in DIN 33 404 [7.69], colours for warning lights and push buttons in DIN IEC 73/VDE 0199 [7.77], and special safety symbols in DIN 4844 [7.40–7.42].

Two-Step Action

If the dangerous situation emerges so *slowly* that operator action can reduce the danger, then a warning should be given before a protective action is initiated.

Between the two steps, there should be a sufficiently large and clearly defined change in the danger variable. For example, if pressure is the danger variable being monitored, a warning could be given at 1.05 p_{normal} and shutdown initiated at 1.1 p_{normal}.

If the dangerous situation emerges too *quickly*, the protective system should react immediately and signal its response clearly. The terms “slowly” and “quickly” must be interpreted in the context of the cycle time of the technical process and the reaction time required [7.243].

Self-Monitoring

A protective system must be self-monitoring; that is, it must be triggered not only when the system breaks down, but also by faults in its own system. This requirement is best satisfied by the *stored energy principle*, because, when this is applied, the energy needed to activate the safety device is stored within the system and any disturbance of or fault in the protective system will release that energy and switch off the plant or machinery. This principle can be used not only in electronic protective systems but also in mechanical, hydraulic and pneumatic systems.

The stored energy principle has been used in the valve shown in Figure 7.20. When the valve opens, the spring is compressed by the operating oil pressure. When the oil pressure reduces, the spring extends and the valve closes. Failure of the spring will not inhibit the closure of the valve because of the particular configuration used. The flow direction selected and the suspended configuration support the requirement of always functioning when danger arises.

A further example of the use of the stored energy principle in a hydraulic system is shown in Figure 7.21. In this protective system, pump 1 with a pressure-regulating valve 2 ensures a constant pre-pressure p_P. The protective system with the pressure p_S is connected to the pre-pressure system by means of an orifice 3. Under normal conditions, all outlets are closed, so that the quick-action stop valve 4 is held open

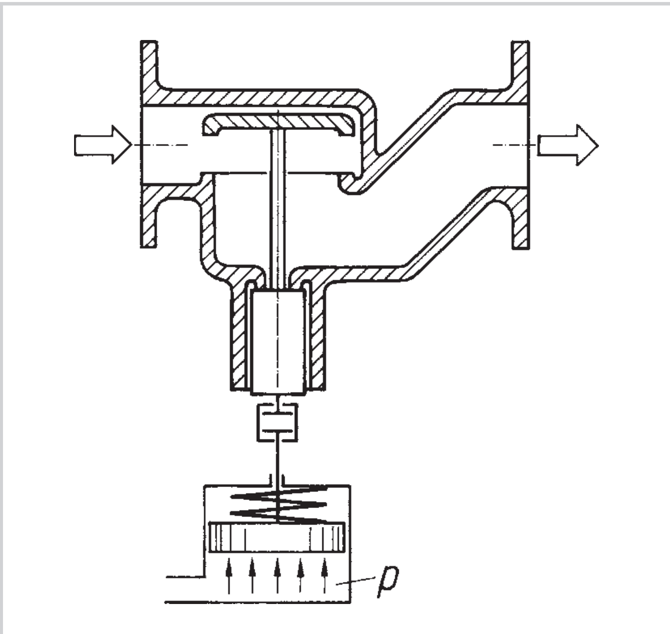


Figure 7.20. Layout of a quick-action valve. In the event of a drop in oil pressure p, the spring force, the flow pressure on the valve face and the weight of the valve act together to guarantee the rapid closure of the valve

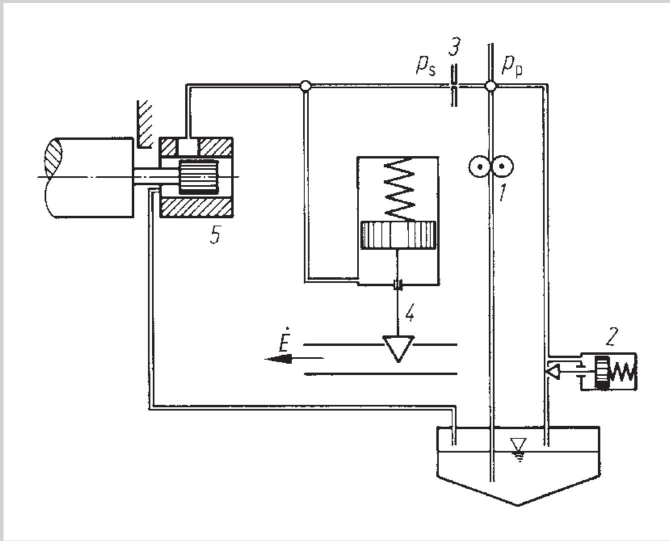


Figure 7.21. Hydraulic protection system employed to prevent incorrect axial shaft positions based on the stored energy principle

by the pressure p_s, allowing energy to be supplied to the machine. In the case of a faulty axial shaft position, the piston valve 5 at the end of the shaft opens, the pressure p_s drops, and further energy supplies are cut off by the quick-action stop valve 4. The same effect is produced by damage to the pre-pressure or protective system, for example by pipe fracture, lack of oil or pump failure. The system is self-monitoring.

A system operating on the *active energy principle*, where energy is only generated in the case of danger, cannot detect a failure in its own system. Therefore, this approach should only be used to provide the warning signals of a protective system when a monitoring system is also available and the system is checked regularly. The possibility that a protective system based on the stored energy principle can cause interruptions that are not caused by a dangerous situation but instead by the protective system itself should be met by increasing the reliability of the system elements, and not through application, for example, of the active energy principle.

Redundancy

The failure of a protective system or device should be seen as a real possibility. Because a single protective system may break down, its mere doubling or replication ensures greater safety: it is unlikely that all the systems will fail at once. A solution that is often applied in protective systems is redundancy based on two from three selection. Three sensors are used to detect the same danger signal (see Figure 7.17). Only when at least two sensors signal the critical value is the protective action—such as machine shutdown—initiated. Thus the failure of a single sensor does not reduce the protective cover, and its failure will not trigger an unnecessary protective action [7.179].

This is however only true provided that the replicated protective systems do not all fail due to a common fault. Safety is considerably increased if the double or multiple systems work independently of one another and are, moreover, based on different working principles (principle redundancy). In this case, common faults—for instance those due to corrosion—will not have catastrophic consequences: the simultaneous breakdown of all such systems is highly improbable.

Figure 7.22 illustrates protective devices employed to prevent excessive pressure in pressure vessels. Mere doubling would not protect against common failures such as corrosion or inappropriate materials. The use of different working principles, however, reduces the possibility of simultaneous failure.

When redundant configurations are linked in parallel or series, the values at which they are triggered should be carefully staggered within an appropriate range. In this manner, primary and secondary protection can be established. In the example in Figure 7.22, the configuration should be chosen such that the safety valve is activated at a lower excess pressure than the shear plate.

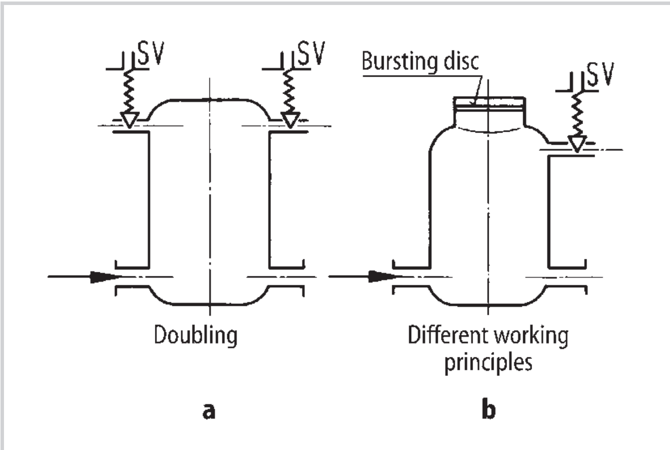


Figure 7.22. Protective devices employed to protect against excessive pressure build-up in pressure vessels: **a** two safety valves (not safe against common faults); **b** safety valve and shear plate (principle redundancy)

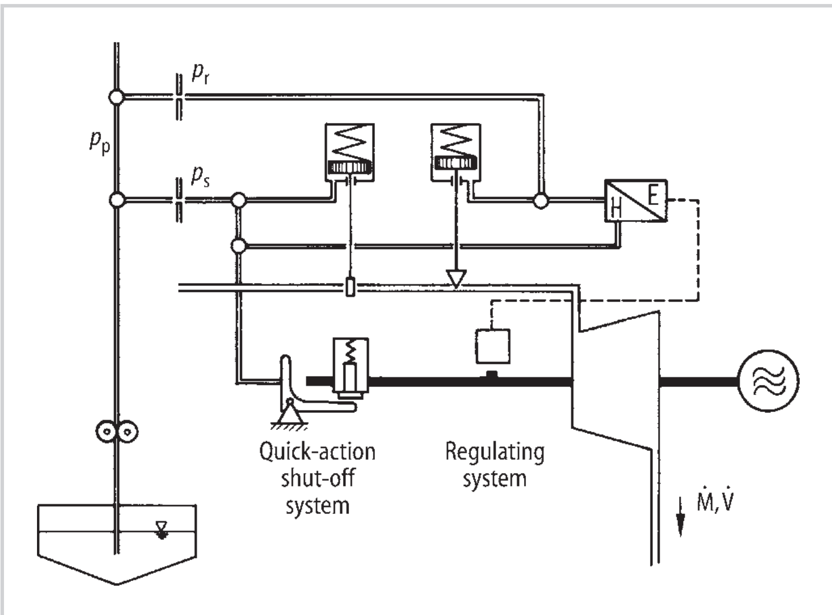


Figure 7.23. Stored energy protective system against overspeeding based on principle redundancy

In many cases the primary protection system can receive its signals from an existing control system, if it has the characteristics of a protective system. This requirement is met in the control of steam turbines shown in Figure 7.23 [7.272]. In the case of overspeeding, the energy supply is cut off by two systems that differ in principle. Increases in speed first invoke the regulating system, whose speed measurement and regulating valve are independent of, and different in principle to, the quick-action shut-off system.

Speed is measured by three identical but independent magnetic sensors. They take their measurements from a gear wheel on the turbine shaft (see Figure 7.24). Their primary purpose is to control the speed of the machine through electronics and hydraulics. In addition, each signal is compared with a reference signal in order to prevent excess speed. This comparison is based on the two from three

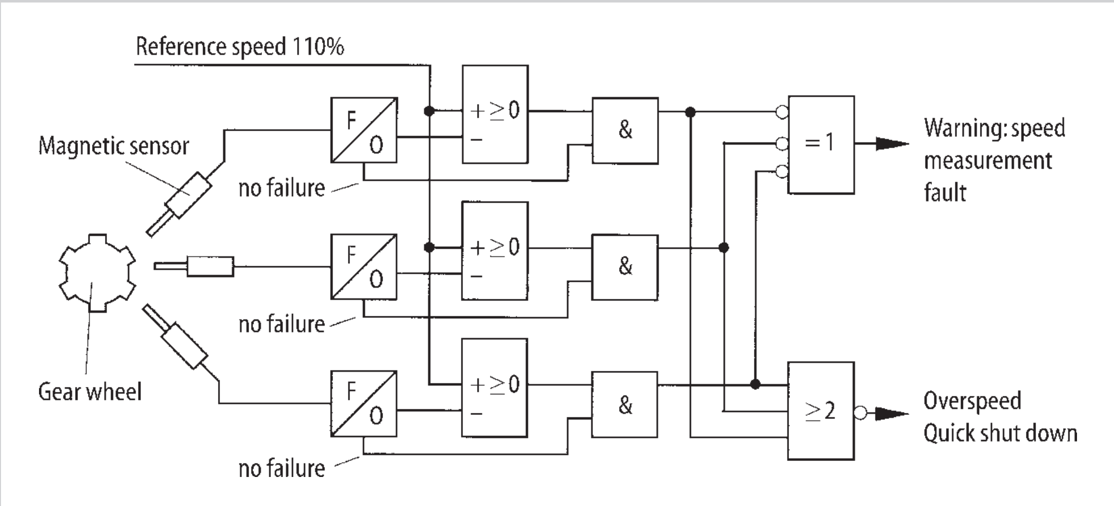


Figure 7.24. Electronic speed control and speed monitoring using a redundant layout based on the two from three principle (simplified representation). Safety is based on the stored energy principle, which is also applied to the quick-action shut-off system

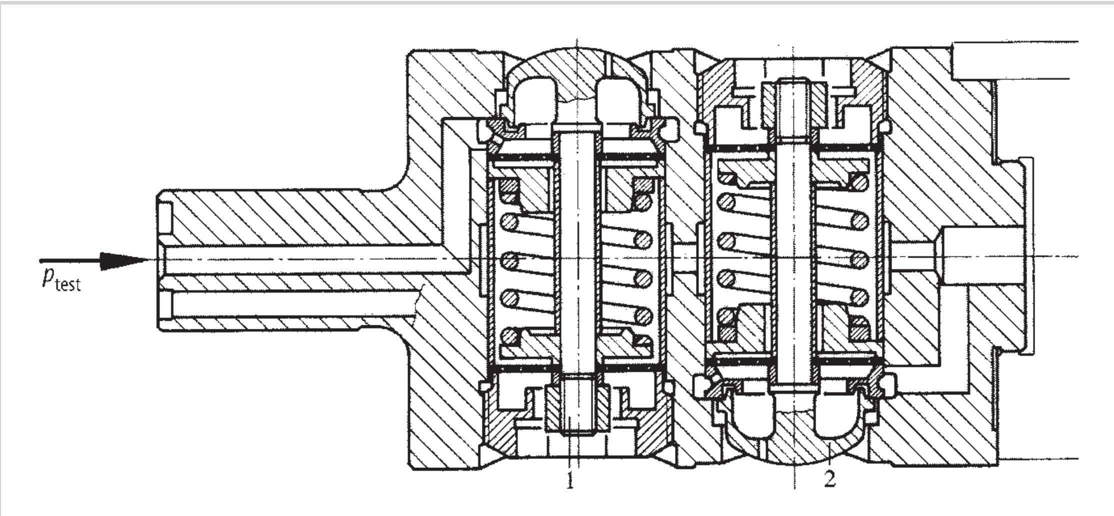


Figure 7.25. Stored energy protective system against overspeeding based on two triggering values

principle. Each measurement circuit is monitored separately, and any failures are signalled. If two fail, the quick-action shut-off system is activated immediately.

The measurement and the activation of the quick-action system, however, are based on a mechanical principle. Figure 7.25 shows quick-action pins that, in the case of excess speed, move out rapidly against their retaining springs and strike a trigger. This in turn activates the quick-action shut-off system hydraulically. The turbine is provided with two such bi-stable devices that trigger at 110% and 112% excess speed respectively (see Section 7.4.4).

A common hydraulic supply to the control and quick-action shut-off system based on the stored energy principle is acceptable because both are based on a common self-monitoring principle.

Bi-Stability

Protective systems and devices must be designed with a clearly defined triggering value. When this value is attained, the protective reaction must be initiated immediately and unambiguously. This can be achieved by using the bi-stable principle (see Section 7.4.4). Below the triggering value, the system is in a stable state. When the triggering value is attained, an unstable condition is created deliberately. This avoids intermediate states and transfers the system rapidly into its second stable state. This bi-stable characteristic must be realised without intermediate states occurring when the triggering value is reached in order to achieve clarity in the behaviour of the protective system or device.

Preventing System Restarts

After a protective system or device has been activated, that system should not automatically return a machine to normal operation, even if the danger recedes. The activation of a protective system is always triggered by an unusual situation. After shutdown, the situation should be checked and evaluated, and the subsequent restart should follow a clearly structured procedure. For example, the safety regulations covering protective systems and devices [7.256], as well as other machines used in production [7.334], prescribe procedures for restarting.

Testability

A protective system or device should allow its functioning to be tested without having to create a situation with real danger. However, it might be necessary to simulate a dangerous situation in order to trigger the protective system. During a simulation, the effects used must be similar to the real danger and all possible danger conditions checked.

In our speed control system example, this means a planned increase in speed up to the excess speed, at which point the protective system triggers. If this is not possible or it is not desirable, it is possible to simulate the centrifugal inertia force by using oil pressure to trigger the system. The machine does not have to be shut down for this simulation. Figure 7.25 shows the oil channel. The oil simulates an increase in the centrifugal inertia force on the quick-action shut-off pins so that they are triggered and their action tested without attaining an excess speed.

With redundant protective systems, it is possible to isolate individual systems from the machine to test them. Any other redundant protective systems can remain active and continue to monitor safety during the test. Care must be taken to ensure that the protective system automatically returns into its fully operational state after test procedures that only check part of the system.

From the previous paragraphs, the following points emerge:

- protection must be retained during testing
- testing must not introduce new dangers
- after testing, the parts tested should return automatically to their fully operational state.

Often a *start-up check* is useful, or even prescribed. This check permits the operation of a machine only after its functions have been tested by activating the protective system. Safety regulations, for example, often prescribe this type of start-up check for power tools with safety devices [7.256].

Protective systems and devices must be *tested regularly*, that is:

- before the first operation
- at regular predetermined intervals
- after every service, repair or modification.

The procedures should be described in operating manuals and the results documented.

Relaxing the Requirements

At this point, one may question whether it is necessary to meet the testability requirement as well as that of self-monitoring. However, even protective systems based on the stored energy principle include elements whose full functionality can only be assessed through testing. Examples include the operation of the quick-action pins in Figure 7.25, and sticking contacts in an electric switch.

Relaxation of the safety system requirements is only permissible when the probability of failure is so small and the consequences of any failure are so limited that the overall risk is acceptable. This will only be the case with redundancy requirements when system tests are easy and carried out regularly. This occurs when these tests are part of normal operation, for example when start-up checks are implemented. This often applies to protective systems associated with safety at work.

If human life is endangered or large-scale damage may occur, leaving out redundancy is neither justified nor economic. Which redundancy is applied, for example two from three selection, replication of the same principle, or principle redundancy, depends on the specific context and the level of risk.

Protective Barriers

The purpose of a protective barrier is to isolate people and objects from the source of danger, and to protect them from a variety of dangerous effects. DIN 31 001 Part 1 [7.58] and Part 2 [7.59] deal mainly with protection against physical contact with dangerous static and moving parts, and against objects and particles that break away. Elaborate illustrations and examples are given in [7.215].

The desired solution principles (see Figure 7.26) prevent contact by providing:

- full enclosure
- cover for a particular side
- fence, used to maintain a safe distance.

Safety distances play an essential role when it is possible to reach through or around fences or barriers. These distances are determined by body dimensions and ranges of reach. DIN 31 001 Part 1 [7.58] gives clear safety distances, depending on body dimensions and posture.

With respect to contact protection and protection against objects and particles that break away, DIN 31 001 Part 2 [7.59] only permits the use of those materials that can fulfil their protective function on the basis of their durability, shape stability, temperature resistance, corrosion resistance, resistance to aggressive substances, and their permeability to those aggressive substances.

4. Designing for Safety

The checklist in Figure 7.3 can prove a great help. Safety criteria must be scrutinised with respect to all the headings listed [7.303].

Function and Working Principle

It is important to establish whether or not the function is fulfilled safely and reliably by the chosen solution. Likely faults and disturbing factors must be taken

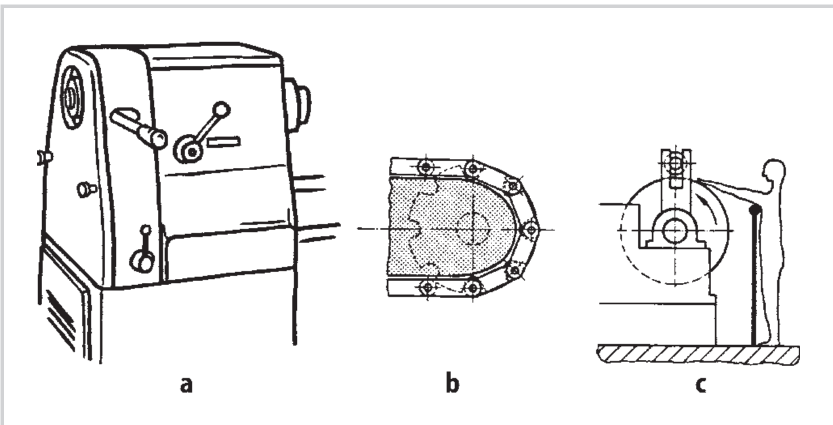


Figure 7.26. Examples of protective barriers: **a** full enclosure; **b** cover for a particular side; **c** fence used to maintain a safe distance

into account as well. The extent to which allowances must be made for exceptional, purely hypothetical, circumstances that could affect the function is not always clear, however.

The correct estimation of the scope and likelihood of a risk should be based on the successive negation of each of the functions to be fulfilled and on an analysis of the likely consequences (see Section 10.2). Sabotage need not necessarily be considered in this context, because measures to prevent human errors are likely to cover most possible circumstances.

What we have to consider and prevent first and foremost are failures due to possible disturbances of the structure, operation and environment of a machine, as well as those caused by operator error. Harmful effects that are not due to technological factors cannot be eliminated by the technical system itself, but the system must be able to survive them and, if possible, limit them.

A further question is whether the direct safety measures we have been discussing are adequate, or whether safety should be increased by additional protective systems and devices. Finally, we might also ask whether the whole project should be abandoned if it proves to be impossible to make adequate safety provisions in a particular case. The answer depends on the degree of safety that has been attained, on the probability of unpreventable damage or accident, and on the magnitude of the possible consequences. Objective standards are often lacking, particularly in the case of new applications. It has been argued that technical risks must be no greater than the risks humans must expect from natural causes [7.138]. However, this is always a matter for discretion. The final decision should, in any case, reflect a responsible attitude towards the human race.

Layout

External loads produce stresses in components. Through analysis we determine their magnitude and frequency (steady and/or alternating loads). The various types of stress produced can be determined by calculation or experiment. The calculated stresses in a component are then, using an appropriate failure hypothesis, converted into an *equivalent stress* σ_E, which should correctly represent the combined direct and shear stresses. The maximum equivalent stress should not exceed the *allowable stress* σ_A. When the two are equal, the material utilisation is 1.0. In general, the ratio of the equivalent stress divided by the allowable stress is smaller than 1.0, because the choice of dimensions is also influenced by standards and other embodiment considerations.

Materials technology provides designers with *material stress limits* σ_L or particular conditions (tension, compression, bending, shear and torsion), beyond which the material will fail or permanently deform. These values are usually obtained from test specimens and not from the components themselves. The strength of a component is also affected by uneven loading, and by its size, surface finish and shape. Only when these are taken into consideration can adequate durability be guaranteed. Thus the component stress limit is usually lower than the material stress limit.

The ratio of the material stress limit (or of the component stress limit) to the allowable stress is the *Safety Factor*, $(SF) = \sigma_L/\sigma_A$. This value must be greater than 1.0. Safety factors are provided in reference manuals for specific situations and types of materials, and the allowable stress σ_A in a component can easily be calculated using these.

The value of a safety factor depends on uncertainties in the determination of the material stress limits; on uncertainties in the load assumptions; on the calculation methods; on the production processes; on the (uncertain) influences of shape, size and environment; and also on the probability and importance of possible failures.

The determination of safety factors still lacks generally valid criteria. An investigation by the authors has shown that published recommended safety factors cannot be classified by type of product, branch of engineering or other criteria such as toughness of material, size of component, probability of failure, etc. Tradition, figures based on one-off and often inadequately explained failures, hunches and experiences are often the basis for numerical data from which no generally valid statements can be derived.

The figures that are given in the literature must therefore be treated with circumspection. Their application usually calls for a knowledge of the individual circumstances and of the special practices or regulations of the branch of engineering in question. In general, however, safety factors smaller than 1.5 should only be used when more precise calculation procedures have been used, experimental data are available, a sufficiently ductile material is used, or there is experience with the specific application. For brittle materials subject to stresses that lead to brittle fracture, the safety factor will be nearer to 2.0.

Toughness—that is, the ability to undergo plastic deformation before failure and thus relieve stress concentrations caused by unevenly distributed loads—is one of the most important safety features any material can have. The usual overspeed spinning tests of rotors with the correspondingly high stresses they set-up, and also the required overpressure tests of pressure vessels—provided that they are built of tough materials—are good examples of the direct safety method aimed at reducing stress concentrations in finished components.

Because toughness is a crucial safety-enhancing property of materials, it is not enough simply to aim at greater yield strength. Since, in general, the toughness of materials decreases with increasing yield strength, it is essential to ensure a minimum toughness, otherwise the benefits of plastic deformation are no longer guaranteed. Also dangerous are those cases in which the material turns brittle with time or for other reasons (for instance, due to radiation, corrosion, heat, or surface coatings). This is particularly true of synthetic materials.

If the safety of a component is calculated merely by the difference between the computed stress and the maximum permissible stress, a vital point is missed.

Of the utmost importance is the loading condition and the effect on the properties of the material due to ageing, heat, radiation, weathering, operating conditions and production processes, for instance welding and heat treatment. Residual stresses must not be underestimated either: brittle (fast) fractures without plastic deformation can occur suddenly and without warning. The avoidance of a build-up of additive stresses, of brittle materials, and of production processes that encourage brittle fractures, is therefore an essential requirement of direct safety.

If plastic deformation is monitored at a critical point, or can be used to impede the function in such a way that the danger can be noticed before humans or machines are endangered, it becomes fail-safe [7.206].

Elastic deformations must not be allowed to disturb the smooth functioning of a machine, for instance through loss of clearance. If this happens, the force transmission paths or the expansions can no longer be determined with certainty and overloading or fracture may ensue. This is just as true of stationary as it is of moving parts (see Section 7.4.1).

By *stability* we refer not only to the basic stability of a machine but also to its stable operation. Disturbances should be counteracted by stabilising effects, that is, by automatic return to the initial or normal position. Designers must ensure neutral equilibrium or that potentially unstable states do not lead to a build-up of disturbances that might get out of control (see Section 7.4.4).

Resonances produce increased stresses that cannot be accurately determined. They must be avoided unless the amplitudes can be sufficiently damped. This applies not only to the stability problem, but also to such associated phenomena as noise and vibration, which impair the efficiency and health of operators.

Thermal expansions must be taken into account under all operating conditions, in particular during unsteady processes, if overloading and impairment of the function are to be avoided (see Section 7.5.2).

Inefficient *seals* are a common cause of breakdown or trouble. Careful choice of seals, provision for pressure relief at critical sealing points and careful attention to fluid dynamics help to overcome these problems.

Wear and the resulting particles can also impede operational safety, and must therefore be kept within tolerable limits. In particular, designers should ensure that such particles do not damage or interfere with other components. They should be removed as near as possible to their point of origin (see Section 7.5.13).

Uniform *corrosion* reduces the designed thickness of components. Local corrosion, particularly of components subject to dynamic loading, may appreciably increase stress concentrations and lead to fast fractures with little deformation. There is no such thing as permanent stability under corrosion—the load capacity of components decreases with time. Apart from fretting corrosion and fatigue corrosion, stress corrosion can also be very serious for certain materials subject to tensile stresses in the presence of corrosive media. Finally, corrosion products can impede the functioning of machines, for instance by jamming valve spindles, control mechanisms, etc. (see Section 7.5.4).

Ergonomics

The application of ergonomic principles to industrial safety involves the careful scrutiny of sources and locations of danger as well as of human–machine relationships. Possible human errors and fatigue must also be included. Machines and products therefore have to be designed ergonomically (see Section 7.5.5).

Table 7.1. Harmful effects associated with various types of energy

Protect humans and environment against harmful effects Headings	 Examples
Mechanical	Relative movement of human and machine, mechanical vibrations, dust
Acoustic	Noise
Hydraulic	Jets of liquid
Pneumatic	Jets of gas, pressure waves
Electrical	Passage of current through body, electrostatic discharges
Optical	Dazzle, ultra-violet radiation, arcs
Thermal	Hot and cold parts, radiation, inflammation
Chemical	Acids, alkalis, poisons, gases, vapours
Radioactive	Nuclear radiation, X-rays

Table 7.2. Minimum industrial safety requirements in mechanical devices

In mechanical devices, protruding or moving parts should be avoided in areas where human contacts may occur
Protective equipment is required for the following, regardless of the operational speed: • for gear, belt, chain and rope drives • for all rotating parts longer than 50 mm, even if they are completely smooth • for all couplings • in cases of danger from flying parts • for potential traps (slides coming up against stops, components pushing or rotating against each other) • descending components (weights, counter-weights) • for slots, for example at material inputs. The gaps between parts must not exceed 8 mm; in the case of rollers, the geometrical relationship must be examined and, if necessary, special guards must be installed *Electrical installation* must always be planned in collaboration with electrical experts. In the case of *acoustic*, *chemical* and *radioactive* dangers, expert advice must be sought for the requisite protection

A great many books and papers have been devoted to this subject [7.26, 7.65, 7.189, 7.255, 7.303]. In addition, DIN 31 000 [7.57] specifies the basic requirements of design for safety, and Parts 1, 2 and 10 of DIN 31 001 [7.58, 7.59] deal with protective equipment. Regulations by various professional bodies, factory inspectorates, etc., must be scrupulously observed in all branches of engineering, and so must a great deal of special legislation [7.115] (see also [7.334]). In this book it is impossible to examine every aspect of industrial safety.

Tables 7.1 and 7.2 provide a introductory guide to the sources of danger and the minimum requirements for industrial safety.

Production and Quality Control

Components must be designed in such a way that their qualities are maintained during production (see Chapter 10). To that end, special quality controls must be instituted, if necessary by special regulations. Through appropriate design measures, designers must help to avoid the emergence of dangerous weak spots in the course of production processes (see Sections 7.3.1, 7.3.2 and 7.5.8).

Assembly and Transport

The loads to which a product will be subjected during assembly and transport must be taken into consideration during the embodiment design phase. Welds carried out during assembly must be tested and, where necessary, heat treated. All major assembly processes should, whenever possible, be concluded by functional checks.

For safe transportation, firm bases, support points and handling points should always be provided and marked clearly. The weights of parts heavier than l00 kg should be marked where they can be seen easily. If frequent dismantling is called for, the appropriate lifting points must be incorporated.

Operation

Operation and handling must be safe [7.57, 7.58]. The failure of any automatic device must be indicated at once so that the requisite actions can be taken.

Maintenance

Maintenance and repair work must only be undertaken when the machine is shut down. Particular care is needed to ensure that assembly or adjusting tools are not left behind in the machine. Safety switches must ensure that the machinery is not started unintentionally. Centrally placed, easily accessible and simple service and adjustment points should be provided. During inspection or repair, safe access should be possible through the provision of handrails, steps, nonslip surfaces, etc.

Costs and Schedules

Cost and schedule requirements must not affect safety. Cost limits and delivery dates are ensured by careful planning, and by implementing the correct concepts and measures, not by cutting corners. The consequences of accidents and failures are generally much greater and graver than the effort needed to prevent them.

7.4 Principles of Embodiment Design

The general principles of embodiment design have been discussed at some length in the literature. Kesselring [7.148] set out principles of minimum production costs, minimum space requirements, minimum weight, minimum losses, and optimum handling (see Section 1.2.2). Leyer discussed the principle of lightweight construction [7.167] and the principle of uniform wall thickness [7.168]. It is obviously neither possible nor desirable to have all of these principles implemented in every technical solution—one of them might be crucial, the rest merely desirable. Which principle should be prioritised in a given case can only be deduced from the task and the company's facilities. By proceeding systematically, elaborating a requirements list, abstracting to identify the crux of the problem, and also by following the checklist given in Figure 5.3, designers transform these principles into concrete proposals that enable them to determine production costs, space requirements, weights, etc. These have to be consistent with the requirements list.

The systematic approach also highlights the question of how, with a given problem and a fixed solution principle, a function can be best fulfilled and by which type of function carrier. Embodiment design principles facilitate this part of the design process. In particular, they help with Steps 3 and 4, but also with Steps 7 to 9 as listed in Section 7.1.

Initially embodiment problems focus predominantly on issues of channelling, combining and storing. For the relatively common task of *transmitting* (channelling) forces or moments, it seems advisable to establish special "principles of force transmission". *Changing* the type or *varying* the magnitude of a force are primarily fulfilled by the appropriate physical effects, but designers must also apply the "principle of minimum losses" [7.148] for energy conservation or economic reasons, which they do by adopting a small number of highly efficient steps. This principle also applies to the efficient conversion of one type of energy into another, whenever this should be required. *Storing* energy involves the accumulation of potential and kinetic energy, be it directly or indirectly through the collection of material. The storage of energy, however, raises the question of the stability of the system, and the consequent application of the "principles of stability and bi-stability".

Often, several functions have to be fulfilled by one or several function carriers. Here the "principle of the division of tasks" may be useful to designers. Its application involves careful analysis of the functions and their assignment to function carriers. This analysis of functions is also helpful for the application of the "principle of self-help" when supplementary effects must be identified and exploited.

When applying embodiment design principles, designers may find that they run counter to certain requirements. Thus, the principle of uniform strength may conflict with the demand for minimum costs; the principle of self-help may conflict with fail-safe behaviour (see Section 7.3.3); and the principle of uniform wall thickness chosen for the purpose of simplifying the production process [7.168] may conflict with the demand for lightweight construction or uniform strength.

These principles represent many strategies that are only applicable under certain conditions. In using them, designers must strike a balance between competing demands. To that end, the present authors have developed what they consider to be important embodiment design principles, which will now be presented. Most are based on energy flow considerations and, by analogy, they apply equally well to the flow of material and of signals.

7.4.1 Principles of Force Transmission

1. Flowlines of Force and the Principle of Uniform Strength

The problems solved in mechanical engineering generally involve forces and/or motions and their connection, change, variation or channelling, and involve the conversion of energy, material and signals. The generally applicable function "channel forces" includes the application of loads to, the transfer of forces between, and the transmission of forces through components and devices. Guidelines are provided in [7.168, 7.278]. In general, designers should try to avoid all sudden changes of direction in the flowlines of force—that is, in the force transmission path—caused by sharp deflections and abrupt changes of cross-section. The idea of "flowlines of force" aids the visualisation of the force transmission paths (load paths) through components and devices, and is analogous to flowlines in fluid mechanics. Leyer [7.167, 7.168] has dealt with the transmission of forces at some length, so we can dispense with a detailed discussion of the problem. Designers are advised to consult these important texts. Leyer, moreover, emphasises the complex interaction between the functional, embodiment and production aspects. The concept of force transmission can be summarised as described below.

Force transmission must be understood in a broad sense; that is, it must include the application, transfer and transmission of bending and twisting moments. First, it is important to remember that *external loads* applied to a component produce axial and transverse forces as well as bending and twisting moments at every section. These set up *stresses* (direct and shear) that produce *elastic* or *plastic deformations* (longitudinal, lateral (Poisson), and shear strains, along with bending and twisting).

The section dimensions transmitting the forces are obtained by "mental dissection" of the components at the point under consideration. The sum of the stresses over these sections produces internal forces and moments which must be in equilibrium with the external loads.

The stresses, determined at the relevant section, are then compared with the material properties of tensile strength, yield strength, fatigue strength, creep strength, etc., with due regard being paid to stress concentrations, surface finish and size effects.

The *principle of uniform strength* [7.278] aims, with the help of appropriate materials and shapes, to achieve uniform strength throughout a mechanical device over its anticipated operational life. Like the principle of lightweight construction [7.167], it should be applied whenever economic circumstances allow.

This important consideration often misleads designers into neglecting the deformations (strains) associated with the stresses. It is, however, these very deformations that often throw light on the behaviour of components and tell us what we need to know about their integrity (see Section 7.4.1).

2. Principle of Direct and Short Force Transmission Path

In agreement with Leyer [7.168, 7.208] we consider the following principle to be of great importance:

- If a force or moment is to be transmitted from one place to another with the *minimum possible deformation*, then the shortest and most direct force transmission path is the best.

This principle, which leads to the minimum number of loaded areas, ensures:

- minimum use of material (volume, weight)
- minimum deformation.

This is particularly true if it is possible to solve a problem using tensile or compressive stresses alone, because these stresses, unlike bending and torsional stresses, produce smaller deformations. When a component is in compression, however, special attention must be paid to the danger of buckling.

If, on the other hand, we require a flexible component capable of *considerable elastic deformation*, then a design using bending or torsional stresses is generally the more economical.

The principle is illustrated in Figure 7.27—the mounting of a machine frame on a concrete foundation—where different requirements demand supports with different stiffnesses. This, in turn, has repercussions on the operational behaviour of the machine: different natural and resonant frequencies, modified response to additional loads, etc. The more rigid solutions are obtained with minimum material and space requirements by means of a short support under compression; the most flexible solution by means of a spring, which transmits the force in torsion. If we look at other design solutions, we find many examples of the same principle: for example, in the torsion bar springs of motor cars, or in flexible pipes that rely on bending or torsional deformations.

The choice of means thus depends primarily on the nature of the task; that is, on whether the force transmission path must be designed for durability with

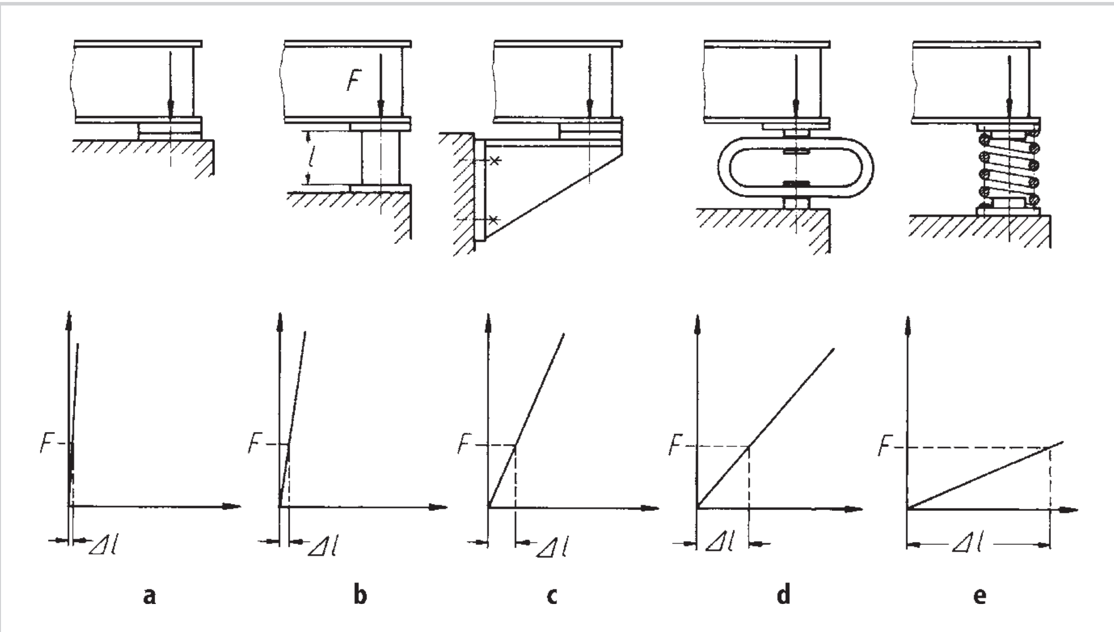


Figure 7.27. Supporting a machine frame on a concrete foundation: **a** very rigid support due to short force transmission path and low stress on the baseplates; **b** longer force transmission path, but still a rigid support with tubes or box sections under compression; **c** less rigid support with pronounced bending deformation (a stiffer construction would involve the greater use of materials); **d** more flexible support under bending stresses; **e** very flexible support using a spring, which transmits the load in torsion. This can be used to alter the resonance characteristics

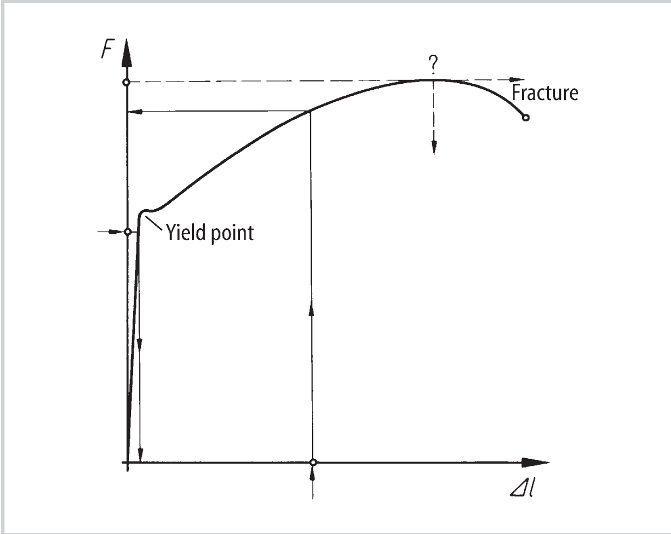


Figure 7.28. Force deformation diagram of tough materials. *Arrows* indicate the cause–effect relationships

maximum stiffness, or whether certain force–deformation relationships must be satisfied first and durability can be treated as a subsidiary problem.

If the *yield point is exceeded*, then the following have to be taken into consideration (see Figure 7.28):

- When a component is *loaded by a force*, it is invariably subjected to deformation. If the yield point is exceeded, then the linear-elastic relationship between the force and the deformation no longer holds. Relatively small changes in the force near the peak of the force–deformation curve may produce unstable conditions leading to fracture, because the load-bearing cross-section may be reduced more rapidly than the strength is increased due to strain hardening. Examples are tie rods, centrifugal inertia forces on a disc, and weights on a rope. The necessary safety precautions must always be taken.
- When a component is *deformed*, then a reaction force is set up. So long as the impressed deformation does not change, the force and the stress remain unchanged as well. If the peak is not reached, the component remains stable so that the yield point can be exceeded without danger. Beyond the yield point, a large change in deformation will lead to only a small change in the force. Admittedly, any preload must not be augmented with further operational loads in the same sense, otherwise the conditions described above will prevail. Further requirements are the use of tough materials and the avoidance of a build-up of multiaxial stresses in the same sense. Examples are highly distorted shrink-fits, preloaded bolts and clamps.

3. Principle of Matched Deformations

Designs matched to the flowlines of force avoid sharp deflections of the transmission path and sudden changes in cross-section, thus preventing the uneven distribution of stresses with high stress concentrations. A visualisation of the flowlines of force, though very graphic, does not always reveal the decisive factors involved. Here, too, the key is the deformation of the affected components.

The principle of matched deformations states that related components must be designed in such a way that, under load, they will deform *in the same sense* and, if possible, *by the same amount*.

As an example, let us take soldered or glued connections in which the solder or adhesive layer has a different modulus of elasticity from that of the material to be joined. Figure 7.29a illustrates the resulting deformation [7.181]. The deformations and the thickness of the solder or adhesive layers have been greatly exaggerated. The load F, which is transmitted across the junction of parts 1 and 2, produces distinct deformations in the overlapping parts, the adhesive layer being subjected to particularly marked deformation near the edges due to differences in the relative deformation of parts 1 and 2. While part 1 bears the full load F at the upper edge of the adhesive layer and is therefore stretched, part 2 does not yet bear a load. The relative shift in the adhesive layer sets up a local shear stress that exceeds the calculated mean value.

A particularly unsatisfactory result is shown in Figure 7.29b where, as a result of opposite and unmatched deformations of parts 1 and 2, the deformation in the adhesive layer is considerably increased. This example makes it clear why provision should be made for deformations to take place in the same sense and, if possible, to be equal in magnitude. Magyar [7.177] has made a mathematical study of the relationships between load and shear stress: the result is shown qualitatively in Figure 7.30.

The same phenomenon also occurs between nuts and bolts in bolted joints [7.328]. The nut (see Figure 7.31a) is in compression and the bolt is in tension, that is, they are deformed in the opposite sense. In the modified nut (see

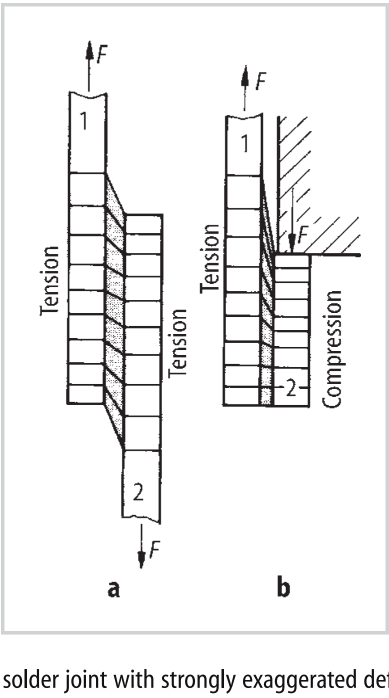


Figure 7.29. Overlapping adhesive or solder joint with strongly exaggerated deformation from [7.181]: **a** Parts 1 and 2 deformed in the same sense; **b** Parts 1 and 2 deformed in the opposite sense

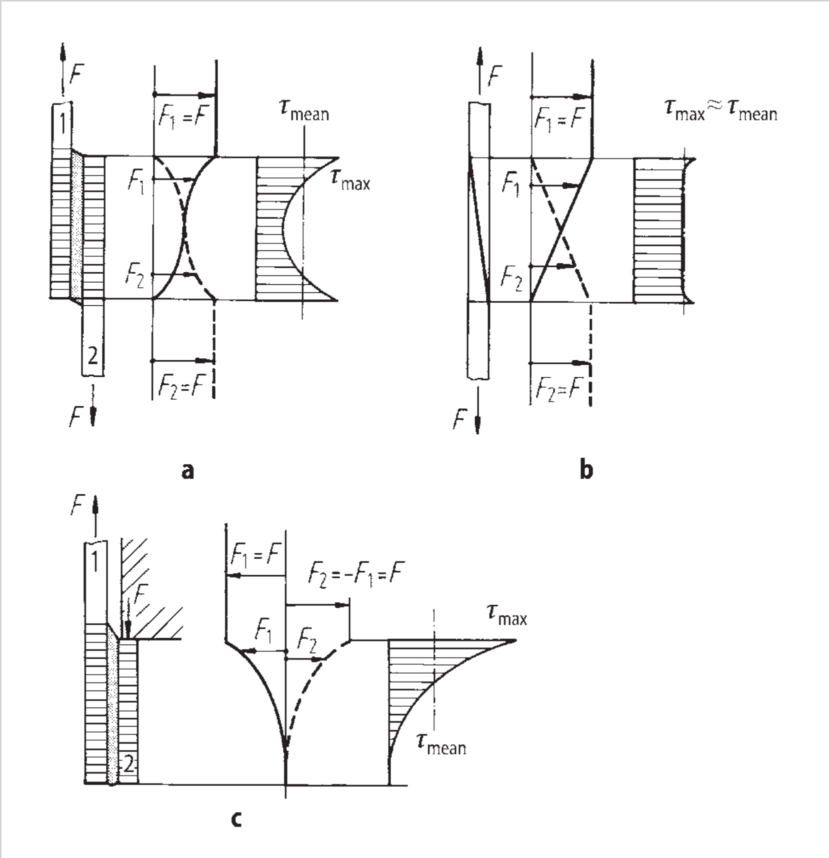


Figure 7.30. Distribution of forces and shear stresses in overlapping joints with layer of adhesive or solder, after [7.177]: **a** overlapped on one side (bending stress neglected); **b** spliced with linearly decreasing thickness; **c** pronounced "deflection of the flowlines of force" with deformations in the opposite sense (bending stress neglected)

Figure 7.31b) a deformation in the same sense is set up in the leading threads, which gives rise to a smaller relative deformation and hence a more even distribution of the load borne by individual threads. Wiegand [7.328] has been able to demonstrate this effect by showing that such nuts have a longer service life. Paland [7.214] has shown more recently that standard nuts are not as unsatisfactory as Maduschka [7.175] has suggested, because the moment $F \cdot h$ produces additional outward deformations of the nut at the contact surface and thus relieves the leading threads of their load. The load-relieving deformation of the nut due to this moment and also to the bending of the threads can be increased considerably by using material with a lower modulus of elasticity. If, on the other hand, the load-relieving deformations are resisted by a very stiff nut or a very small lever arm h, then the type of load distribution described by Maduschka would ensue.

As a further example, let us take a shaft–hub connection formed by a shrink fit. In essence, this too involves the deformation of two components [7.125]. In transmitting the torque, the shaft experiences a torsional deformation that decreases as the torque is transferred to the hub. The hub, for its part, is deformed in accordance with the transmitted torque.

Figure 7.32a shows that the maximum relative deformation occurs at A. In the case of alternating torques, this may lead to fretting corrosion; moreover, the right-

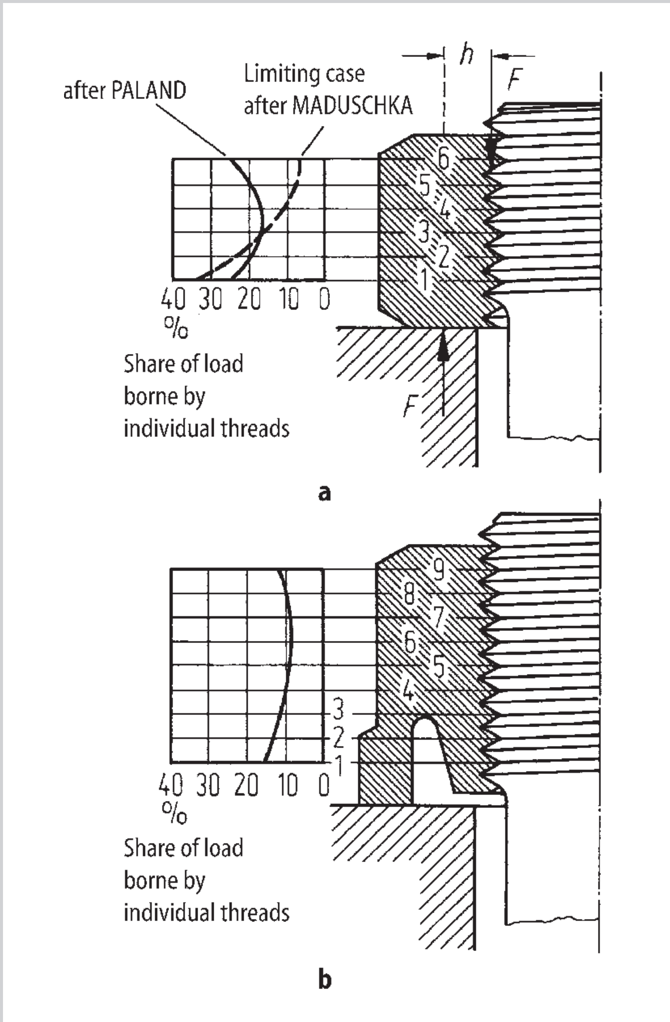


Figure 7.31. Nut shapes and load distribution, after [7.328]: **a** standard nut: limiting case after Maduschka [7.175] and case after Paland [7.214] allowing for deformation due to moment $F \cdot h$; **b** modified nut with matched deformations in the tension part

hand end, to all intents and purposes, contributes nothing to the transfer of the torque.

The solution shown in Figure 7.32b is much better because the resulting deformations are in the same sense. The best solution appears when the torsional stiffness of the hub is matched to that of the shaft. The transfer of torque then takes place along the whole length of the connection, ensuring uniform distribution of force flowlines and thus avoiding stress concentrations.

Even if the shrink fit were replaced with a keyed connection, the layout depicted in Figure 7.32a would, because the torsional deformations are in the opposite sense, set up very high contact stresses in the neighbourhood of A. The layout depicted in Figure 7.32b will, on the other hand, ensure an even contact stress distribution because the deformations are in the same sense [7.188].

The principle of matched deformations can also be applied to bearings, as in Figure 7.33. The embodiment of the bearings should ensure matched deformations between bearing and shaft, or provide for adjustment possibilities.

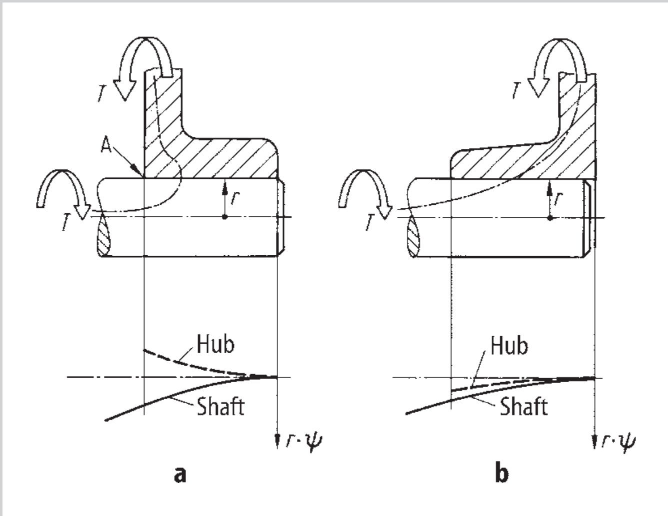


Figure 7.32. **a** Shaft–hub connection with strong "force flowline deflection". Torsional deformations of shaft and hub in opposite sense (ψ = angle of twist). **b** Shaft–hub connection with gradual "force flowline deflection". Torsional deformations of shaft and hub in the same sense

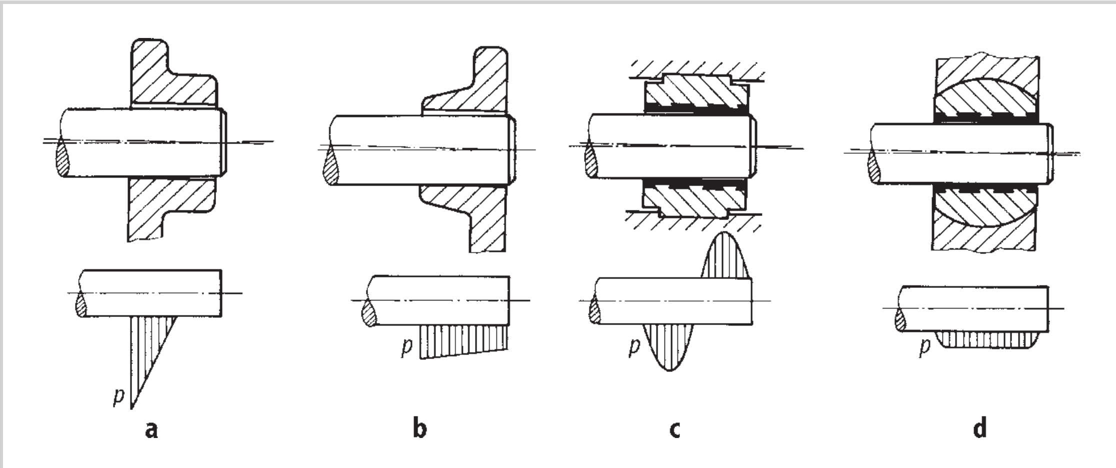


Figure 7.33. Force transmission in bearings: **a** edge compressing because of insufficient adaptation of the bearing to the deformed shaft; **b** more even bearing pressure because of matched deformations; **c** lacking adjustment to shaft deformation; **d** more even bearing pressure because of adaptability of bearing bush

The principle of matched deformations must be taken into account, not only in the transfer of forces from one component to another, but also in the division or combination of forces or moments. A well-known problem is the simultaneous propulsion of wheels that have to be placed at a considerable distance from one another, for instance in crane drive assemblies. In the layout shown in Figure 7.34a, the left side has a relatively high torsional stiffness due to the short force transmission path, and the right side a relatively low torsional stiffness because of its greater path length. When the torque is first applied, the left wheel will be set in motion, while the right wheel remains stationary until the right hand part of the shaft has twisted sufficiently to transmit the torque. The drive assembly has a tendency to run skew.

It is essential to provide the same torsional stiffness to both parts of the shaft so as to ensure an appropriate division of the initial torque. This can be achieved

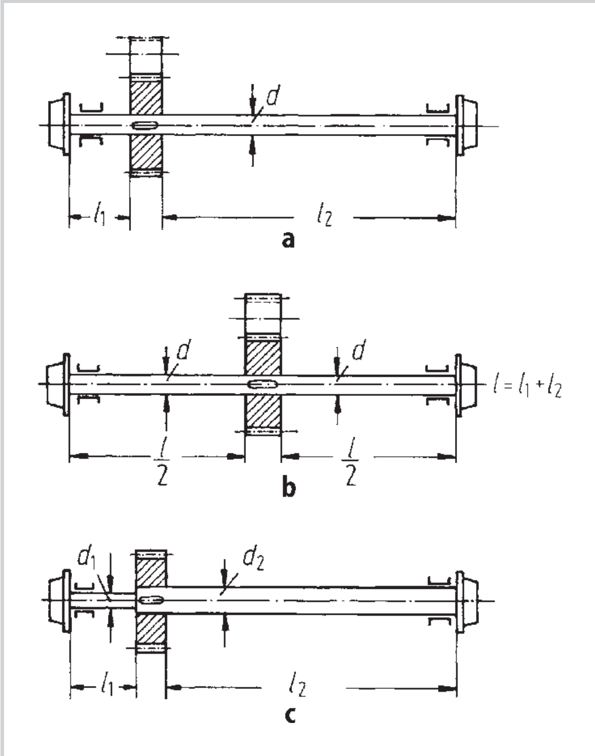


Figure 7.34. Application of the principle of matched—here equal—deformations in crane drives: **a** unequal torsional deformation of lengths l_1 and l_2; **b** symmetrical layout ensures equal torsional deformation; **c** asymmetrical layout with equal torsional deformation due to adaptation of torsional stiffnesses

in two distinct ways if the input torque is taken in one position only: either by symmetrical layout (see Figure 7.34b); or by adaptation of the torsional stiffness of the appropriate parts of the shaft (see Figure 7.34c).

4. Principle of Balanced Forces

Those forces and moments that serve the function directly, such as the driving torque, the tangential tooth force, and the load torque in a gearbox, can, in accordance with the definition of a main function, be described as *functionally determined main forces*.

In addition, there are many forces or moments that do not serve the function directly but that cannot be ignored, for instance:

- the axial force produced by a helical gear
- the force resulting from a pressure difference, for instance across the blades of a turbine or across a control valve
- tensile forces for producing a friction connection
- inertia forces due to linear acceleration or rotation of components
- fluid flow forces, inasmuch as they are not the main forces.

Such forces and moments accompanying the main ones are called *associated forces*, and may either produce a useful auxiliary effect or else appear merely as an unwanted effect that has be taken into account.

Associated forces place additional loads on the components and require an appropriate layout, or must be taken up by further surfaces and elements, such as stiffening members, collars, bearings, etc. As a result, weights are increased and further frictional losses may be incurred. For that reason, the associated forces must, whenever possible, be balanced out at their place of origin, thus obviating the need for a heavier construction or for reinforced bearing and transfer elements.

As has been shown in [7.204], this balance of forces is essentially ensured by two types of solution:

- balancing elements
- symmetrical layout.

Figure 7.35 shows how the associated forces can be balanced in a turbine, helical gears and a cone clutch, with the help of the principle of direct and short force

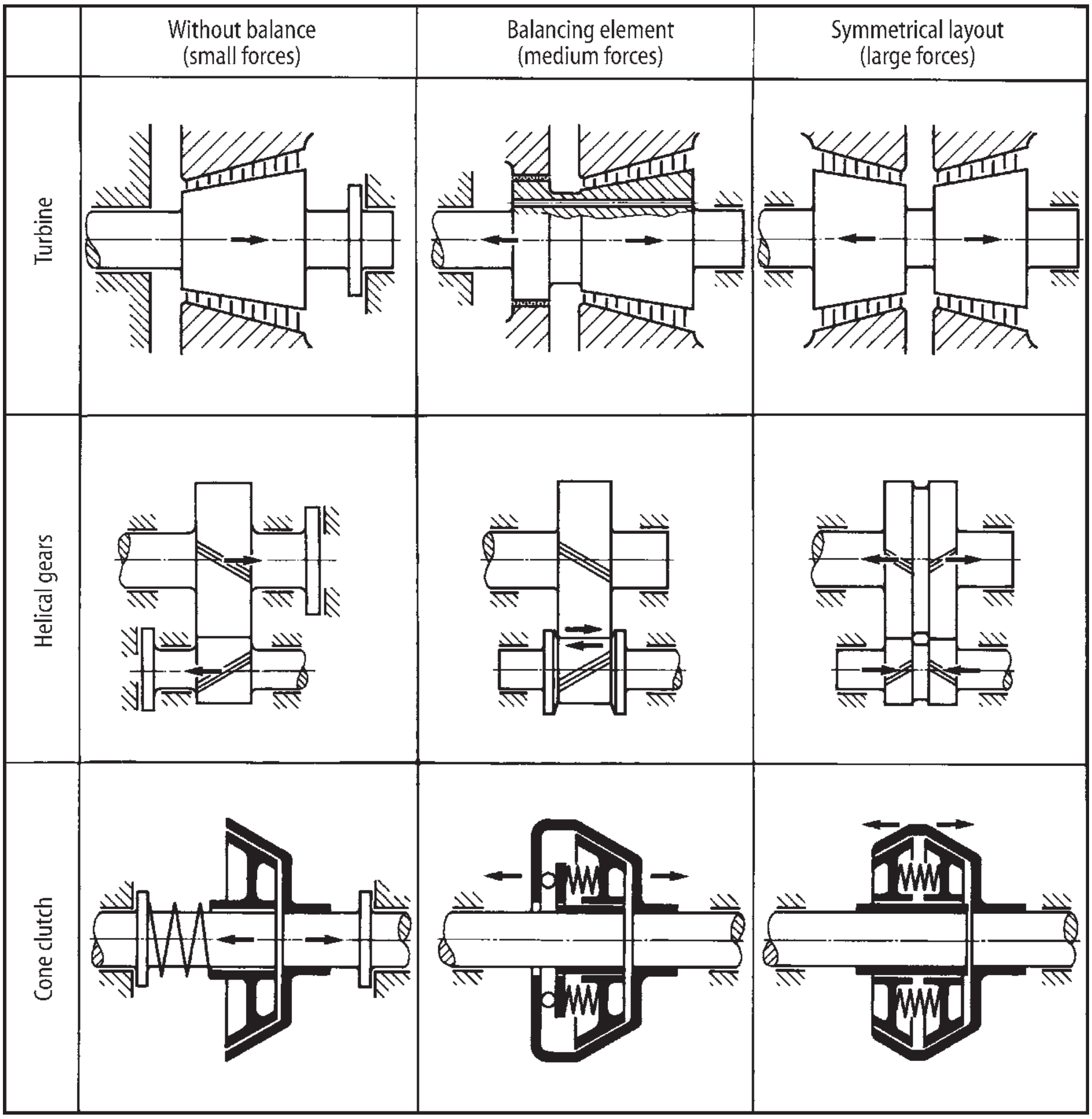


Figure 7.35. Fundamental solutions for balancing associated forces, illustrated via a turbine, helical gears and cone clutch

transmission path. As a result, no bearing position is loaded additionally and the designs are highly economical.

When it comes to the balancing of inertia forces, we find that a rotationally symmetrical layout is inherently balanced. The same solution principle is applied for reciprocating masses, as we know from automobile engineering. If the number of cylinders is too small to ensure a perfect balance, either special balancing elements, weights or shafts [7.228] are introduced, or cylinders are arranged symmetrically, as for instance in opposed cylinder engines.

As a general rule (which, however, can be ignored if there are overriding reasons for doing so), balancing elements should be chosen for relatively small or medium forces, and a symmetrical layout for relatively large forces.

5. Summary of Force Transmission Principles

Earlier we discussed the value of using the descriptive idea of flowlines of force when considering the transmission of forces during the embodiment of assemblies and components. The flowlines should fulfil the following criteria:

- the flowlines must always be closed
- the flowlines should, in general, be as short as possible, which can best be achieved by direct force transmission
- sharp deflections of the flowlines and changes in their "density" resulting from sudden changes in cross-section must be avoided.

In the case of complex force transmission situations, the definition or visualisation of a flowline envelope can be useful. This is the working zone outside of which the forces have no effect. The smaller the envelope, the shorter the force transmission paths. Figure 7.36 shows different concepts of a rotary bending test rig with the respective flowlines envelopes indicated.

The following principles complement the concept of flowlines:

- The *principle of uniform strength* which ensures, through the careful selection of materials and shapes, that each component is of uniform strength and contributes equally to the overall strength of a device throughout its service life.
- The *principle of direct and short force transmission path*, which ensures minimum volume, weight and deformation, and which should be applied particularly if a rigid component is needed.
- The *principle of matched deformations*, which ensures the matching of deformations of related components, so that stress concentrations are avoided and the function can be reliably fulfilled.
- The *principle of balanced forces*, which ensures, with the help of balancing elements or a symmetrical layout, that the associated forces accompanying the main ones are reacted as close as possible to their place of origin, so that material quantities and losses can be kept to a minimum.

In many situations, these principles cannot be applied to their full extent and often have to be applied in combination.

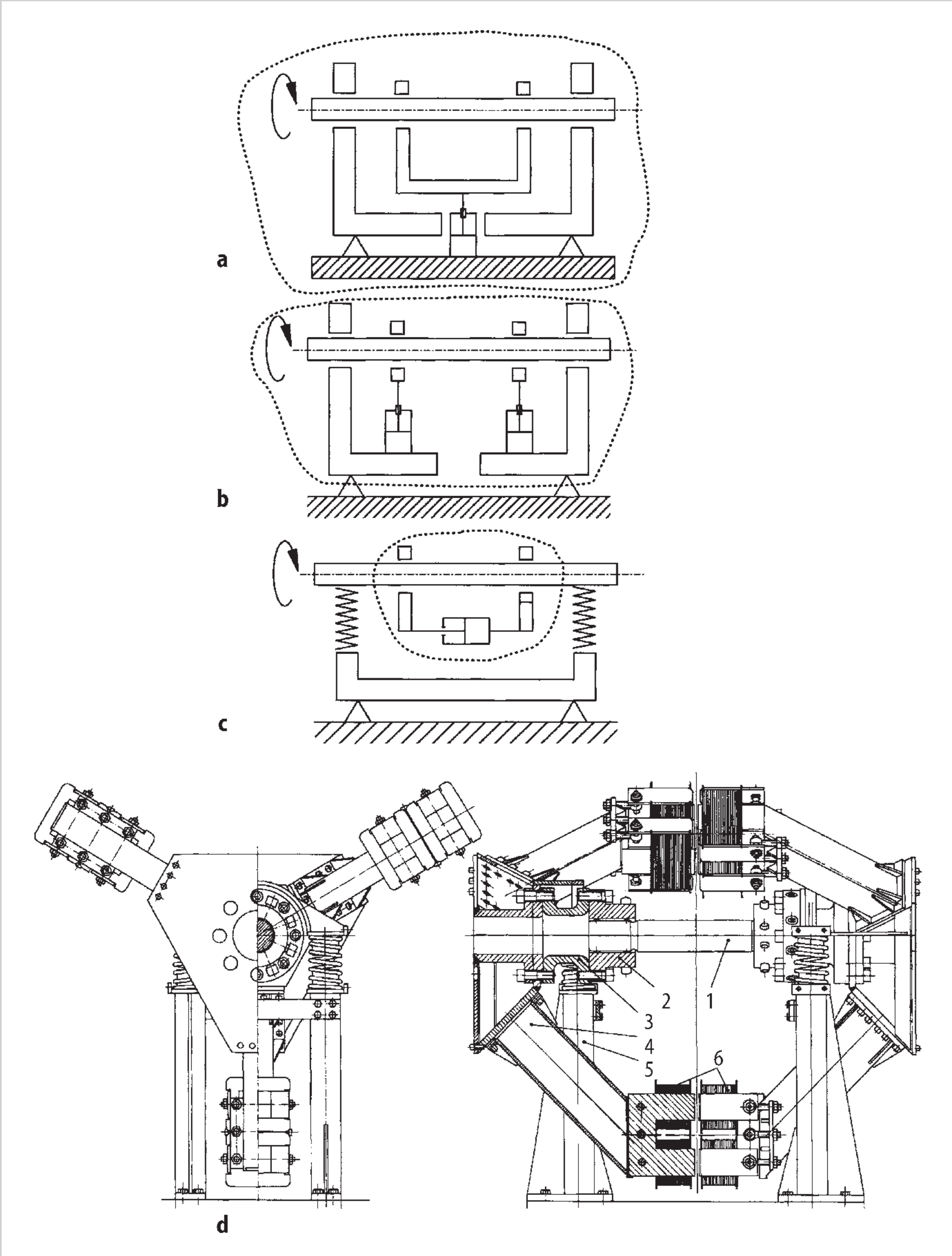


Figure 7.36. Force flow envelope (working zone of the forces) for a rotary bending test rig [7.330]. **a** Working zone includes the foundations; **b** working zone includes the supports; **c** working zone excludes the supports; **d** the test rig actually built using principle **c**, but with magnetic force excitation: 1 test shaft, 2 mounting flange, 3 connector, 4 support arm, 5 foundation supports, 6 magnet pair

7.4.2 Principle of the Division of Tasks

1. Assignment of Subfunctions

Even during the setting up and variation of the function structure, it is important to determine to what extent several functions can be replaced by a single one, or whether one function can be subdivided into several subfunctions (see Section 6.3).

These questions reappear in the embodiment phase, when the problem is to fulfil the requisite functions with the choice and assignment of suitable function carriers. We ask:

- Which subfunctions can be fulfilled with one function carrier only?
- Which subfunctions must be fulfilled with the help of several, distinct function carriers?

So far as the number of components and the space and weight requirements are concerned, a single function carrier fulfilling several functions would, of course, be the best. In terms of the production and assembly processes, however, this may prove disadvantageous, if only because of the complicated shape of the resulting component. Nevertheless, for economic reasons, an attempt should always be made to fulfil several functions with a single function carrier.

Numerous assemblies and components can fulfil several functions simultaneously or successively, as in the following examples:

- A shaft on which a gearwheel has been mounted transfers the torque and the rotating motion simultaneously, and, at the same time, takes up the bending moments and shear forces resulting from the normal tooth force. It also locates the gears axially and, in the case of helical gears, carries the axial force components from the teeth. In conjunction with the body of the gearwheel, it provides sufficient stiffness to ensure correct mating of the teeth.
- A pipe flange connection makes the connection and separation of the pipes possible, ensures the sealing of the joint, and transmits all forces and moments in the pipe resulting from residual tension, from thermal expansion and from unbalanced pipe loads.
- A turbine casing provides the appropriate inlet and outlet flow areas for the fluid, provides a mounting for the stationary blades, transmits the reaction forces to the foundation, and ensures a tight seal.
- A wall of a pressure tank in a chemical plant must combine a retaining with a sealing function and stave off corrosion, while not interfering with the chemical process.
- A deep groove ball bearing, apart from its centering task, transmits both radial and axial forces and occupies a relatively small volume.

The combination of several functions in a single function carrier may often prove economically advantageous, but may have certain drawbacks. These do not usually appear unless:

- the capacity of the function carrier has to be increased to the limit with respect of one or several functions
- the behaviour of the function carrier must be kept absolutely constant in one important respect.

As a rule, it is impossible to optimise the carrier of several combined functions. Instead, designers have recourse to the *principle of the division of tasks* [7.207], by which a special function carrier is assigned to every function. Moreover, in borderline cases, it may even be useful to distribute a single function over several function carriers.

The principle of the division of tasks:

- allows much better exploitation of the component concerned
- provides for greater load capacity
- ensures unambiguous behaviour, and hence fosters the basic rule of clarity (see Section 7.3.1).

This is because the separation of tasks facilitates optimum design in respect to every subfunction and facilitates more accurate calculations. In general, however, the constructional effort becomes correspondingly greater.

To determine whether the principle of the division of tasks can be usefully applied, the *functions must be analysed* with a view to determining if the simultaneous fulfilment of several functions in one carrier introduces constraints or mutual interferences. If it does, then it is best to settle for individual function carriers.

2. Division of Tasks for Distinct Functions

Examples from various fields illustrate the advantage of the division of tasks for distinct functions.

In large gearboxes, as found for instance between a turbine and a generator, it is advisable, because of thermal expansion of the foundations and bearings and also because of the torsional oscillations, to use a radially and torsionally flexible shaft whilst maintaining the shortest possible axial length on the output side [7.203]. However, because of the forces between the gear teeth, the transmission shaft must be as rigid as possible. Here the principle of the division of tasks leads to the following arrangement: the gearwheel is fitted to a stiff hollow outer shaft with the shortest possible distance between the bearings, while the radially and torsionally flexible component takes the form of an inner torsion shaft (see Figure 7.37).

Modern pressure-fed boilers are built with a membrane wall, as shown in Figure 7.38. The furnace must be gas-tight. Moreover, optimum heat transfer to the water demands thin walls with large surface areas. Beyond that, thermal expansion and pressure differences between the furnace and its environment must also be taken into consideration, and so must the weight of the walls. This complex problem is solved with the help of the principle of the division of tasks. The tubular walls with their welded lips constitute the sealed furnace. The forces resulting

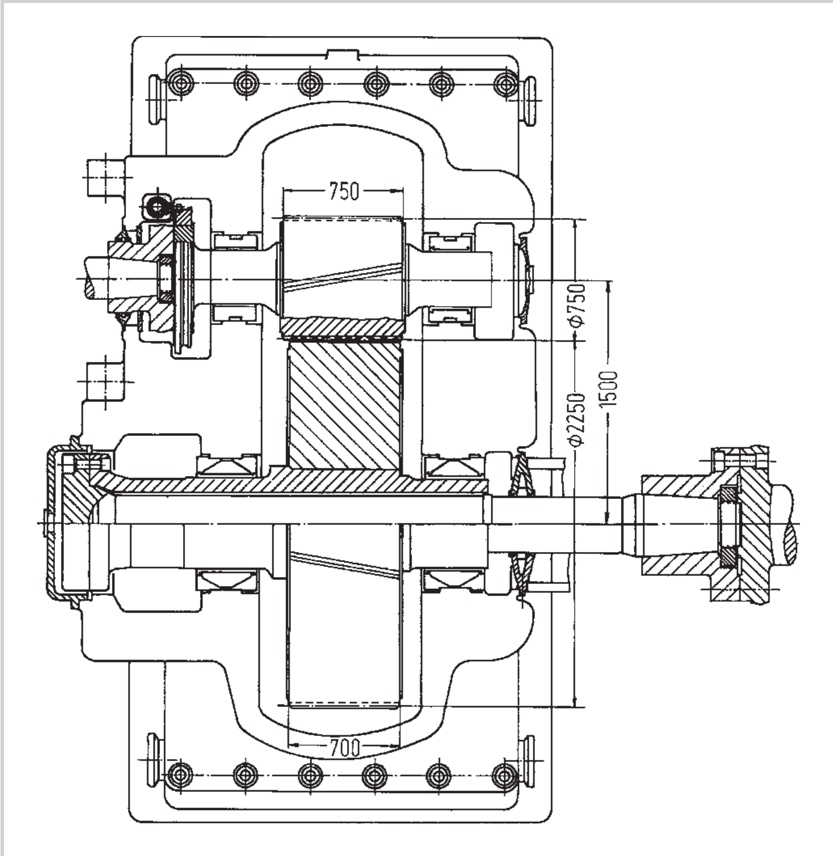


Figure 7.37. Large gearbox with an output torsion shaft; the bearing forces are transmitted over a stiff hollow shaft; the inner torsion shaft is radially and torsionally flexible, after [7.203] (Siemens-Maag)

from the pressure differences are transferred to special supports outside the heated area, which also carry the weight of the—usually suspended—walls. Articulated arms between the tubular wall and the supports allow for unimpeded thermal expansion. Thus every part can be designed in accordance with its special task.

The clamp connection in a superheated steam pipe shown in Figure 7.39 has also been designed based on the principle of the division of tasks. The sealing and load-carrying functions are assigned to different function carriers: the sealing function is performed by the welded membrane seal, which is axially loaded by the tension in the clamp. Tensile forces or bending moments should not be carried by the seal, whose function and durability would thereby be destroyed, so the load-carrying function is performed by the clamp which, in turn, is designed on the principle of the division of tasks. The clamp is made up of segments, which transmit forces and bending moments by means of a close-tolerance fit, and shrink rings hold the clamp segments together via friction in a simple and effective manner. Every part can he optimally designed for its particular task and is easily analysed.

The casings of turbines must ensure a tight seal under all operational and thermal conditions if they are to conduct the working fluid with minimum loss and turbulence. They must also provide an annular area and a support for the stationary blades. During temperature changes, sectioned casings with an axial flange have a particular tendency to distort and to lose sealing power due to

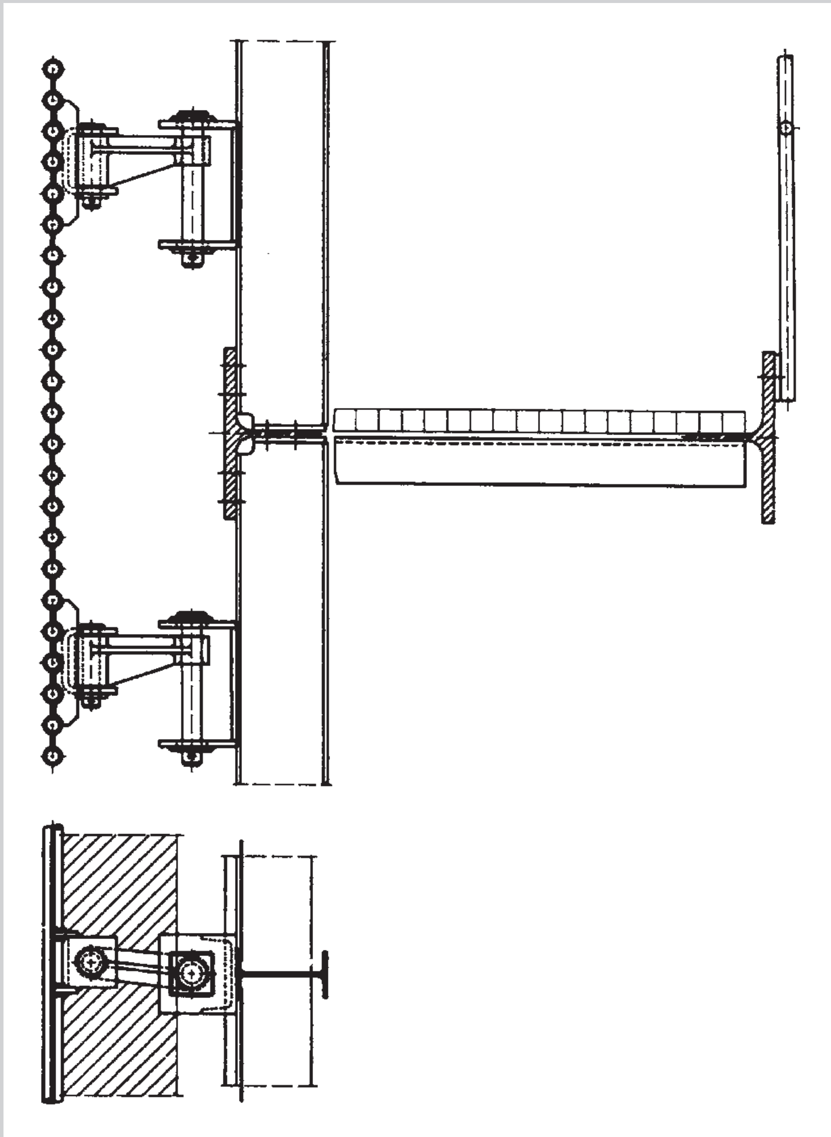


Figure 7.38. Section of boiler with membrane walls and separate supports (Babcock)

marked changes in shape at the inlet and outlet [7.224]. This effect can be offset by a separate blade carriers, that is, by a division of tasks. The annular area and stationary blade attachment can be designed regardless of the larger casing with its inlet and outlet sections. The outer casing can then be designed exclusively for durability and sealing (see Figure 7.40).

A further example is provided by the synthesis of ammonia, which involves feeding nitrogen and hydrogen into a container under high pressures and temperatures. If the hydrogen were allowed to come into direct contact with a ferritic steel container, it would penetrate into and decarbonise the latter, producing decomposition at the grain boundaries with the formation of methane [7.117]. The solution is again based on the division of tasks. The sealing function is provided by an inner casing of austenitic steel which is resistant to hydrogen, while support and strength are provided by a surrounding pressure chamber constructed from high-tensile ferritic steel, which is not resistant to hydrogen.

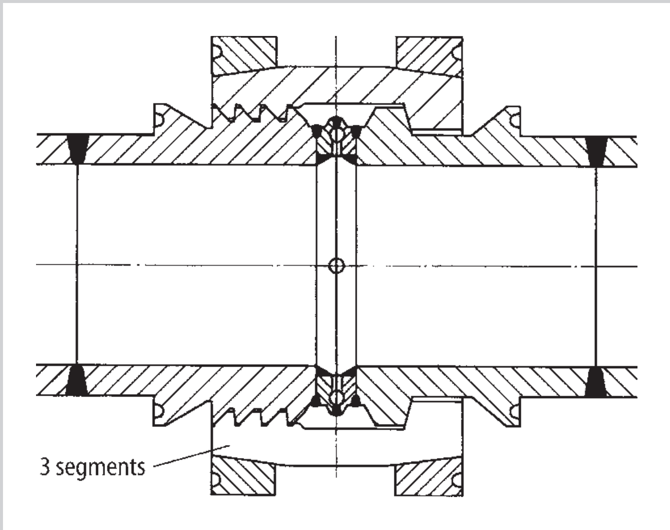


Figure 7.39. Clamp connection in a superheated steam pipe (Zikesch)

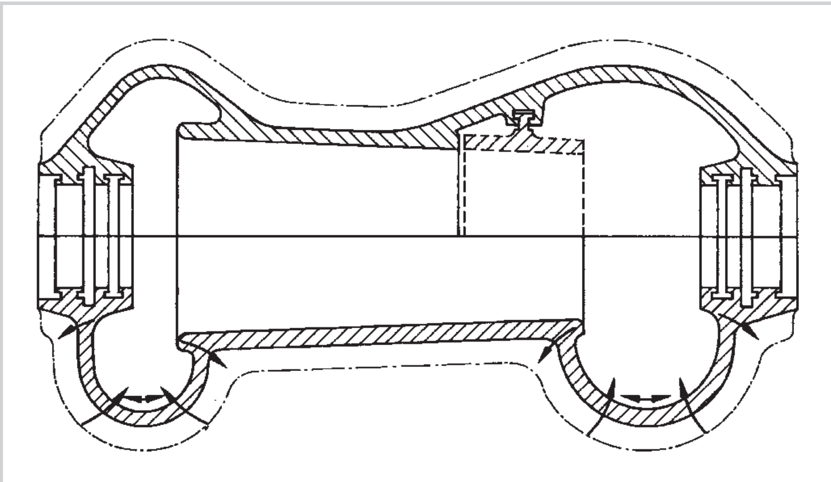


Figure 7.40. Axially divided turbine housing, after [7.224]: lower half conventional; upper half with separate blade carrier

In the electrical circuit-breaker illustrated in Figure 7.41, two or even three contact systems are provided. The breaker contacts 1 take the arcing current during the closing or opening of the switch, and the main contacts 3 carry the current under normal conditions. The breaker contacts 1 are subject to burning—that is, to wear and tear—and must be designed accordingly, while the main contacts must be designed to carry the full working current.

The division of tasks is also illustrated in Figure 7.42: the Ringfeder connectors carry the torque, while the corresponding cylindrical surfaces ensure the central location and seating of the pulley, which the Ringfeder connector cannot provide by itself when high accuracy is required.

A further example is provided by the design of rolling element bearings in which the service life of the locating bearing is increased by the clear separation of the transmission paths of radial and axial forces (see Figure 7.43). The outer race of the deep-groove ball bearing is not supported radially, and hence transmits axial forces only, while the roller bearing transmits radial forces only.

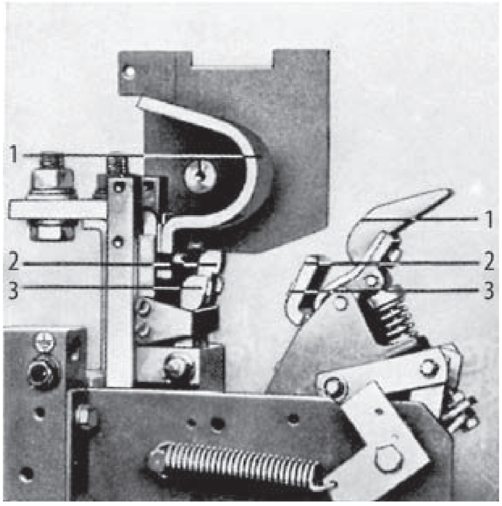


Figure 7.41. Arrangement of contacts in a circuit breaker (AEG): 1 breaker contacts; 2 intermediate contacts; 3 main contacts

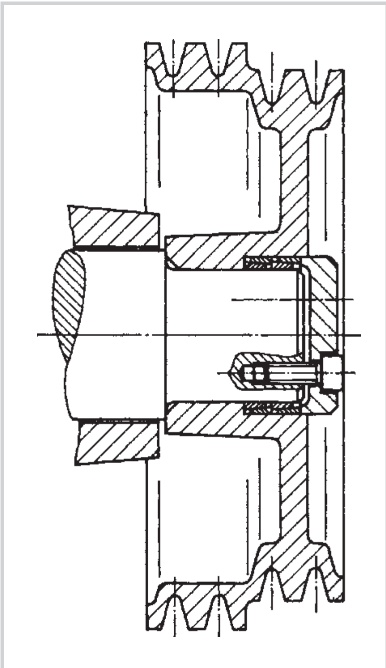


Figure 7.42. Ringfeder connector plus centralising surfaces

The principle of the division of tasks has been applied consistently to the construction of composite flat belts. They are made up, on the one hand, of a synthetic material capable of carrying high tensile loads and, on the other hand, of a chrome leather layer on the contact surface which provides a high coefficient of friction for the transfer of the load.

Yet another example is provided by the rotor blade attachment in a helicopter (see Figure 7.18).

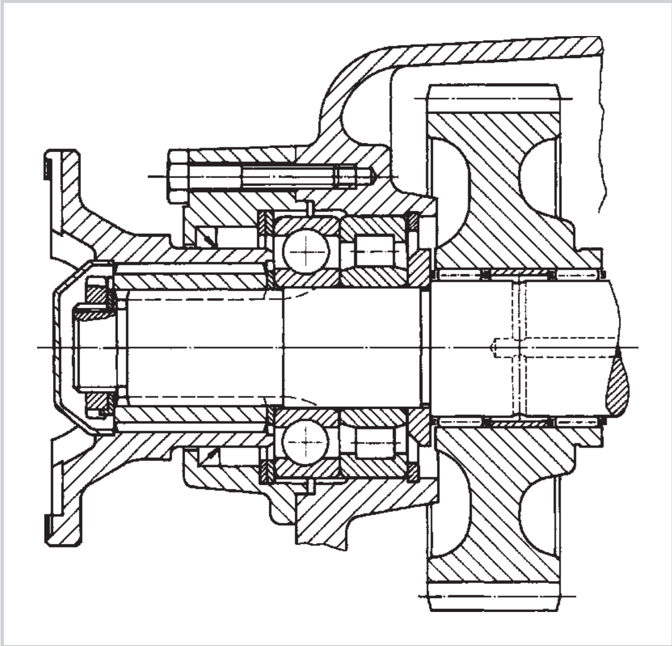


Figure 7.43. Locating bearing with separate transmission paths for radial and axial forces

3. Division of Tasks for Identical Functions

If increases in load or size reach a limit, a single function can be assigned to several, identical function carriers. In other words, the *load can be divided* and then recombined later. There are numerous examples of this.

The load capacity of a V-belt cannot be increased at will by increases in its cross-section (number of load-carrying strands per belt) because, for a given pulley diameter, an increase in the belt height *h* (see Figure 7.44) leads to an increase in the bending stress. As a result of the ensuing deformation, the rubber (which has hysteresis properties and is also a poor conductor of heat) becomes overheated and this reduces its life. A disproportionally wide belt, on the other hand, loses the stiffness needed to take up the normal forces acting on the wedge-shaped surfaces of the pulley. An increase in load-carrying capacity can, however, be obtained by dividing the overall load into part loads, each appropriate to the load limit and normal life of the individual belts (multiple arrangement of parallel V-belts).

The coefficient of thermal expansion of superheated steam pipes made of austenitic steel is approximately 50% higher than that of pipes made of the usual ferritic steel. Such pipes, moreover, are particularly stiff. At constant inner pressures and fixed material property limits, the ratio of outer to inner pipe diameter remains constant if the inner diameter is changed. However, while the throughput at constant flow velocities varies as the square of the inner diameter, the bending and torsional stiffnesses vary as its fourth power. The substitution of z pipe lines for a single large pipe would admittedly lead to increased pressure and heat losses for the same flow area, but would reduce the stiffness resisting thermal expansion by $1/z$. With four or eight pipelines, the individual reaction forces would then be no more than 1/4 or 1/8 of that present in a single pipe [7.29, 7.279]. In addition, the reduction in wall thickness leads to a reduction in thermal stresses.

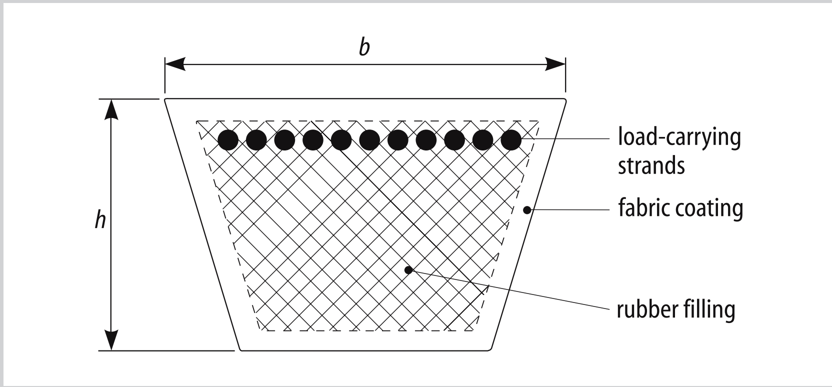


Figure 7.44. Cross-section of V-belt

Gearboxes, and epicyclic gearboxes in particular, make use of the principle of the division of tasks (or rather of forces) in the form of multiple meshing, which will increase the transmission capacity of the gearbox provided that the thermal effects can be kept within reasonable limits. In the symmetrical layout of epicyclic gearboxes based on the principle of balanced forces (see Section 7.4.1), even the bending moment in the shaft is eliminated because the forces produced by the gears cancel out. However, the torsional deformation is increased because of the greater load capacity (see Figure 7.45). In large gearboxes, this principle is applied to advantage in the form of multiple drives equipped with spur gears, which have external teeth only and hence are more easily produced. As Ehrlenspiel [7.96] has shown, it is possible to increase the load capacity with the number of force transmission paths, though not in direct proportion, because each step introduces a different flank geometry with a slightly greater flank loading. Basic arrangements are depicted in Figure 7.46.

One problem with the principle of the division of tasks is the uniform participation of all of the elements in the fulfilment of the function, that is, the provision of a *uniform distribution of forces or loads*. In general, this can only be achieved if:

- the participating elements adjust themselves automatically to balance out the forces
- appropriate flexibility is specially provided in the force transmission paths.

In the case of multiple V-belt drives, the tangential forces produce slight extensions of the belts which help to offset any dimensional errors in the lengths of the belts or in the pulleys, or any lack of parallelism in the shaft, and thus ensure equal load sharing.

In the case of the multiple pipeline discussed above, the individual pipe loss coefficients, the relationships between inflow and outflow, and also the geometry of the pipe layouts must be kept similar, or else the individual loss coefficients must be small and not greatly affected by the flow speeds.

In the case of multiple gears, either a strictly symmetrical arrangement must ensure equal stiffnesses and temperature distributions throughout the gearbox, or special flexible or adjusting elements [7.97] must ensure the equal participation of all of the components.

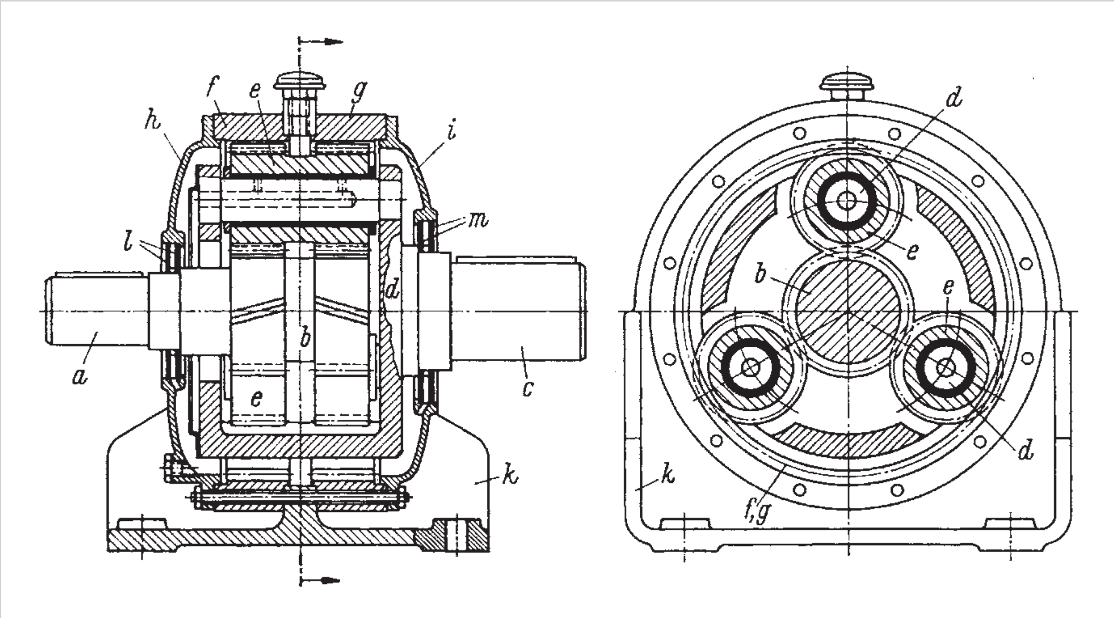


Figure 7.45. Epicyclic gearbox with balanced forces, after [7.97]

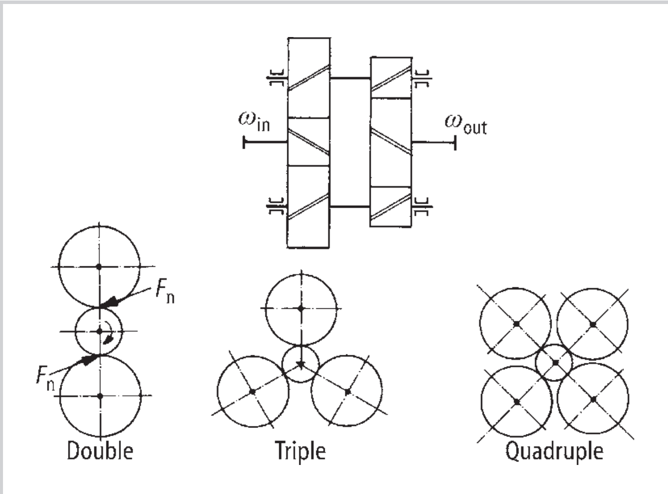


Figure 7.46. Basic arrangements of multiple gears, after [7.203]

Figure 7.47 illustrates a flexible arrangement. Further balancing components, such as elastic and articulated joints, are described in [7.97].

All in all, the principle of the division of tasks provides for increases in the maximum load capacity or for wider applications. By spreading tasks over several function carriers, we also gain a clearer picture of the relationship between forces and their effects, and, what is more, we can increase the output, provided only that a balanced division of forces is maintained by adjustable or self-regulating elements.

In supporting structures (such as bearing supports) where force transmission is divided, a more balanced load distribution can be achieved by adjusting the stiffness. During the stiffness analysis, the location and direction of the external forces have to be considered carefully, because they influence the deformation

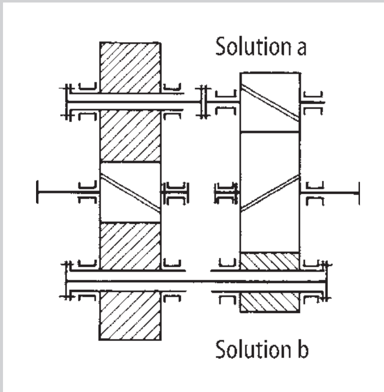


Figure 7.47. Balanced forces in multiple gears by means of flexible torsion shafts, after [7.203]

behaviour. This analysis can be facilitated by the use of Finite Element (FE) methods (see the principle of matched deformation in Section 7.4.1).

In general, the application of the principle of the division of tasks increases the number of components, which must be offset by greater overall economy or safety.

7.4.3 Principle of Self-Help

1. Concepts and Definitions

In the last section we discussed the principle of the division of tasks and showed how it could help to increase load capacity and provide a clearer definition of the behaviour of the components. To that end, we analysed the various subfunctions and assigned them to function carriers chosen such that they neither influence nor interfere with one another.

The same analysis can also be used in conjunction with the *principle of self-help* to achieve, through the appropriate choice of system elements and their arrangement, a mutual supportive interaction that improves the fulfilment of the function.

Under *normal conditions* (normal loading), self-help provides for greater effect by arranging the forces to work in the same direction as each other, or for relief by arranging the forces to offset each other. In *emergency situations* (overloading), self-help provides for greater protection and safety. In a self-helping design, the *overall effect* is made up of an initial effect and a supplementary effect.

The *initial effect* sets off the physical process required by the solution, but is insufficient on its own.

The *supplementary effect* is obtained from the functionally determined main forces (gearbox torque, sealing force, etc.) and/or from the associated forces (axial force produced by helical gears, centrifugal inertia force, force due to thermal expansion, etc.), provided, of course, that the two sets of forces are clearly correlated. A supplementary effect may also be obtained by appropriate changes to the type and distribution of the force transmission paths in order to increase load capacity.

The idea of formulating the self-help principle was first suggested by the Bredtschneider–Uhde self-sealing cover, which is particularly suitable for pressure vessels [7.237]. Figure 7.48 shows how it works. A relatively small force provided by the central bolt *2* suffices to press the cover *1* against the metal seal *5*. The initial effect of this force ensures that the parts make the proper contact. With increasing operational pressure, a supplementary effect is produced, which ensures that the sealing force between cover and tank is increased appropriately. The internal pressure thus provides the required sealing force automatically.

Inspired by this self-sealing solution, the principle of self-help was formulated in [7.206, 7.209] and further analysed and elaborated by Kühnpast [7.161].

It may be useful to specify the quantitative contribution of the supplementary effect S to the overall effect O in producing the degree of self-help:

$$\chi = S/O = 0 \dots 1$$

The gain from self-help solutions can be expressed in terms of one or several technical characteristics: efficiency, service life, use of materials, technical limit, etc. The self-help gain is defined as:

$$\gamma = \frac{\text{technical characteristic with self-help}}{\text{technical characteristic without self-help}}$$

Whenever the application of the self-help principle calls for a greater effort on the part of designers, then it must bring clear technical or economic advantages.

Identical design approaches may turn out to be *self-helping* or *self-damaging*, depending on the layout. Take the case of an inspection cover (see Figure 7.49). So long as the pressure inside the tank is greater than the pressure outside, the layout shown on the left is self-helping, because the pressure on the cover (supplementary effect) increases the sealing effect (overall effect) of the initial tension-screw force (initial effect).

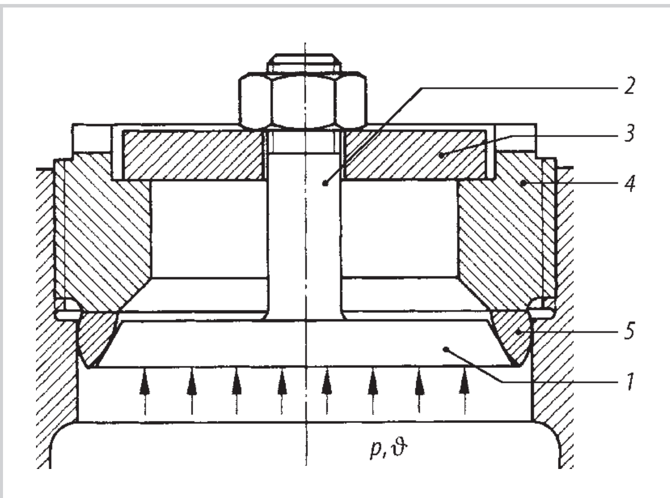


Figure 7.48. Self-sealing cover: *1* cover; *2* central bolt; *3* cross member; *4* element with sawtooth thread, *5* metal sealing ring; p = internal pressure, ϑ = temperature

The layout shown on the right, by contrast, is self-damaging because the pressure on the cover decreases the sealing effect O of the initial tension-screw force I. If, however, the tank were kept at below-atmospheric pressure, the left layout would be self-damaging, the right layout self-helping (see also Figure 7.50).

This example shows that the degree of self-help depends on the resultant effect: in the present case the effect on the sealing force resulting from the elastic forces, and not on the simple addition of the force exerted by the screw and the force acting on the cover. Figure 7.50 can also be considered to be a force–deformation diagram

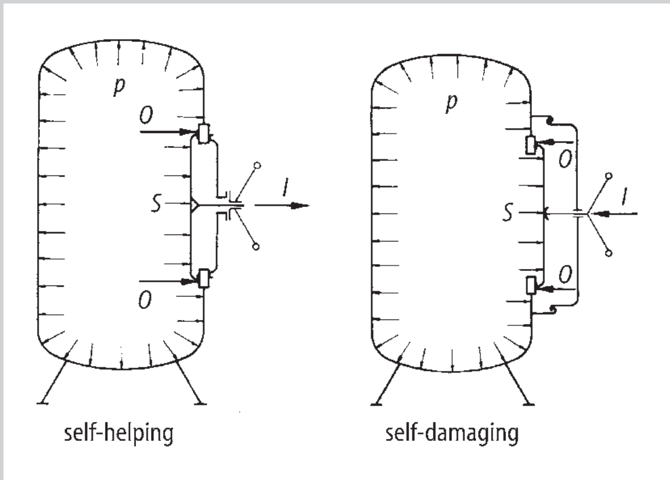


Figure 7.49. Layout of an inspection cover. I = initial effect; O = overall effect; p = internal pressure

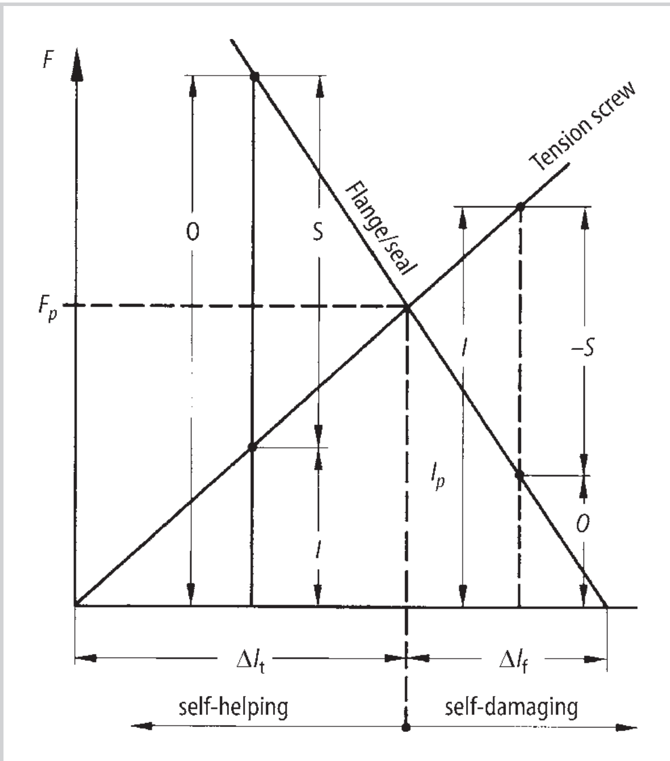


Figure 7.50. Force diagram for Figure 7.49: F = forces; F_p = preload; Δl = change in length; subscript t = tension screw; subscript f = flange/seal

Table 7.3. Summary of self-help solutions

	Normal load		Overload
Type of self-help	Self-reinforcing	Self-balancing	Self-protecting
Supplementary effect due to	Main and associated forces	Associated forces	Altered force transmission path
Important features	Main or associated forces act in the same sense as other main forces	Associated forces act in the opposite sense to main forces	Force transmission path altered by elastic deformation; limitation of function permissible

of a bolted connection with a preload and a working load. The conventional bolted flange connection may be called self-damaging inasmuch as, under operational conditions, the overall effect—that is, the flange sealing—becomes smaller than the preload. Also, the loading of the bolts is increased at the same time. If possible, therefore, only self-reinforcing arrangements that increase the overall effect while reducing the loading of the bolts should be chosen (Figs. 7.53a–d illustrate such arrangements).

For practical purposes, it is useful to classify self-helping solutions in accordance with Table 7.3.

2. Self-Reinforcing Solutions

In self-reinforcing solutions, the supplementary effect is obtained directly from a main or associated force and it adds to the initial effect to produce a greater overall effect.

This group of self-helping solutions is the most common. Under part-load conditions, it ensures greater service life, less wear, higher efficiency, etc., because the components are only loaded to an extent needed to fulfil the function at any particular moment.

As a first example, let us consider a continuously adjustable friction drive (see Figure 7.51).

The preload spring *a* presses the freely movable cup wheel *c* on the drive shaft *b* against the cone wheel *d*, thus providing the initial effect. Once a torque is applied, the roller follower *e* attached to shaft *b* is pressed against the cam *f* formed on the cup wheel *c*, where it produces a normal force F_n that can be resolved into a tangential force F_t and an axial force F_a, which, for its part, increases the contact force F_c applied to the cone wheel in a fixed proportion to the applied torque T:

$$F_a = T/(r \cdot \tan\alpha)$$

The force F_a represents the supplementary effect gained from the torque. The overall effect is obtained from the spring preload force F_p plus the axial force F_a,

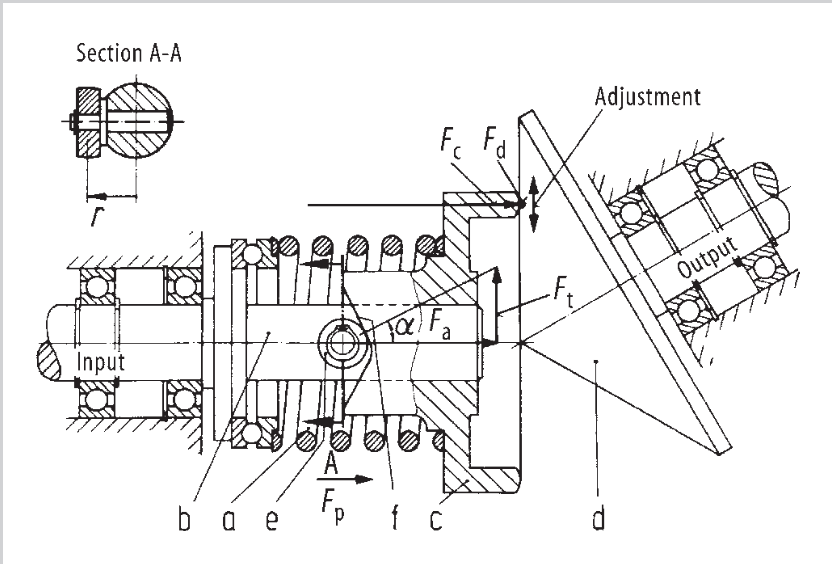


Figure 7.51. Continuously adjustable friction drive: *a* preload spring; *b* drive shaft; *c* cup wheel; *d* cone wheel; *e* roller follower; *f* cam formed on the cup wheel; *r* radius on which F_t and F_a act

which varies as the torque T (see Figure 7.52). The tangential driving force F_d on the cone, which determines the transmittable torque, is therefore:

$$F_d = (F_p + F_a) \cdot \mu$$

and the degree of self-help is:

$$\chi = S/O = F_a/(F_p + F_a)$$

It is obvious that the contact pressure between the wheels, which helps to determine the wear and the service life of the drive, must not exceed what is strictly necessary. A conventional solution (no self-reinforcement) would have demanded an axial force produced exclusively by the spring preload corresponding to the maximum torque, and would therefore have necessitated maximum pressure being applied to the contact area under all load conditions. As a result, the bearings would also have had to carry a considerably greater load, which would have led to a reduced service life or would have demanded a much heavier construction.

A rough calculation shows that if the actual loading is, say, 75% of the nominal maximum load, then the bearing load would be reduced by about 20% which, because of the exponential relationship of service life to load, can lead to the life of the bearings being doubled. In that case, with $n = 3$ the self-help gain with respect to the service life becomes:

$$\gamma_L = \frac{\text{Life with self-help}}{\text{Life without self-help}} = \left(\frac{C/0.8P}{C/P}\right)^n = 1.25^3 = 2$$

A typical example is provided by the SESPA drive [7.157].