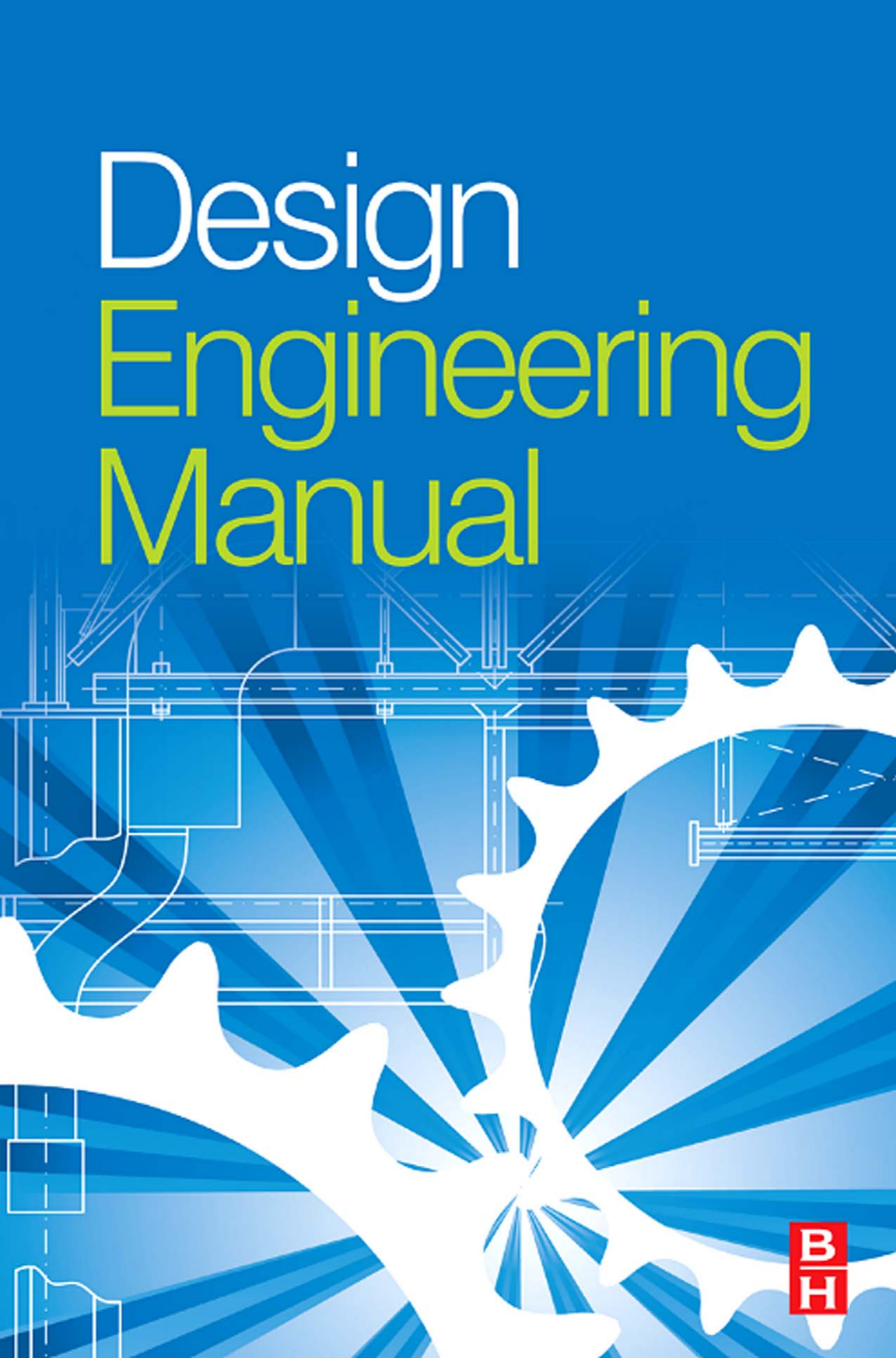
Design Engineering Manual



Design Engineering Manual

Note from the Publisher

This book has been compiled using extracts from the following books within the range of Design Engineering books in the Elsevier collection:

Annacchino (2007) *The Pursuit of New Product Development*, 9780750679930
Ashby (1999) *Materials Selection in Mechanical Design*, 9780750661683
Attwood, Deeb, and Danz-Reece (2003) *Ergonomic Solutions for the Process Industries*, 9780750677042
Booker, Raines, and Swift (2001) *Designing Capable and Reliable Products*, 9780750650762
Childs (2004) *Mechanical Design*, 9780750657716
Gilb (2005) *Competitive Engineering*, 9780750665070
Mills (2005) *Plastics: Microstructure and Engineering Applications*, 9780750651486
Mital (2008) *Product Development*, 9780750683098
Rosato and Rosato (2003) *Plastics Engineered Product Design*, 9781856174169
Smallman and Ngan (2007) *Physical Metallurgy and Advanced Materials*, 9780750669061
Smith (2005) *Reliability, Maintainability and Risk*, 9780750666947
Tooley (2006) *Electronic Circuits*, 9780750669238
Williams (2007) *EMC for Product Designers*, 9780750681704

The extracts have been taken directly from the above source books, with some small editorial changes. These changes have entailed the re-numbering of sections and figures. In view of the breadth of content and style of the source books, there is some overlap and repetition of material between chapters and significant differences in style, but these features have been left in order to retain the flavour and readability of the individual chapters.

Units of measure
Units are provided in either SI or IP units. A conversion table for these units is provided at the front of the book.

Design Engineering Manual

Amsterdam • Boston • Heidelberg • London • New York • Oxford
Paris • San Diego • San Francisco • Sydney • Tokyo
Butterworth-Heinemann is an imprint of Elsevier

Butterworth-Heinemann is an imprint of Elsevier
Linacre House, Jordan Hill, Oxford OX2 8DP, UK
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA

First edition 2010

Copyright © 2010 Elsevier Ltd. All rights reserved

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloguing-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 978-1-85617-838-9

For information on all Butterworth-Heinemann publications
visit our web site at elsevierdirect.com

Printed and bound in the UK

10 11 11 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries
www.elsevier.com | www.bookaid.org | www.sabre.org
ELSEVIER BOOK AID International Sabre Foundation

Contents

Section 1 INTRODUCTION 1

1.1 Introduction 3

Section 2 THE DESIGN PROCESS 17

2.1 The structure of the design process 19

2.2 Designing to ensure quality 43

2.3 Solving the requirements problem 57

2.4 Design specification 77

Section 3 PRODUCT DEVELOPMENT 101

3.1 The business objective 103

3.2 The market opportunity 135

Section 4 DESIGN AND MATERIALS 171

4.1 Materials and the environment 173

4.2 Materials and industrial design 187

4.3 Forces for change 197

Section 5 NEW MATERIALS 205

5.1 Advanced alloys 207

Section 6 DESIGN AND ERGONOMICS 231

6.1 Personal factors 233

6.2 Physical factors 249

6.3 Environmental factors 273

Section 7 RELIABILITY AND MAINTAINABILITY 293

7.1 Quality and reliability 295

7.2 Designing capable components and assemblies 319

Section 8 PRODUCT SAFETY AND LEGAL LIABILITY 369

8.1 Project management 371

8.2 Contract clauses and their pitfalls 377

8.3 Product liability and safety legislation 385

Section 9 DESIGN TECHNIQUES: ELECTRONICS 389
9.1 Circuit simulation 391
9.2 The PIC microcontroller 401
9.3 Introduction to EMC 415
9.4 The EMC directive 431

Section 10 DESIGN TECHNIQUES: MECHANICS 453
10.1 Introduction to mechanical design 455
10.2 Mechanical elements 473

Section 11 DESIGN TECHNIQUES: PLASTICS 481
11.1 Overview 483
11.2 Product design 507
11.3 Material and shape selection 583

Index 599

Section **One**

Introduction

This page is intentionally left blank

Chapter 1.1

1.1

Introduction

Anil Mital, Anoop Desai, Anand Subramanian and Aashi Mital

1.1.1 Introduction

Successful companies in the business world constantly operate in a state of innovation in terms of products they manufacture, frequently introducing new products or modifying and improving existing products as needed and desired by the customers. The overall process of conceptualizing a product and designing, producing, and selling it is known by a generalized and comprehensive process called *product development*. In this chapter, we discuss the initial steps of the product development process. The marketing and sales aspects of the product development process, while important, are mentioned in this chapter only in passing.

The key in new product development is the information that indicates what people want, what features of the product are considered absolutely essential, what price they are willing to pay for it, what features are desirable but can be sacrificed for a lower price, current and potential competitors, and likely changes in the market size. Knowing what the market needs is essential in order to develop innovative new products; this knowledge is what leads to developing a successful business strategy. Any product development strategy that is not based on market needs will lead to failure.

Before a successful product can be developed, someone has to come up with, or develop, an idea for conceptualizing it. There cannot be just one idea; several promising ideas need to be developed and analyzed before the detailed plans for a new business activity can be generated. Figure 1.1.1 shows the progression of actions in the development of a new business activity.

In developing the overall business strategy, a company has to develop and manage its entire product portfolio. Such a portfolio includes not only new-to-market products but also modifications of the existing line of products as well as products that are in the maturity part of

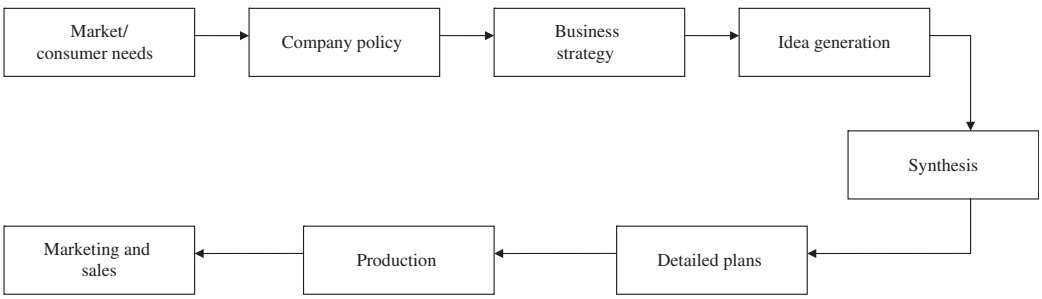


Fig. 1.1.1 Progression of actions in a new business activity.

Product Development; ISBN: 9780750683098
Copyright © 2008 Elsevier Ltd. All rights of reproduction, in any form, reserved.

sales. Concurrently, the company has to ensure that research and development of new technological platforms continue so that the transition from one S curve to another can take place smoothly.

In the following sections, we discuss the attributes of a successful product development process, what successful new products have in common, what steps are necessary to develop a successful portfolio of products, how to identify customer and market needs, and how to develop plans for a new product development.

1.1.2 Attributes of successful product development

Products that sell well and make a healthy profit (measured by the minimum attractive rate of return a business establishes for itself) reflect a successful product development process. However, sales grow slowly; consequently, it takes time to assess profitability. As a result, we must rely on the definition of manufacturing to establish the attributes of a successful product development process.

A business should develop the high-quality products the market desires quickly, economically, easily, and efficiently. This definition leads to the following attributes that define a successful product development process:

1. **Cost:** both the cost of producing the product and the total cost of developing it.
2. **Quality:** the quality of the product.
3. **Product development time:** from assessing market needs to product sale.
4. **Development of know-how:** the ability to repeat the process for future products.

The product cost determines its selling price and, to a large extent, its market attractiveness. This is not to say that price is the sole determinant of what the buyers find attractive about a product; cheaper but inferior-quality products tend to fall by the wayside. The price determines profitability; however, it is in this context that product cost is important. Product cost is a function of both fixed costs, such as tooling and capital equipment costs, and variable costs, such as material and labor costs. How much money the business spends on developing the product, from concept to prototype, also determines the profitability. A product does not become profitable until the development costs are fully recovered. Figure 1.1.2 shows the relationship between the cumulative cash flow and product life cycle.

Unless a product satisfies customers' needs and is considered dependable, it will not succeed in the marketplace. The quality of the product, therefore, is the ultimate determinant of the price customers are willing to pay for it. The share of the market a product gains is reflected by its quality.

How quickly a product makes it to the market also determines the overall economic returns on the investment and can be used as a measure of the success of the development effort. The quickness to the market, however, cannot be at the cost of the quality.

In contrast to "one-time wonders," development of successful innovative products one after the other

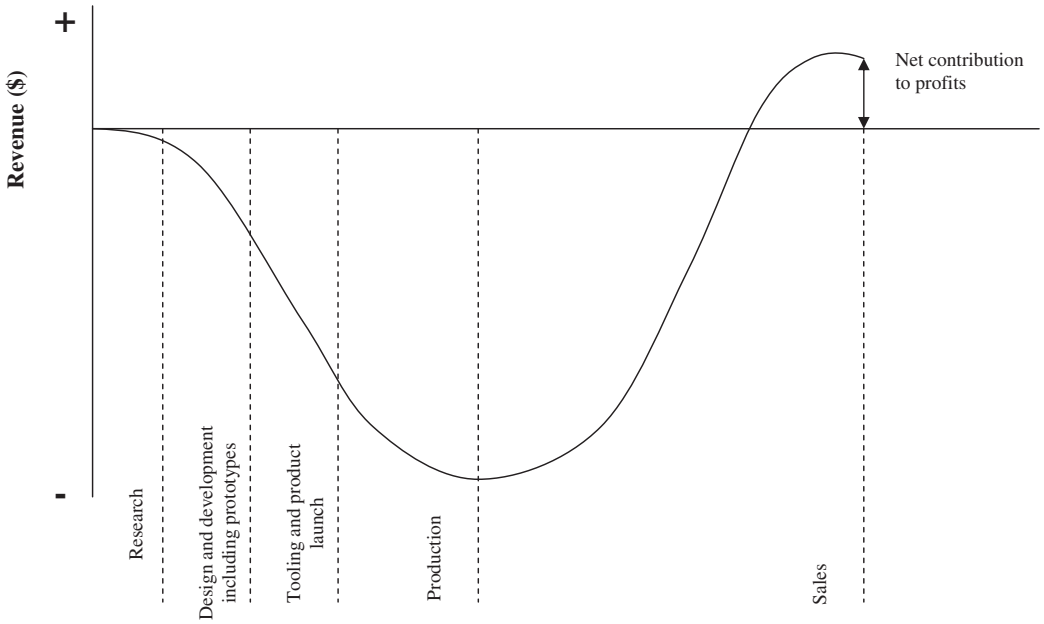


Fig. 1.1.2 Relationship between cumulative cash flow and product life cycle (modified from Corfield, 1979).

reflects the know-how a business has acquired over a period of time. Such companies are able to perform efficiently, effectively, and economically. This, in turn, is reflected in reduced development time, lower developmental costs, and products that capture significant market share and become profitable.

1.1.3 Key factors to developing successful new products

To succeed, a business must develop and market new products. However, not all new products that are developed succeed in the marketplace. What separates successes from failures? Cooper (1993, 1996) and Montoya-Weiss and Calantone (1994) identified several factors from a large number of studies that make new products winners. We discuss each of these factors briefly.

1.1.3.1 Uniqueness

Products that succeed in the marketplace are unique and clearly superior to reactive "me-too" products, which lack any distinguishable characteristics and provide only marginal benefits. The winners

- Provide excellent value for the money spent.
- Have excellent quality in comparison to their competition as perceived by customers.
- Meet customers' needs more fully than competing products and have unique features.
- Have highly visible useful benefits.

The product development process, therefore, must aim at developing products that are superior in value, distinct in features, and provide clear and unique benefits to the user.

1.1.3.2 Customer focus and market orientation

Focus on customer wants is critical to the development of successful products. Such focus improves success rates and profitability. To achieve a strong market orientation during the product development, businesses must

- Develop a thorough understanding of the nature of the market.
- Understand the competition.
- Devote resources to activities that determine customers' wants.
- Develop a relationship between product attributes and user needs.
- Seek customer input throughout the product design, development, prototyping, testing, and marketing (e.g., Boeing in the development of its 777 and 787 aircrafts).

The purpose of a strong market orientation is to leave nothing to chance by seeking customer inputs and incorporating them in product design.

1.1.3.3 Doing the homework

Work preceding actual product design is critical in determining whether a product will be successful. This includes the decision to proceed with the project, a quick study of the market for the product, technical assessment of the capabilities and requirements, market research, and the financial analysis (developing the per forma: capital needed, sources, potential sales, etc.). According to Cooper (1996), only about 7% of developmental money and 16% of development effort are devoted to these "critical" activities. This lack of attention to predevelopment work significantly increases the chances of product failure; company's name, reputation, and size of the sales force do not necessarily help if the "homework" is skipped.

1.1.3.4 Sharp and early product definition

Predevelopment work, or homework, leads to a sharp and early product definition and is essential for reducing the time to market. A product definition includes

- An outline of the concept and the benefits to be provided.
- A list of product attributes and features, ranked in the order "essential" to "desirable."
- A description of the potential users and attributes of the market (size, demographics, etc.).
- An outline of the business strategy.

In the absence of a clear product definition, the chances of failure increase by a factor of 3 (Cooper, 1996). A sharp product definition forces attention to predevelopment work and sets clear goals for product development. It also forces all parties involved in the development to commit themselves to the project.

1.1.3.5 Execution of activities

Product development teams that succeed consistently do a better job across the activities identified under homework and market orientation. These teams do not skip market studies and do undertake trial sells (using test markets to see how the products will fare). There is no rush to market to capture that illusive share which maximizes profits. Some exceptions to this practice exist, however; for instance, Sony, which primarily believes in "creating" a market rather than identifying one. In general, the completeness, consistency, and quality of

predevelopment work are crucial to reducing development time and achieving profitability. The quality of execution is not limited to predevelopment work, however; it has to be an integral part of all development activities, from concept development to delivery to market.

1.1.3.6 Organizational structure and climate

For product development teams to succeed, they must be multifunctional *and* empowered. This means that

- Teams comprise members from all basic functions: research and development, engineering design, production, quality, sales and marketing, and so forth.
- Each member of the team represents the team and his or her "function," not the department and its "territory."
- Teams devote most of their time to project planning and product development.
- The team members share excellent communication and are in constant contact with each other.
- The entire team is accountable for the entire project.
- The team is led by a strong and motivating leader.
- The company management strongly supports the project, the team, and the team leader.

Although these points seem obvious, many businesses do not get the message. It is important that the operating climate be supportive, recognize effort, and reward success. The urge to punish failure or discourage risk taking should be avoided. Further, the top management should trust the team and the team leader and avoid micromanaging the project; once it has appointed a strong team leader and picked qualified team members, it should provide proper encouragement and make the necessary resources available. The corporate structure should encourage employees to provide ideas for new products, new technological platforms, and new ventures.

1.1.3.7 Project selection decisions

Many companies are involved in too many projects at one time, scattering valuable resources among many candidate projects. However, not all projects are likely to materialize. Product selection helps narrow down the choice of projects so that resources may be directed to those projects most likely to succeed and become profitable. This requires making tough "go" and "no-go" decisions, where projects that have only marginal value are "killed" so that those with merit may get the necessary resources and focus so that good products are developed.

1.1.3.8 Telling the world you have a good product

Having a good product is not enough; it must be promoted properly in the marketplace. New products must be launched at appropriate forums and adequate resources must be allocated to market it. The launch and marketing efforts must be supported by a professional staff who can troubleshoot and service the product promptly if needed (ideally, if the product is designed properly and has high-quality built in, this would not be an issue). It would be foolhardy to assume that a good product will sell by itself by word of the mouth. The launching of the new iPhone by Apple is a case in point. It was launched with big publicity and the launch was covered by major media worldwide.

1.1.3.9 Role of the top management

As stated under the section "Organizational structure and climate," the primary role of the top management is to support the product development team and provide it with the necessary resources. The management must realize that lack of time, money, and human resources are the main causes of failure. The top management must also clearly articulate the strategy for the business as it pertains to the development of the new product. It must define the goals for the new product. These goals typically include types of product, percentage of market share to be captured, profits from the new product, technologies on which to focus, focus of research and development, and long-term goals. It is also worth emphasizing, again, that micromanagement can have a very negative effect, so can pushing the development team in the direction of a favorite project.

1.1.3.10 Speed without compromising quality

The quickness with which a product makes it to the market is an important determinant of profitability. The advantage of speed is lost, however, if it means compromising product quality or the quality of executing essential activities. Since time to market is important, it can be reduced without sacrificing quality by

- Performing many activities concurrently; the lines of communication among the team members and the management must be kept open.
- Mapping out the entire project development on a timescale, ensuring that all activities are given adequate time and the precedence of activities is not violated (some activities cannot be performed unless other activities are completed first, e.g., estimating

product cost without first completing details of materials and processes).
- Not violating this activity–time map.

It must be realized that violation of the time line represents a lack of discipline, needed resources, or both. Regardless, this means delaying the delivery of the product. Also, there is no assurance that time line will be violated only once. What stops it from being violated again and again once the process starts? It is better to redirect more resources to the project than violate the time line; the time line must be considered sacred.

1.1.3.11 Availability of a systematic new product process

Cooper (1993) outlined a stage-gate process adopted by many companies. This process formalizes the new product development process, from concept development to launch, by dividing it into logical steps (stages) with strict go and kill decision criteria (gates). These criteria are established by the project team and generally are listed in terms of deliverables for each stage. Each stage can include several concurrent activities, but each activity must meet certain criteria to proceed to the next stage. These criteria, or gates, serve as the quality control checkpoints (ensuring the quality of execution of activities) and cannot be violated or deferred to the next stage. According to Cooper, this sort of strict action to enforce the product development plan results in many advantages, including improved teamwork, early detection of failure, higher success rate, better launch of the product, and a shorter time to market.

1.1.3.12 Market attractiveness

The market for launching the new product should be attractive; however, this is easier said than done. Nevertheless, some market attributes can help identify an attractive market. Among the desirable market attributes are

- The market is large and the product is essential for customers.
- The market is growing rapidly.
- The market economic climate is pro-product (positive).
- The market demand for the new product is not cyclic (seasonal) or unstable.
- The customers are receptive to adopting the new product (typically younger customers) and can easily adopt the product in their lifestyle.
- The customers are more eager to try new products and less concerned about its price.
- The customers have sufficient disposable income.

While it is desirable to have an attractive market, the success of a new product is less sensitive to external environment than to what the development team does to understand the market and customer needs and incorporate them into the product design. These positive actions are more responsible for success.

1.1.3.13 Experience and core competencies

It is very unlikely that a business will succeed right away in a totally new area of expertise. Some experience in the basic technologies needed, management capabilities, knowledge of product category, market needs for the nature of the product, resources necessary for developing newer products in the area is necessary for success. This requirement may be termed *synergy* or *familiarity with the business*. In general, the stronger the fit between the requirements of the new product development and core competencies (expertise) of the business, the greater are the chances of success. Specifically, the fit must be in terms of

- Technical expertise, both in terms of production capabilities and future research and development.
- Management capabilities.
- Marketing, selling, and customer service resources.
- Market and customer needs.

In general, it is easier for a business to succeed if it is on a familiar territory. However, this does not mean that a business does not or should not venture in different fields. If such a need arises, it should be pursued cautiously. One way to acquire core competencies in another area is to acquire existing businesses in that area. This kind of action has become quite commonplace in the world today. Numerous examples of businesses acquired other businesses in the same and totally different fields. Sony's venture in the entertainment business, TATA business house, involving ventures ranging from automobiles to management consulting, and GE's involvement in manufacturing activities ranging from aircraft engines to medical devices are some prominent examples of companies acquiring core competencies by buying other businesses.

1.1.3.14 Miscellaneous factors

Some of the factors listed in this section have unexpected effect on the success of a new product. Among these factors are order of entry, innovativeness, and the nature of benefits.

The order of entry has mixed results as far as success in the marketplace is concerned. While a very innovative new product may have some initial success, in general it

is better to introduce a high-quality product rather than be the first on the market. A poor-quality product may capture some market initially, but the bad experience associated with it can have a lasting effect on the customers, which may not be possible to overcome, ever. Obviously, it is highly desirable that an innovative, high-quality product be the first to market—such as, for instance, the Chrysler mini van. This example also represents a situation where the competition simply followed with me-too type products instead of offering innovative variations, such as mini vans with two sliding doors; it was left up to Chrysler to come up with that variation. There is a myth that highly innovative products are risky. In fact, if a new product offers an innovative solution to customers' needs, there is no reason for it to fail. Some caution here is necessary: products that have less innovation or too much innovation are less likely to succeed than products with a moderate degree of innovation; customers are more likely to accept such products, treating products with little innovation as me-too products and those with a very high degree of innovation as fancy gadgets.

A business simply cannot introduce a product and, on the basis of price advantage alone, expect to succeed. Unless a new product provides good value for the price, it is bound to fail, as price alone is an inadequate benefit for success.

In summary, successful consumer products have the following attributes:

1. They offer entirely new benefits that existing products do not.
2. They offer a new secondary benefit in addition to new primary benefit.
3. They are comparative to what the competition offers.
4. They eliminate an important negative in existing market products.
5. They offer a higher-quality product than available in the market.
6. They harness contemporary societal trends.
7. They offer a price advantage in comparison to the competition.

1.1.4 Strategy for new product development

The primary objective in establishing a strategy and a business plan for developing a new product is to ensure that all concerned parties "buy into" the effort and a consensus is reached on the fundamental inputs to the plan. As mentioned earlier, these inputs include information regarding the market, sources of capital, business per forma, information about the nature of the product, and information about the market. However, before a development plan can be put together, certain activities must be performed in order to develop an overall new product development strategy. These activities include

1. Determining the company's growth expectations from the new products.
2. Gathering information of interest regarding capabilities, market, and the customers.
3. Determining what opportunities exist.
4. Developing a list of what new product options exist.
5. Setting criteria for inclusion of new product(s) in the company's portfolio of products.
6. Creating the product portfolio (new, modified, and existing).
7. Managing the product portfolio to maximize profitability.

1.1.4.1 Determining the company's growth expectations from new products

A company's mission typically provides some insight into its business objectives. The business objectives and the overall business plan delineate the role the company expects new products to play in its growth. The company 3M, for instance, expects products developed in the last 5 years to contribute 25% to its profits. The role of new products can be a similarly worded target, indicating what new products are expected to contribute to the overall business goals. Setting such a target is important in deciding what resources to direct to new product development. It also helps in reviewing what are company's technical and financial capabilities, what product concepts are within company's ability to develop and are attractive to its customers, what are the risks and how these risks can be spread by diversifying product portfolio, and how well the company's short-term and long-term goals are being met.

1.1.4.2 Gathering strategic information

While the company may already have information regarding its customer base, market needs, its business and technical capabilities, and the competition, it helps in periodically updating this information. New market research, information on emerging competition, development of new markets, updating internal documents on customer needs, and so on must be carried out from time to time. Most useful is the compilation of all this information in a meaningful form, for instance, comparing the company's technical capabilities and product and sales profiles with those of its competition (benchmarking). This building of corporate knowledge base is not

a one-time effort but a dynamic process and allows the company to constantly update its strategic and business plans.

1.1.4.3 Determining existing opportunities

As mentioned already, the challenge is to present the information gathered in a meaningful form so that strategic and business plans may be revised and new opportunities identified. The presentation of information should be such that different product options and opportunities are easily identified. Two tools are helpful in this process: a matrix scoring model and a map of the opportunities.

The matrix scoring model is useful in situations where a number of options are available and the best one must be chosen. An example of this kind of analysis is comparing different sites for locating a facility using a number of selection criteria. In choosing a potential product concept for development from among several possibilities, the concepts can be compared using a variety of criteria with weights assigned to each criteria. The scores for each criteria of each concept are added and the total is compared to make the final selection. Table 1.1.1 shows how the matrix scoring model works.

The scoring scale used in Table 1.1.1 ranges from 1 (poor) to 10 (excellent) and is somewhat arbitrary; a 5-point, 7-point, or scale with fewer or more gradations can be used. A larger scale with more gradations increases the sensitivity of the evaluation process; shorter scales with fewer gradations reflect lower evaluation sensitivity. The weights chosen for different criteria indicate relative importance of criteria with each other. One can use a 10-point total, a 100-point total, or different total points for weight as long as the distribution among the criteria is relative. The method also allows using as many criteria as one chooses as long as it is realized that more criteria reduce the relative importance of each, as the weight then gets distributed over a larger number.

The second method for identifying opportunities for new products requires developing a map of all events and trends and linking them on a time horizon. This method was developed by Motorola (Willyard and McClees, 1987) and helps a company to identify new product opportunities. Figure 1.1.3, for instance, shows opportunities for developing energy-efficient products by linking events and trends associated with global warming.

The event map can be used to make decisions regarding which products a company should pursue to achieve forecasted revenues (generated from existing products, modified and improved products, products from existing technological platforms, and totally new products). This is known as *gap analysis*; gaps in the forecasted revenue are to be filled by sales from new products (Figure 1.1.4).

1.1.4.4 Developing a list of new product options

After mapping trends and events and reviewing sales goals for new products, the company is ready to develop a list of new product ideas and options. These have to be consistent and compatible with core competencies and the strategic information gathered. The list should be as complete as possible so that all available product options may be considered before selection is finalized. The possible options should be listed in an easy to compare format (Table 1.1.1 shows one option). Important information includes, but is not limited to, information on the financials (investment, cost of production, etc.), risks, available technologies, production capabilities, status of the concept, uniqueness of the features, production goals, expected profitability, product life span, potential for derivatives, development team expertise, and synergy with existing products and programs.

1.1.4.5 Setting criteria for product inclusion in the portfolio

The company expects that including a new product in its portfolio of products will increase its sales revenue and profitability. Typically, all businesses expect a minimum

Table 1.1.1 Matrix scoring model

		Product concept scores		
Criteria	**Weight (*w*)**	**A**	**B**	**C**
Financial	3	$3 \times 3 = 9$	$2 \times 3 = 6$	$2 \times 3 = 6$
Customer needs	4	$8 \times 4 = 32$	$5 \times 4 = 20$	$7 \times 4 = 28$
Production ease	2	$4 \times 2 = 8$	$3 \times 2 = 6$	$5 \times 2 = 10$
Core competency	2	$3 \times 2 = 6$	$4 \times 2 = 8$	$8 \times 2 = 16$
Total score		55	40	60

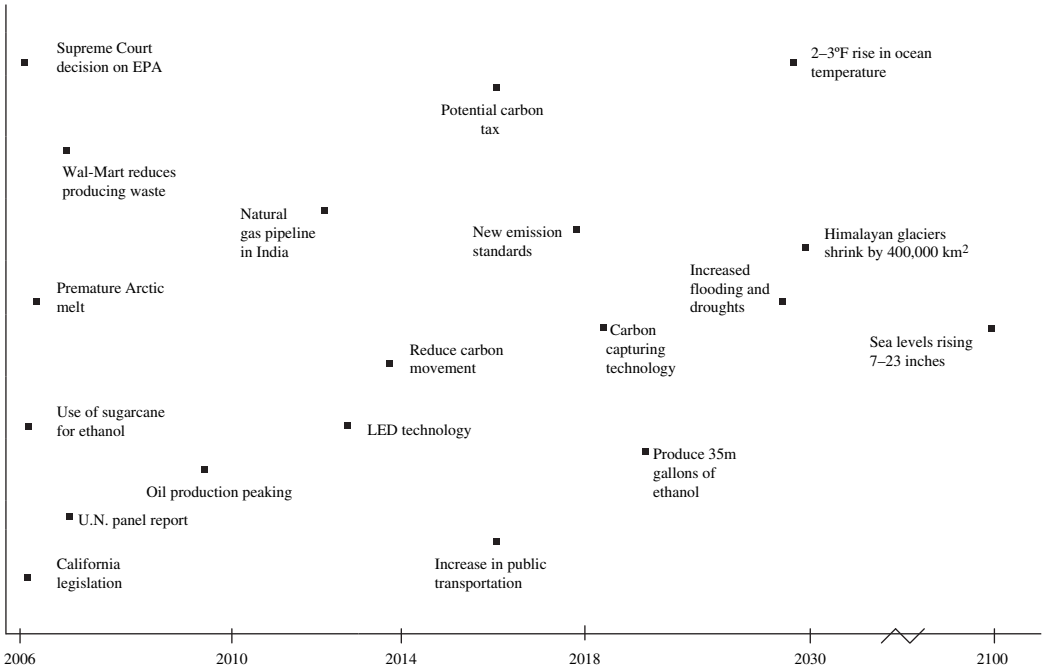


Fig. 1.1.3 Mapping of events and trends in global warming to point out opportunities for energy-efficient and alternative energy products (times are approximate; not all events are shown; some events are anticipated).

return on investment (ROI). If a product option fails to meet the economic criteria established by the company, it should not be considered any further.

In addition to economic criteria, the company should also look at how well the new product option conforms to its short-term and long-term goals. Can the company develop other products from the technology developed for this product? Is the product concept so new that it exposes the company to unacceptable risks? Are the investment requirements disproportionate? Would the

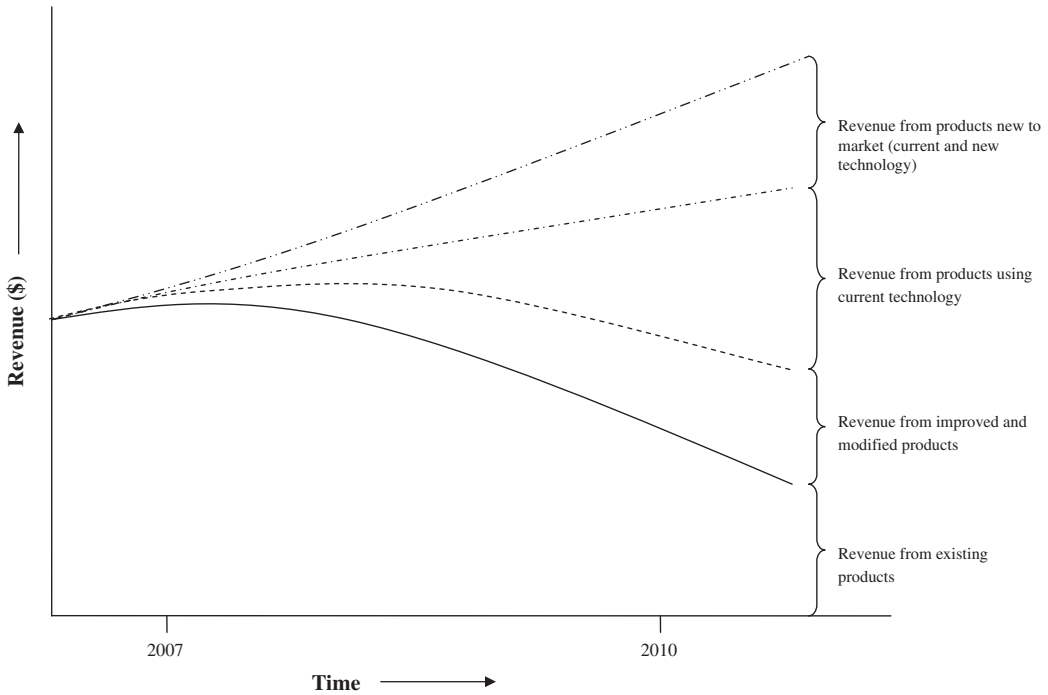


Fig. 1.1.4 Forecasted revenue to be generated from sales of new products.

option open new markets to the company? Answers to these kinds of questions help the company to develop portfolio criteria. For the final selection, both economic and portfolio criteria must be established and considered.

1.1.4.6 Creating the product portfolio

As shown in Figure 1.1.4, the company's product portfolio includes existing, modified, and new products. New products that meet the selection criteria fulfill the new product target, address customer and market needs, promote the company's mission, and meet its business objectives should be included in the portfolio. The final decision should be taken by appropriate people from management and the project development team and should be based on the best information available. A consensus among the participants is necessary so that everyone buys into the process.

1.1.4.7 Managing the portfolio

Managing the portfolio typically includes assembling the right product development teams, making resources available, ensuring that research and development efforts are focused in developing technological platforms from which new products can be developed, developing appropriate marketing and sales strategies, and the like.

1.1.4.8 Developing new product plans

Once the preceding activities (also known as *strategic development activities*) are completed, it is time to develop plans for a new product. The first step in this process is the development of a statement of customer needs. This, in turn, requires understanding the customer. In this regard, customer connection and the early definition of customer value are the two most important practices that separate high-performance companies from low-performance ones (Deck, 1994).

1.1.4.8.1 Understanding consumers and their needs

The key question here is, will the new product excite consumers into spending money? To answer this question, we have to understand the customers. We have to know at least the following:

- What are their critical needs and how well are these being met at present?
- Who are the consumers, what products do they use, how do they use products, and under what conditions are the products used?
- How are current market products received and perceived with respect to meeting their needs?
- How do consumers know that a product is working, what signals covey that a product is working?
- Are products being used for physical reasons (e.g., hammering a nail), emotional reasons (e.g., feel good products), or both?

Market research techniques help in understanding consumer needs. These techniques can be broadly classified into two categories: qualitative techniques and quantitative techniques.

Under qualitative techniques, the following techniques generally are grouped: focus group interviews, one-on-one or in-depth interviews, and in-home visits. The focus group approach is used most frequently. Typically, a group comprising 6–10 individuals of similar background and demographics, led by a moderator, discusses a topic. The one-on-one, or in-depth, interview focuses on one individual at a time in order to learn his or her habits, motivations, needs, and so forth. In-home visits allow for a closer interaction with consumers in their own environment. However, to be effective, households have to be representative of the general population or the population that is to be the focus of the new product development.

Qualitative techniques, while providing information that can be quite enlightening and creative in nature, are limited by small group size or numbers. In such cases, it is better to use quantitative techniques. Among the widely used quantitative techniques are brand image research, segmentation research, and conjoint analysis. The brand image research, using a "important" to "not at all important" scale, attempts to determine which attributes are most important to customers, their opinion of all key brands regarding these attributes, and which attributes tend to best predict the overall brand opinion. In segmentation research, consumers are grouped into segments and their response patterns examined. Conjoint analysis is based on the concept that a brand, product, or concept can be considered as a bundle of attributes that make some contribution to overall customer acceptability. In contrast to qualitative techniques, quantitative techniques provide objective and reliable information leading to understanding the consumer.

To test the concept idea among the consumers, it is important to know how many consumers are likely to buy the product in order to try it. This number of consumers who will try the product is determined as follows:

$$\text{Consumer trial} = \text{interested universe} \times \text{consumer awareness} \times \text{retail distribution}$$

The interested universe is determined by using the purchase interest scale (a weighted 5-point scale from "definitely would buy" to "definitely would not buy"). This universe is adjusted downward by multiplying by factors for the expected consumer awareness and retail distribution the company expects to achieve.

Table 1.1.2 Laundry detergent purchase trends in the United States (Information Resources, Inc., 1985)

Brand	Market share (%)	Repeat purchase (%)
Tide	25	71
Wisk	10	62
Bold	8	58
Era	6	55
Cheer	5	55
A&H	5	57
All	5	53
Ajax	2	50
Dash	1	54

Table 1.1.3 Customer needs statement for rechargeable electric toothbrush

Small, compact, good fit in hand, nonslip grip
Attractive modern styling
Easy to charge; charge should last at least 7 days
Solid base; to double as charger
Both 110 and 220 volt operation
Brushing head to have rotational and reciprocating movement
Interchangeable and variable-size cleaning heads
Price to be less than $100
Attractive colors and packaging
Brand name

Table 1.1.4 Customer needs statement for blender

Ability to puree
Ability to mix ingredients evenly
Ability to crush ice
Sturdy base to prevent tipping
Variable-speed motor (three to five speeds) and quiet operation
Attractive styling and availability in assorted colors
Easy to use
Easy to disassemble and clean
Dishwasher-safe components
Detachable, nonrusting blades
Weight less than 2.5 pounds
Motor not to overheat quickly
Pulse grind operation
Easy operating controls
Wide-mouth jar for easy loading
Pour spout on jar for easy pouring
Easy to read English and metric graduations on jar
40–60 ounce capacity
Spill-proof lid
Clear jar for easy visibility
Cord storage area
110/220 volt operation
Price under $35

It is also important to know how many consumers who try a new product will buy it. Consumer behavior research indicates that consumer buying behavior is both regular and predictable (Uncles et al., 1994). Their loyalty, however, is greater in the case of big brand name customers than small brand name customers, as shown by laundry detergent purchase data gathered by Information Resources, Inc., shown in Table 1.1.2.

The market research techniques described here should lead to a clear and concise statement of customer needs from a customer perspective. Table 1.1.3 shows an example of customer needs statement for electric toothbrush. Another example of customer needs statement is shown in Table 1.1.4.

1.1.4.8.2 Understanding the market

To properly evaluate the potential for new product success, it is necessary to understand the market. This market understanding should clarify how the product will benefit both the consumer and the company and should focus on the following factors:

1. Market fit with the overall mission of the company.
2. Synergy between the market and the company.
3. Attractiveness of the market.

While the first factor is obvious, synergy between the market and the company is determined by answering a series of questions, such as

Is the market new to the company?
Will the company have to learn a new business?

How well do its management, talent, and skills apply to this market?
Can the company use its technological skills in this market? Its production facilities?
Will the company have to learn new technologies?
Have the technologies been acquired by acquisition?
How well does the acquired management fit with the market?
Are significant capital investments required to enter the market?
Can the existing marketing and sales forces and strategies be used? Or, must new ones be developed?

A poor fit between the core competencies of the company (knowledge, experience, and capability bases) and unfamiliarity with the market is a recipe for disaster. Davis (1996) provides a couple of examples: Fruit of the Loom, a well-known garment manufacturer, introduced the Loom laundry detergent in 1977 and discontinued it in 1981 and Procter & Gamble acquired the Orange Crush business but had to withdraw from it as it never understood the role of bottlers in the market. Both examples indicate that it is important for survival that the relationship between the old and new business be properly understood.

Attractiveness of the market also is determined by answering a series of questions:

Is the market a large one? How large? Geographical and populations sizes?
What are the past market trends? Growing? Stagnant? Shrinking?
What competitors already are in the market? What are their strengths and weaknesses?
Does one competitor dominate the market? Which one?
What are the cost and sales profiles of the leading competitor? Marketing and sales strategies?
Technologies? Patents? Reaction to new competition?
How do the consumers perceive the competition? How well are their needs being met?
What market share is held by the competition?
What are the trends in consumer needs? What are the features of the new product?
How are the pricing and features likely to attract the consumers?
Is the timing of new product introduction good?

Responses to these questions can be weighted to determine the market attractiveness and the market share one expects to gain, initially and over a period of time.

1.1.4.8.3 Product attributes and specifications

Once the assessment of customer needs and market conditions is complete, it is time to develop product specifications. The elements that should be included in product specifications include the following (Rosenthal, 1992):

- **Performance:** primary operating characteristics of the product.
- **Features:** characteristics of the product.
- **Reliability:** mean time between failures.
- **Durability:** product life estimate.
- **Serviceability:** ease of repair, part replacement, maintenance.
- **Aesthetics:** look, feel, sound.
- **Packaging:** packaging requirements, labeling, handling.
- **Perceived quality:** subjective reputation of the product.
- **Cost:** manufacturing, servicing.

The details pertaining to all elements are tabulated, resulting in product specifications. Table 1.1.5 shows product specification for electric toothbrush.

Table 1.1.5 Product specifications for rechargeable electric toothbrush

Performance: Effective cleaning effects; 2-minute cleaning cycle time Long-lasting battery life (rechargeable); at least 30 minutes
Features: Capability to reach different areas with ease Small, lightweight design; not more than 1.3 ounce Ergonomic grip, comfortable; 1.5-inch circumference Exceeds the American Dental Association requirements for storage and replacement Timer Waterproof assembly
Reliability: Effective cleaning each use Cleaning head to last 12 weeks
Durability: All components to last 2 years under normal usage
Serviceability: Easy to replace head design Easy to replace floss design
Aesthetics: New ergonomic styling concept
Packaging: Small and compact box-packaging Lightweight package design and packing material; not more than 1 pound Attractive labeling and graphics Radio-frequency identifier (RFID) for added security and easy inventory tracking
Cost: Manufactured cost less than $80 Service and warranty cost less than $15

1.1.4.8.4 Schedules, resources, financials, and documentation

Developing an agreeable schedule is the next step in the product development planning effort. While the schedule and achievable milestones should be realistic, attention should be paid to competition and profitability realities. Tools such as the program evaluation and review technique, critical path method, and Gantt charts should be used to determine a realistic working schedule. The schedule also should be kept in mind in allocating and phasing resources (types, quantity, timing, etc.).

The next logical step is the development of financial data, also known as *product per forma* in some circles. The kinds of financial details needed are

- Developmental cost: investment, hours, people, and so forth.
- Cost of developing prototypes.

Table 1.1.6 Product plan for rechargeable electric toothbrush

Customer needs:
Small, compact, attractive styling
Easy to use
Easy to clean
Price less than $100

Key product attributes:
Rotating and reciprocating head design
Overall weight less than 1.3 ounce
Overall length not more than 8 inches
Polymer-based ergonomic contoured handle
Soft bristles, pressure sensor limiting force 1.5 newtons
Dual head: front, brushing; rear, tongue cleaning
Rechargeable battery with 1-hour life cycle
LCD screen for charge display, brushing cycle time, sanitization status
Quiet (less than 60 dBA)

Product financials:
Development costs, $1,781,000
Tooling and capital, $3,500,000
Manufacturing cost, $70.70
Distribution and administration costs, 15%
Margin for profit, 20%

Market and competition:
Gain > 25% of market share
Penetrate all leading retail chains

Development schedule:

Phase	Completion
1. Customer needs	February
2. Product concept	May
3. Product design	August
4. Prototype development	September
5. Manufacturing	October
6. Product release	December

Resource requirements:

Weeks	Phase
6	Marketing and product management
24	Design engineering
12	Computer-aided design
16	Manufacturing engineering
10	Quality and test engineering

Key interfaces:
Marketing
Research and development

Table 1.1.7 Product plan for blender

Customer needs:
Ability to puree, crush ice, mix, grate
Fast, quiet, and safe operation
Easy to use, disassemble, clean
Price less than $35
Heavy-duty motor
Clear, wide-mouth jar with spout

Key product attributes:
Mixes ingredients evenly
Overall weight less than 2 pounds
Overall height not more than 15 inches
Quiet (75 dBA) and fast operation
5-year life
Available in various colors
Automatic shutoff if motor reaches 150°F
Dishwasher safe, nonrust blade

Product financials:
Development costs, $575,000
Tooling and capital, $4,070,000
Manufacturing cost, $21
Service cost, $10
Distribution and administration costs, 15%
Rate of return, 10.4%

Market and competition:
Gain 5% of market share
Advertise on TV
Market to young
Use wholesale outlets

Development schedule:

Phase	Completion
1. Customer needs	January
2. Product concept	February
3. Product design	March
4. Prototype development	May
5. Manufacturing	August
6. Product release	December

Resource requirements:

Hours	Phase
500	Marketing
5000	Design engineering
4000	Computer-aided design
8000	Manufacturing engineering
2000	Quality and test engineering

Key interfaces:
Marketing in India and China
Research and development

- Capital costs, tooling costs, setup costs, training costs, and the like.
- Direct labor, materials, and overhead costs.
- Packaging, distribution, marketing, and sales costs.
- Product manufacturing cost estimate, selling price determination.
- Production volume, revenue, profits, and so on.
- Return on investment, profitability, time frame for recovering investment, and the like.

Putting all this information in a concise form, preferably on a single sheet of paper, is a product plan. Tables 1.1.6 and 1.1.7 show examples of product plans.

1.1.5 Summary

In this chapter, we attempted to lay out a step-by-step procedure for developing new-to-market products. The critical steps in this procedure are develop a consumer- and market-friendly business strategy, identify consumer needs, and recognize market conditions. The outcome of the process is a concise, preferably a single page, product development plan. It should be recognized that this process requires the participation of a number of individuals with a multitude of expertise. It would be foolhardy to think that one individual, single-handedly, could accomplish this task. No matter how capable an individual, the key lies in recognizing that people must buy into the final outcome and the outcome must have synergy with the company's mission and business objective.

It is also worth mentioning at this point that information technology now is the dominant technological item with which the younger generation defines itself. How this technology is handled in the development of new products and how it is customized, to a considerable extent, determines the success of a company in developing products people like. For instance, cellular telephones today are fashion items: what sort a person has defines that person. People nowadays have a tendency to replace their cell phones long before they wear out. In that regard, cell phones are like cars. These products not only define people, they bring them together and, for the younger generation, serve as symbols of independence, lifestyle, and mobility. Cars and cell phones are examples that suggest how the development of products may proceed in the future. Here, one should consider at least two phenomena: the product features are as much a function of social factors as technological factors and there is no convergence in product design.

The first phenomenon dictates that, as technology develops and becomes more affordable, features once limited to high-end items, such as touch screen and cruise control in cars, become widely available. Color screens now are common features in cell phones and cameras. The second phenomenon, convergence in product design, suggests that there is no point in looking for an ideal product. Should all cars converge in the direction of a single ideal car—and look identical? How about telephones, cameras, televisions, lamps, furniture? People are different with different needs, tastes, and preferences. As long as individuals differ, there will be a need for diverging products. The question is, how do we manufacture divergent products that address consumer needs and make these products profitable?

References

Cooper, R. G. *Winning at New Products: Accelerating the Process from Idea to Launch*. Reading, MA: Addison-Wesley, 1993.

Cooper, R. G. New Products: What Separates the Winners from the Losers. In *The PDMA Handbook of New Product Development*. New York: John Wiley, 1996.

Corfield, A. P. National Economic Development Office. *Product Design* (The Corfield Report). London: National Economic Development Office, 1979.

Davis, R. E. Market Analysis and Segmentation Issues for New Consumer Products. In *The PDMA Handbook of New Product Development*. New York: John Wiley, 1996.

Deck, M. Why the Best Companies Keep Winning the New Product Race. *R&D* (November 1994): 4LS–5LS.

Montoya-Weiss, M. M., and R. Calantone. Determinants of New Product Performance: A Review and Meta-Analysis. *Journal of Product Innovation Management* 11, no. 5 (1994): 397–417.

Rosenthal, S. *Effective Product Design and Development*. Homewood, IL: Business One Irvin, 1992.

Uncles, M. D., K. A. Hammond, A. S. C. Ehrenberg, and R. E. A. Davis. A Replication Study of Two-Brand Loyalty Measures. *European Journal of Operational Research* 76 (1994): 375–384.

Willyard, C. W., and C. W. McClees. Motorola's Technology Roadmap Process. *Harvard Business Review* (1987): 13–19.

This page is intentionally left blank

Section **Two**

The design process

This page is intentionally left blank

Chapter 2.1

2.1

The structure of the design process

Anil Mital, Anoop Desai, Anand Subramanian and Aashi Mital

2.1.1 What is design?

Design is the act of formalizing an idea or concept into tangible information. It is distinct from *making* or *building*. Taking the concept for an artifact to the point just before the process of converting it into a physical, or embodied, form begins may be termed as the process of designing it. According to Caldecote (1989), design is the process of converting an idea into information from which a product can be made.

From an engineering perspective, the application of scientific concepts, mathematics, and creativity to envision a structure, a machine, system, or artifact that performs a prespecified function is the definition of *design*. *Design* is used pervasively; its meaning being somewhat different for an engineer than an industrial designer. While an engineer is more concerned about the arrangement of parts, the mechanics of the arranged parts, and their functionality when put together, an industrial designer is more concerned about the appearance of an artifact. Since, in designing consumer products, both form and function are important, both disciplines (engineering and industrial design) are crucial in the development of the final information from which a product can be made. The degree to which a product design depends on engineering or industrial design is determined by the product itself. A product that relies mostly on aesthetics, such as textile products, greeting cards, and furniture, is within the design spectrum of an industrial designer while products that are function dominant, such as automobile engines, building foundations, and gear trains, are within the domain of engineers. Consumer products depend on both engineers and industrial designers for success, on engineers for function and on industrial designers for aesthetics. The degree to which each discipline dominates the design varies from product to product. Figure 2.1.1 shows the design spectrum for both disciplines.

In general, the relative cost of products determines the extent to which a discipline contributes in the overall design. The design of a fighter plane is going to depend more on the principles of aerodynamics than on requirements of appearance or what pleases the eye. The design of fabric, on the other hand, is dominated by needs of appearance. Most consumer products, however, have significant contribution of industrial and engineering design content. Figure 2.1.2 shows the relative cost shares of engineering and industrial designs in a typical product design. Generally, engineering design costs tend to be 10- to 100-fold or more than industrial design costs, due to the very nature of engineering design: investment in functionality, reliability, and the like.

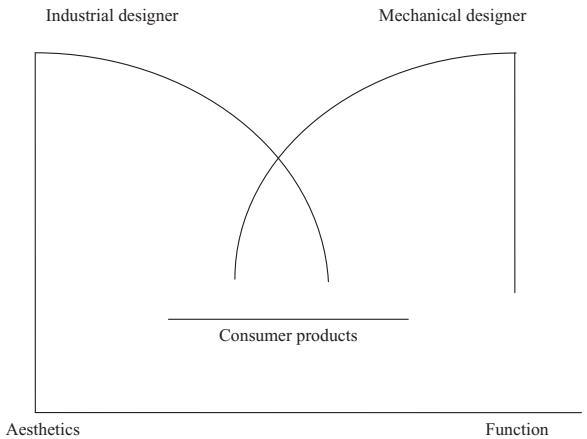


Fig. 2.1.1 The design spectrum for most consumer products (adapted from Caldecote, 1989).

Product Development; ISBN: 9780750683098
Copyright © 2008 Elsevier Ltd. All rights of reproduction, in any form, reserved.

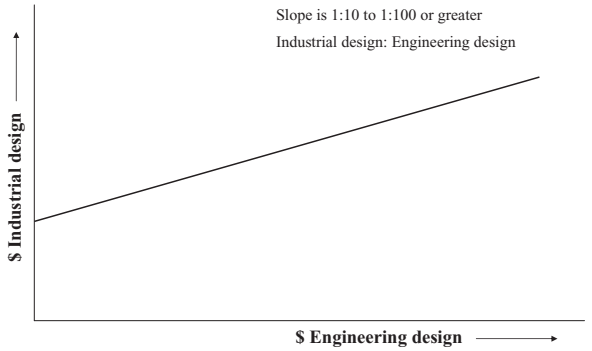


Fig. 2.1.2 Relative Costs of Engineering and Industrial Designs (adapted from Caldecote, 1989).

The outcome of the design process is the information that can be used to build it. The format of this information has changed over the years. It used to be informal drawings, leading to formal blueprints, to design drawings, to drawings in electronic format (computer-aided design (CAD)) that can be stored on electronic media. Some of the information formats, such as CAD drawings, can be fed directly to machines to produce the object (e.g., CAM–CAD interfaces).

2.1.2 The changing design process

In a primitive society, people designed things without being conscious of the effort. Stone-Age tools, doors to mud dwellings, and protection for the feet are examples of products that were developed with no formal drawing or awareness of the design process. Things were created without anyone designing them; the existence of technology was not a necessity.

The modern design process emerged with the growth of the industrial society. While design by drawing or sketching existed in some form or the other for more than 5000 years, it became more formalized with time. Figure 2.1.3 shows the sketch of a flying machine by Leonardo da Vinci. With the growth of the industrial society, this process has become more sophisticated. Figure 2.1.4, for example, shows a typical engineering drawing from the twentieth century. In the last few decades, humans have been aided by computers; computer-aided drawings, such as the one shown in Figure 2.1.5, have become commonplace. These days, the design drawings need not even be on paper; storage of drawings on electronic media has become routine.

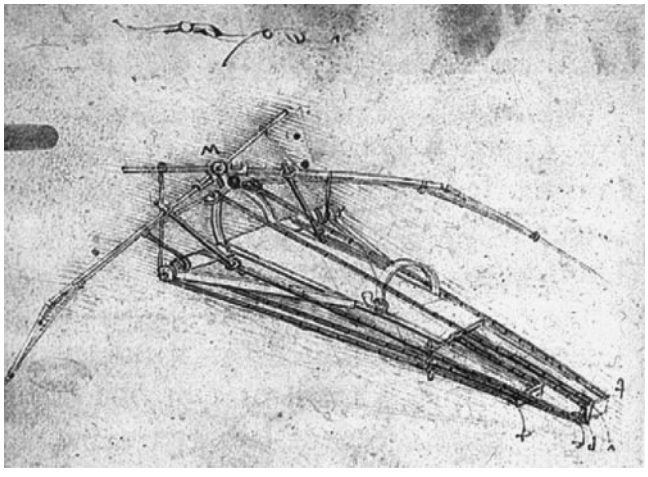


Fig. 2.1.3 Leonardo da Vinci's flying machine.

It has been argued, for instance by Jones (1970), that design by drawing provided a designer much greater flexibility to manipulate the design compared to the preindustrial society craftsmen. This flexibility allows manipulation of the design without incurring the high cost of building the product then changing it. The design-by-drawing process, however, does not guarantee success. According to Alexander (1969), this self-conscious process is limited, creates misfits and failures, and cannot replace centuries of development and adaptation in the preindustrial society that led to necessary inventions. Further, the extent of innovation required is beyond the average design-by-drawing designer and, therefore, a new design process, something totally different from design by drawing, is needed. Jones (1970) recommends that we shift our emphasis from product design to system design to avoid the design failures that created large, unsolved problems such as pollution and traffic congestions. Jones recommends a hierarchy of design levels, from components to products to systems to the community.

The views of Jones are foreshadowed by Schon (1969; quoted by Cross, 1989) who argued that we are amidst the postindustrial emergence as indicated by the following elemental shifts:

- from component to system to network;
- from product to process;
- from static organizations and technologies to flexible ones;
- from stable institutions to temporary systems.

In the context of design process changes, this means design becomes a central corporate function open to all in the corporation, instead of just the designer. This externalization of the design process allows all stakeholders (e.g., users and business) to see what is going on and to contribute to it in a way that is beyond the capabilities, knowledge, and experience of the designer. In a way, this shift is as significant as the shift from craftwork to design by drawing. For the corporation, such a shift means that it need not commit itself to a single product line or technology; its commitment could be to a major human function and the technologies and organizational relationships necessary to carry it

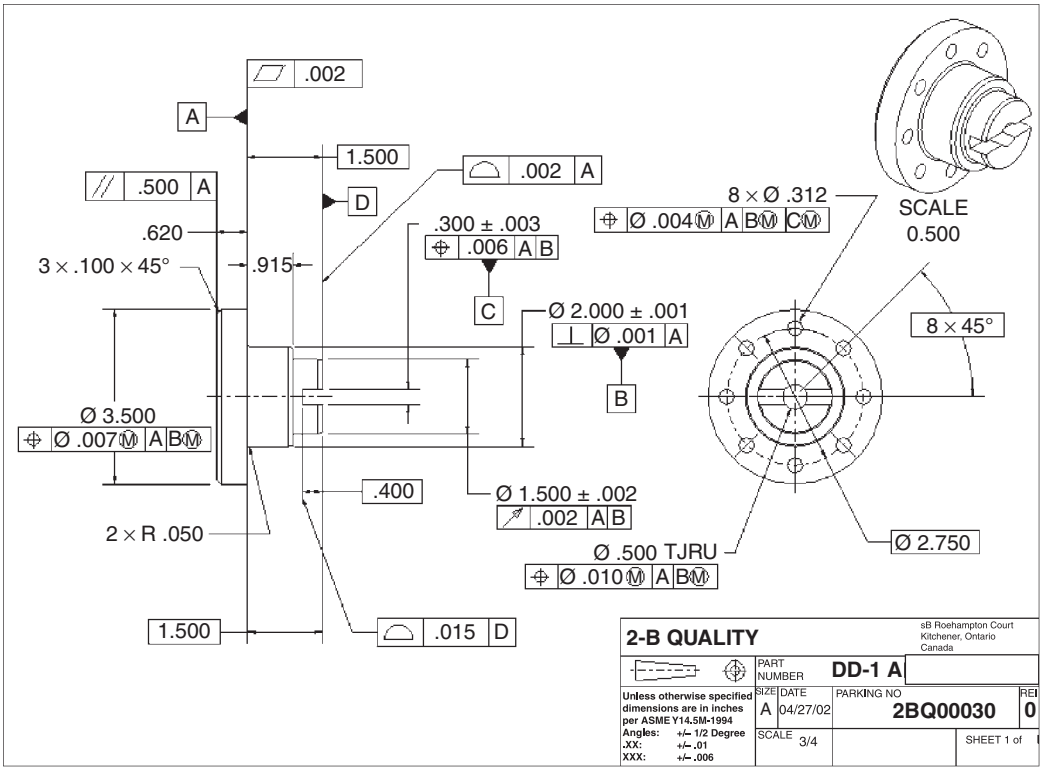


Fig. 2.1.4 A typical twentieth-century engineering drawing.

out. Both Schon and Jones argued that we are in the midst of this changing design process, and increasingly the designers have to focus on designing the systems and subsystems (e.g., transportation systems instead of automobiles) instead of just products, which become obsolete. Corporations such as General Electric, Siemens, and Boeing, to some degree, reflect this changing design process philosophy.

2.1.3 Design paradigms

Design paradigms are models, or quintessential examples of designed solutions to problems. The term *design paradigm* is used within engineering design to indicate an archetypal solution. Thus, a Swiss Army Knife is a design paradigm illustrating the concept of a single object that changes configuration to address a number of problems. Design paradigms can be used either to describe a design solution or as an approach to design problem solving. In this section, we briefly discuss the following important concepts (based on Petroski, 1994) through some unfortunate but classic examples:

- the need for a model,
- the need for redundancy,
- the scale effect,
- avoiding starting problem analysis in the middle,
- avoiding confirming a false hypothesis,
- avoiding tunnel vision.

2.1.3.1 The need for a model

In the sixth-century Crete, large stone columns needed to be moved across long distances. These columns were too heavy for the axles of four-wheel carts, putting too much pressure on roads. A six- or eight-wheel cart presented problems in distributing the load over the axles. Sledges were an alternative, but wider wheels were considered better, widest being the best. Stone columns being the widest were thought to provide a workable solution. Chersiphron, an architect, cut out the center of the column at each end, fitted lead/iron pivots into the cut-out center, and put a wooden frame around the column for pulling it.

The scheme worked well for round columns but not for rectangular ones. Chersiphron's son, Metagenes, used the entire column as an axle by building wide wheels around it at each end. Since this concept worked well too, Paconius, an almost contemporary Roman engineer, bid on the job of moving the pedestal for a statue of Apollo. The pedestal was a stone block, 12 by 8 by 6 feet — about 50 tons of stone. The job was unusual, and Paconius had his own

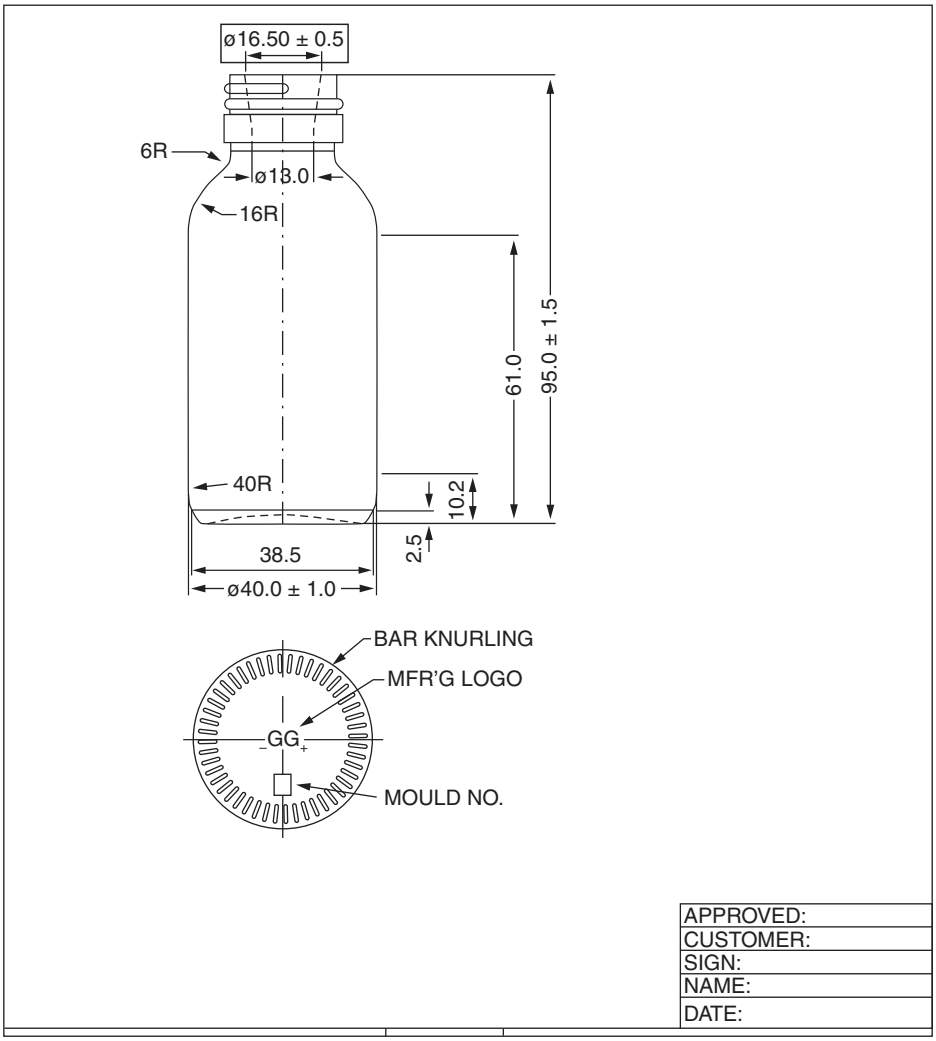


Fig. 2.1.5 A contemporary computer-generated design drawing.

scheme for handling the stone block. He built a great 15-foot-diameter horizontal wooden spool around the pedestal, then he wrapped the spool with rope. The end of the rope came over the top of the spool and was attached to several yokes of oxen. As the oxen pulled the rope, the spool was to roll forward, playing out the rope. It seemed to make sense. As the rope uncoiled, it did indeed cause the wheels to turn, but it could not draw them in a line straight along the road. Hence, it was necessary to draw the machine back again. Thus, by this drawing to and fro, Paconius got into such financial embarrassment that he went bankrupt. Had Paconius built a model, he could have tested his concept and could have avoided the embarrassment as well as the financial ruin.

2.1.3.2 The need for redundancy

In 1961, the first of many "tower blocks" began to be built to accommodate the housing needs of the thousands of local people of West Ham in east London, an area badly mauled by World War II. The design chosen was the Larsen-Nielsen method of using precast reinforced blocks "slotted" into place on site, then bolted and cemented together. This was seen as a safe, quick way to provide new homes while minimizing on-site construction. To avoid shoddy construction and expensive work delays, all walls, floors, and stairways were precast.

On the morning of May 16, 1968, a freak gas explosion caused the collapse of one corner of a 23-story block of Ronan Point apartments in Clever Road, Newham, in east London. The construction of Ronan Point began on July 25, 1966, and the building was handed over to Newham Council on March 11, 1968; it cost approximately £500,000 to build. It was 80 by 60 feet in area, 210 feet high, and consisted of 44 two-bedroom apartments and 66 one-bedroom apartments, five apartments per floor. However, at 5:45 A.M. Thursday morning, an explosion occurred in Apartment 90, a southeast corner apartment on the 18th floor of the new building, blowing out

sections of the outer wall. The modern design apparently proved to have a major fault (insufficient support) which allowed a domino-style collapse of wall and floor sections from the top of the building to the ground (Figure 2.1.6). Officially, the design faults were investigated and took the blame for much of the disaster. It was determined that there was no redundancy; there was a need to effectively join all components of the structure to take up the load during such an eventuality, such as gas explosion.

2.1.3.3 The scale effect

In the late 1800s, a railway bridge across Scotland's Firth of Tay swayed and collapsed in the wind. Seventy-five passengers and crew on a passing night train died in the crash. It was the worst bridge disaster in history. So, when engineers proposed bridging the even wider Firth of Forth, the Scottish public demanded a structure that looked like it could never fall down. Chief engineers Sir John Fowler and Benjamin Baker came up with the perfect structural solution: a cantilevered bridge, with a span of 8276 feet. The Firth of Forth Bridge (Figure 2.1.7) is made of a pair of cantilevered arms "sticking out" from two main towers. The beams are supported by diagonal steel tubes projecting from the top and bottom of the towers. These well-secured spans actually support the central span. This design makes the Firth of Forth Bridge one of the strongest — and most expensive — bridges ever built. The bridge was opened in 1890.

The Firth of Forth Bridge, which was thought to have been overdesigned, was the basic model for the Quebec Bridge. This bridge across the St. Lawrence River was the brainchild of the Quebec Bridge Company. In 1903, the Quebec Bridge Company gave the job of designing of the bridge to the Phoenix Bridge Company. The company also contracted a renowned bridge builder, Theodore Cooper from New York, to oversee the engineering design and construction.

The peculiarities of the site made the design of the bridge most difficult. Because the St. Lawrence was a shipping lane, the 2800-foot bridge was required to have an 1800-foot single span to allow the oceangoing vessels to pass. Further, the bridge was to be multifunctional and 67 feet wide to accommodate two railway tracks, two streetcar tracks, and two roadways. The key to the cantilevered bridge design was the weight of the center span.

In late 1903, P. L. Szlapaka of the Phoenix Bridge Company, laid out the initial drawings of the bridge. His design, which had the background of the overdesigned Firth of Forth Bridge, was approved with very few changes. The estimated weight of the span was calculated based on these initial drawings. In 1905, the working drawings were completed and the first steel girder was bolted into place. These working drawings took over 7 months for final approval. In the meantime, the work had begun. It was not until Cooper received the drawing that he noticed that the estimated weight of the span was off, on the low side, by almost 8 million pounds. Cooper

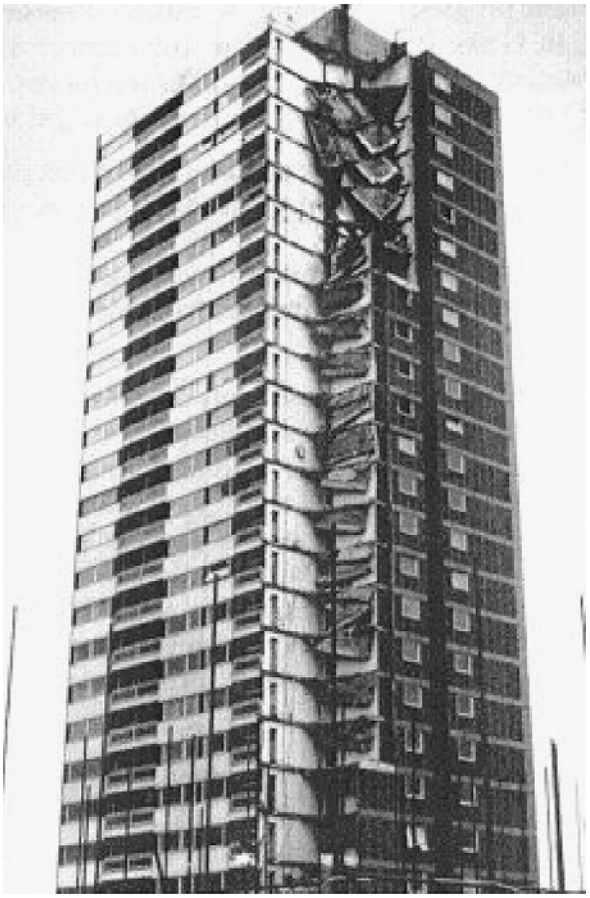


Fig. 2.1.6 The collapse of the Ronan Point apartment complex in the United Kingdom.

Fig. 2.1.7 The Firth of Forth Bridge was completed in 1890 in the United Kingdom.

had two choices: condemn the design and start over or take a risk that there would be no problem. Telling himself that the 8 million pounds was within engineering tolerances, Cooper let the work continue (Figure 2.1.8). After all, he wanted to be known as the designer of the greatest bridge in the world. Also factored into the decision was that the Prince of Wales (later to be King George V) was scheduled to open the bridge in 1908 and any delay in construction would upset the planning.

On June 15, 1907, an inspecting engineer noted that two girders of the anchor were misaligned by a quarter of an inch. Cooper called this a "not serious" problem. In the inspection report in August 1907, it was noted that the girders had moved out of alignment a bit more and "appeared bent." Although this condition was a bit more disconcerting, the work continued. On August 27, 1907, the warning bells finally went off when the inspection team noted that, over the weekend, the girders had shifted a "couple of inches" and were more obviously bent. At 5:32 P.M. on August 29, the girders trembled with a grinding noise and gave way. The bridge structure plunged over 150 feet taking with it the lives of 75 workers (Figure 2.1.9).

The members of the Royal Commission of Inquiry investigating the collapse wrote in their 1908 report, "A grave error was made in assuming the dead load for the calculations at too low a value . . . This error was of sufficient magnitude to have required the condemnation of the bridge, even if the details of the lower chords had been of sufficient strength." The lower chord members had a rectangular section of 5 feet, 7.5 inches, not much smaller than the overdesigned Firth of Forth Bridge chord's circular intersection. The load at the center span was much greater than anticipated, and even though the lower chord section was comparable to the Firth of Forth Bridge chord, it was not sufficient due to the size of the middle span. Clearly, the scale of the center span should have been considered and the chord section enlarged.

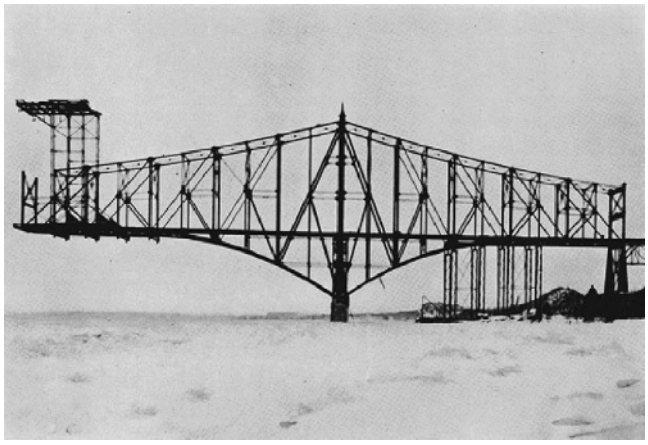


Fig. 2.1.8 The Quebec Bridge under construction in 1907.

Fig. 2.1.9 The collapsed Quebec Bridge.

Extrapolation from other bridge data simply did not work in this case. The bridge finally was completed with the help from the British engineer who worked on the Firth of Forth Bridge.

The significance of the scale factor also is reflected when we wonder why we do not have any giants. To be able to stand up, the bones of a 40-foot giant must support a load that would require bones of much bigger size or of a different material; larger bones would interfere with the limb movements.

2.1.3.4 Avoiding starting problem analysis in the middle

Liberty ships were cargo ships built in the United States during World War II. Eighteen American shipyards built 2751 Liberties between 1941 and 1945. They were British in conception but adapted by the United States, cheap and quick to build. The ships were used primarily for carrying troops.

The initial design was modified by the United States Maritime Commission to conform to American construction practices and to make it even quicker and cheaper to build. The new design replaced much riveting, which accounted for one-third of the labor costs, with welding. No attention was paid to how welding would affect the structure.

Early Liberty ships suffered hull and deck cracks. Almost 1500 instances of significant brittle fractures were recorded. Nineteen ships broke in half without warning. Investigations focused on the shipyards, which often used inexperienced workers and new welding techniques to produce large numbers of ships quickly. A researcher from the Cambridge University demonstrated that the fractures were not initiated by welding but instead by the

grade of steel used, which suffered from embrittlement. It was discovered that the ships in the North Atlantic were exposed to low temperatures that changed the mechanism of cracking from ductile to brittle, causing the hull to fracture easily. The predominantly welded (as opposed to riveted) hull construction then allowed cracks to run large distances unimpeded. One common type of crack nucleated at the square corner of a hatch that coincided with a welded seam, with both the corner and the weld acting as stress concentrators. Had the effect of welding on structural stiffness been considered, the problem could have been avoided.

Similar was the case with NASA's shuttle booster joint. The design was based on Titan III joint and included two O-rings instead of one. This was thought to be safer. The outcome indicated otherwise. After the 1986 space shuttle disaster (The Challenger), the booster rocket was redesigned with three O-rings (Figure 2.1.10).

Both examples indicate that it is unwise to solve a problem by beginning in the middle; it is important not to loose site of the goal and start the process at the very beginning.

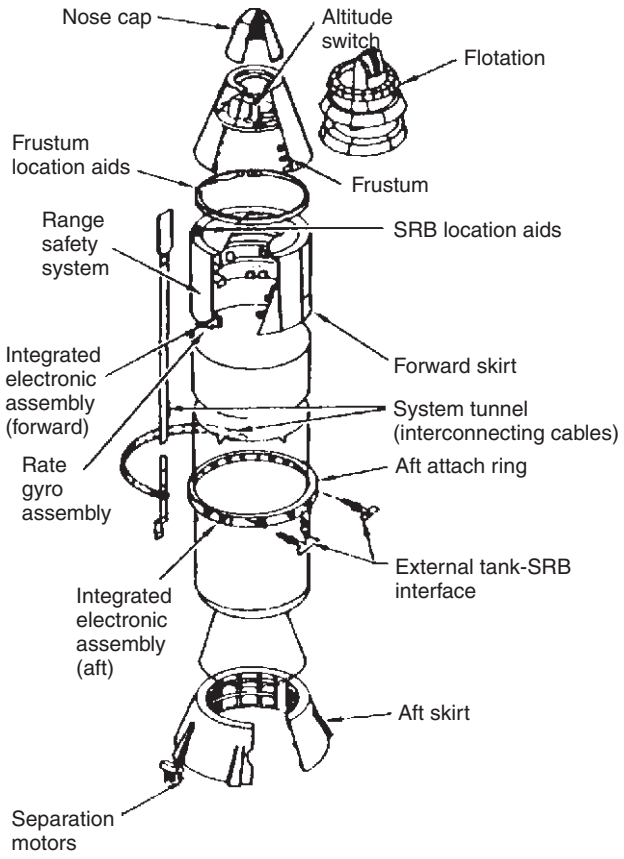


Fig. 2.1.10 The redesigned solid rocket booster for the space shuttle (NASA).

2.1.3.5 Avoiding confirming a false hypothesis

It is easy to validate a hypothesis by means of several examples. However, only one counterexample is necessary to invalidate it. It is inevitable in a data-driven statistical study that some false hypotheses are accepted as true. In fact, standard statistical practice guarantees that at least 5% of false hypotheses are accepted as true (the probability of type I error being 5%). Therefore, out of the 800 false hypotheses 40 will be accepted as true, that is, statistically significant. It is also inevitable in a statistical study that some true hypotheses are not accepted as such. It is hard to say what the probability is of not finding evidence for a true hypothesis, because it depends on a variety of factors, such as the sample size. Our goal, therefore, should be to collect as much data as possible to reduce the probability of accepting a false hypothesis or to come up with a proper explanation, such as why a failure occurs. It is not uncommon in the field of medicine to find a clinical study whose conclusions are proven to be false when a larger study is undertaken later. In engineering design, similar occurrences are known to happen.

Jet transportation age began on May 5, 1952, when the de Havilland Comet 1 (Figure 2.1.11) began scheduled flights from London to Johannesburg. The Comet had a cruising speed of 490 mph at 35,000 feet and a range of 1750 miles with a payload of 44 passengers. The cabin was pressurized equivalent to 8000 feet at an altitude of 40,000 feet. This gave a pressure differential of 8.25 psi (56 kPa) across the fuselage, twice the value previously used. De Havilland conducted "many tests" to ensure structural integrity of the cabin. However, three accidents occurred where Comet aircraft disintegrated in flight and all Comet 1 aircraft were subsequently withdrawn from service:

- G-ALYV after leaving Calcutta May 1953. Violent storms were thought to be involved and some wreckage was recovered. No firm conclusions were drawn as to the cause.
- G-ALYP over Elba January 1954 after 1286 cabin pressurization cycles. Little wreckage was recovered and no major problems were found in fleet inspection. Fire was assumed to be the most likely cause and modifications were made to improve fire prevention and control. The aircraft was returned to service.
- G-ALYY flying as SA 201 after leaving Rome April 1954. About 70% of the aircraft was recovered and reconstructed at Farnborough. The engines were recovered more-or-less intact, showing that engine disintegration was not the cause of the accident, and neither was any evidence of fire found.

Fig. 2.1.11 The unfortunate de Havilland Comet airplane of BOAC.

The Comet G-ALYU, which had experienced 3539 flying hours and 1221 cabin pressurization cycles, was subjected to full-scale flight simulation testing at Farnborough. The fuselage was hydraulically pressurized in cycles, while the wings were flexed with jacks to simulate the flight loads. Water was used for this pressurization because calculations had indicated that the energy release under cabin rupture with air as the pressurization medium was equivalent to the explosion of a 500-pound-force bomb in the cabin. The cabin also was supported in water to avoid extraneous weight effects. After the equivalent of a total of 3057 (1836 simulated cycles) flight cycles, a 2-mm crack near the escape hatch grew to failure. This was repaired, and after 546 flight cycles, a 4.5-meter section of the cabin wall ruptured due to fatigue cracking. It was concluded that explosive cabin failure had caused the loss of the three Comet aircraft. Developing a detectable crack 6-mm long consumed some 95% of the cyclic life.

The Royal Navy was charged with getting the relevant fuselage piece of G-ALYP from the sea (using simulation trials, based on the way the aircraft was now thought to break up in flight) to establish the likely position of this part of the aircraft on the seabed. This showed the unmistakable sign of fatigue. The fatigue crack was associated with the stress concentrations of the rather square rear ADF window cutout (stress of 315 MPa at edge of window) and with a bolt hole around the window (although the stress at the bolt position was only 70 MPa).

The chief designer at De Havilland had wanted to glue the windows in position, but the tooling for the square shape was too difficult to make. A lower stress concentration shape would have been easier to manufacture.

The manufacturer had performed fatigue tests of the forward cabin area at about 10 psi (with cracking occurring at 18,000 cycles), but these were carried out after static tests of up to 16.5 psi (twice operating pressure) had previously been applied. Cracks also were known to be present after manufacture, and the remedy was to drill 1.6-mm holes at the crack tip to "arrest" them (such an arrested crack was present near the rear ADF window, which had not propagated until the final failure). In the end, the following causes were identified:

1. New technology introducing new load cases (high-altitude flight for turbojet engines requiring cabin pressurization).
2. Mismatch between service loads and fatigue test procedure.
3. Possible contribution from out-of-plane bending loads (biaxial stresses) resulting from the following design failures:
   - Improperly understood failure mode assessment procedures necessitated by implementation of new technology.
   - Poor configuration due to wing root engine placement (very few other aircraft have had engines in this position), affecting uprating potential, fire hazard, and structural integrity in the event of engine disintegration.

2.1.3.6 Avoiding tunnel vision

The 1940 Tacoma Narrows Bridge, a very modern suspension bridge with the most advanced design, collapsed in a relatively light wind. The final investigation revealed that the designers did not pay attention to the light weight of the road and supporting system in their concern for the vertical and horizontal "flexibility."

The state of Washington, the insurance companies, and the United States government appointed boards of experts to investigate the collapse of the Narrows Bridge. The Federal Works Administration appointed a three-member panel of top-ranking engineers. In March 1941, the panel of engineers announced its findings. "Random action of turbulent wind" in general, said the report, caused the bridge to fail. This ambiguous explanation was the beginning of attempts to understand the complex phenomenon of wind-induced motion in suspension bridges. Three key points stood out:

1. The principal cause of failure of the 1940 Tacoma Narrows Bridge was its "excessive flexibility."
2. The solid plate girder and deck acted like an aerofoil, creating "drag" and "lift."
3. Aerodynamic forces were little understood, and engineers needed to test suspension bridge designs using models in a wind tunnel.

"The fundamental weakness" of the Tacoma Narrows Bridge, said a summary article in *Engineering News*

Record, was its "great flexibility, vertically and in torsion." Several factors contributed to the excessive flexibility: The deck was too light. The deck was too shallow, at 8 feet (a 1 to 350 ratio with the center span). The side spans were too long, compared with the length of the center span. The cables were anchored at too great a distance from the side spans. The width of the deck was extremely narrow compared with its center span length, an unprecedented ratio of 1–72.

The pivotal event in the bridge's collapse was the change from vertical waves to the destructive twisting, torsional motion. This event was associated with the slippage of the cable band on the north cable at midspan. When the band slipped, the north cable became separated into two segments of unequal length. The imbalance translated quickly to the thin, flexible plate girders, which twisted easily. Once the unbalanced motion began, progressive failure followed. Wind tunnel tests also concluded that the bridge's lightness, combined with an accumulation of wind pressure on the 8-foot solid plate girder and deck, caused the bridge to fail.

2.1.4 The requirements for design

Designing is the application of technical and scientific principles to arrange components of a device. When the device is adapted and embodied to achieve a specific result, it must satisfy the six requirements as outlined by Pye (1989). These requirements are as follows. The requirements of use are

1. It must correctly embody the essential principle of arrangement.
2. The components of the device must be geometrically related to each other and to the objects, in whatever particular ways suit these particular objects and this particular object.
3. The components must be strong enough to transmit and resist forces as the intended results require.
4. Access must be provided.

The requirement for ease and economy is

5. The cost of the result must be acceptable.

The requirement of appearance is

6. The appearance of the device must be acceptable.

Design is the process of satisfying these requirements.

As Pye states, these design requirements are conflicting in nature; therefore, the final design, no matter what, cannot be perfect. It is the designer's responsibility to determine, in consultation with the client, the degree and location of "failures" (compromises); for instance, the conflict between the requirements for economy and durability, between speed and safety, between usability and functionality. Since the final design is the result of many compromises, design may be considered a problem-solving activity; the requirement of appearance makes it an art. Therefore, design is both a problem-solving activity and an art.

2.1.5 The design process

In this section, we discuss the steps of the actual design process. However, before we discuss these steps it would be prudent to discuss the problem that confronts the designers.

2.1.5.1 Problem confronting the designers

A product has certain properties that make it useful to people. The properties can be physical, such as size, weight, or strength, or chemical, such as composition, heat tolerance, or rust resistance. Some of the properties are intrinsic, some are extrinsic, and some are the result of the physical form of the product (geometrical form). Table 2.1.1 shows the various intrinsic, extrinsic, and design properties of a product. As a result of these properties, the environment in which it operates, and the geometrical form it has, a product can perform certain functions. The fulfillment of these functions satisfy human wants and needs and helps the product achieve one or several values. The achievement of these values is what makes a product useful to people. Figure 2.1.12 shows this progression.

Knowing the form of a product, it is possible to derive its properties, the functions it can perform, the human needs it can satisfy, and the values it will achieve. The process of design, however, does not require predicting properties and functions from the form, for the form is unknown. Rather, it is to achieve an embodied form that, by virtue of its intrinsic and extrinsic properties, performs certain functions that satisfy human needs. In other words, the challenge for a designer is to move from right to left in Figure 2.1.12. Transition from function to form, to a considerable degree, depends on the ability, imagination, and creativity of the designer. This, then, is the problem that confronts the designers: to embody properties in a geometrical form such that the embodied form, when used as intended in the specified environment, can perform the intended functions.

While a specific product can perform only certain functions, it is possible to come up with a number of forms that perform the same set of functions. Conceiving these forms and choosing the final one is the challenge for designers. While design methods can help, creativity and imagination are crucial in the transition from function to form.

Table 2.1.1 Properties of a product (adapted from Hubka and Eder, 1988)

Design structure	Internal strength	External operational properties	System space requirement
Form	Manufacturing properties	Ergonomic properties	Durability, life
Tolerance	Corrosion resistance	Aesthetic properties	Weight/mass
Surface	Durability	Distribution properties	Maintenance
Manufacturing methods		Delivery and planning properties	Operation
Materials		Law conformance properties	Surface quality
Dimensions		Manufacturing properties	Color
		Economic properties	Appearance
		Liquidation properties	Storage space
		Function	Transportability, packing
		Functionally determined properties	Delivery deadline
			Laws, regulations, standards, codes of practice
			Quality
			Operational costs
			Price
			Wastes
			Recycling
			Function
			Reliability

2.1.5.2 Steps of the engineering design process

The basic engineering design process is not unlike the engineering problem-solving process, as designing is not much different from solving any problem. In fact, "designing" is a special form of problem solving, as the cycle of activities that must be undertaken to come up with a design is similar. Hall (1968) outlined these basic activities as

- **Problem definition:** Studying needs and environment.
- **Value system design:** Stating objectives and criteria.

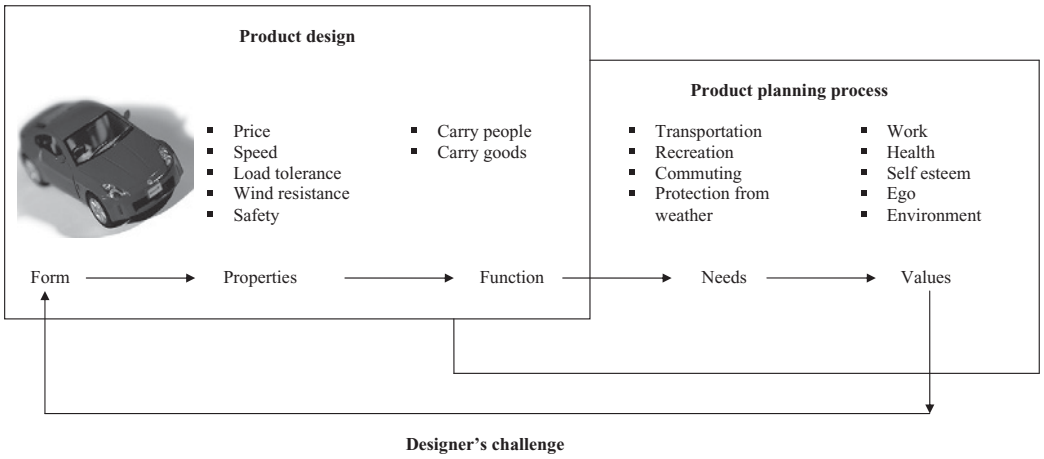


Fig. 2.1.12 Link between a product's form, its properties, its function, and human needs and values.

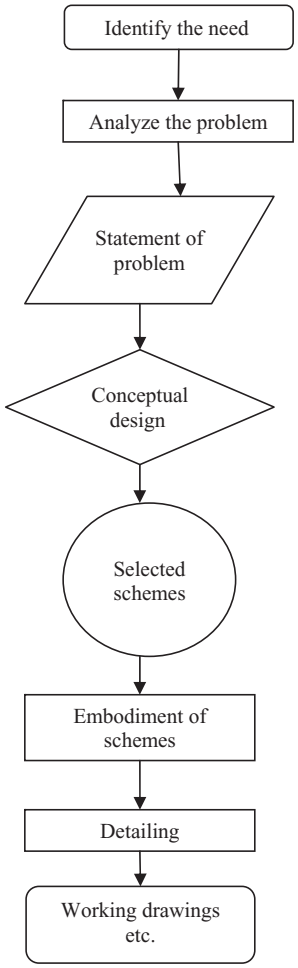


Fig. 2.1.13 French's (1971) basic engineering design process.

- **Systems synthesis:** Generating alternatives.
- **Systems analysis:** Analyzing alternatives.
- **Selecting the best system:** Evaluating alternatives against selected criteria.
- **Planning for action:** Specifying the selection.

These activities are reflected in many models of the basic design process. Figure 2.1.1.13 shows one of the earliest such models, developed by French (1971). A variation of the model, somewhat more detailed and shown in Figure 2.1.14, is presented by Pahl and Beitz (1984). The activities in Figure 2.1.13 are replaced by phases in Figure 2.1.14. The German professional engineers' body, Verein Deutscher Ingenieure (VDI) produced a general design process guideline, VDI 2221, shown in Figure 2.1.15 (1987). This is a somewhat detailed variation of the engineering design process model shown in Figure 2.1.14. The VDI 2221 model outlines a procedure that first emphasizes analyzing and understanding the problem in detail, then breaking the problem into subproblems, finding solutions to these subproblems, and finally combining these solutions into an overall solution. All phases of this model require coming up with a variety of solutions (divergence in the design phase). From this variety of solutions, the best option is picked (convergence in the design phase). This concept is shown in Figure 2.1.16.

The principles used for the solution to subfunctions generally are basic scientific and engineering principles. Pahl and Beitz (1984) provided examples of solution principles to subfunctions, as shown in Figure 2.1.17.

2.1.5.3 Defining the problem and setting objectives

A problem is the result of an unfulfilled need. Unless the need is clearly defined, the problem cannot be formulated; unless a problem is properly formulated, there is either no solution or no guarantee that the proposed solution will solve the problem. It is typical, however, that the need, when expressed initially, is vague. The problem, therefore, is ill-defined.

Since the start of the problem, and therefore the design solution, generally is vague, the first step is to state the general objective and gradually clarify it. It is possible that, in the process, the initial objective may change or be altered significantly. Such a change or alteration reflects a better understanding of the problem and, eventually, a more fitting design solution.

As the objective changes, from a broad goal to specific goals, the means to accomplish the result may change as well. With each stage of the change, the objective should be reiterated in clear and precise language. As the objective becomes more specific or is broken down into subobjectives, the criteria for evaluating design solutions emerge. These are classified as design specifications.

The objective tree method is a convenient format for developing and clarifying objectives. Beginning with a general statement or broad objective, the intent is to prepare a sequence of objectives and subobjectives. Some of these could be scaling (e.g., price has to be as low as possible) and some nonscaling (e.g., the price must be under $100); some could be in the form of requirements (e.g., the operating temperature must not exceed 210°F); and some could be standards (requirements imposed by an external authority, such as the local, state, or federal government or industry agreements). The objectives are arranged in a tree format, showing hierarchical relationships and interconnections.

Figure 2.1.18 shows an objectives tree for a comfortable, safe, and attractive child car seat. The figure represents a top-down approach, starting from a higher-level objective and progressing to lower-level objectives. There is no clear end point in the top-down approach, but achieving a lower-level objective is a means to achieving the higher-level objective. When the lower-level objectives are used as a means to achieving

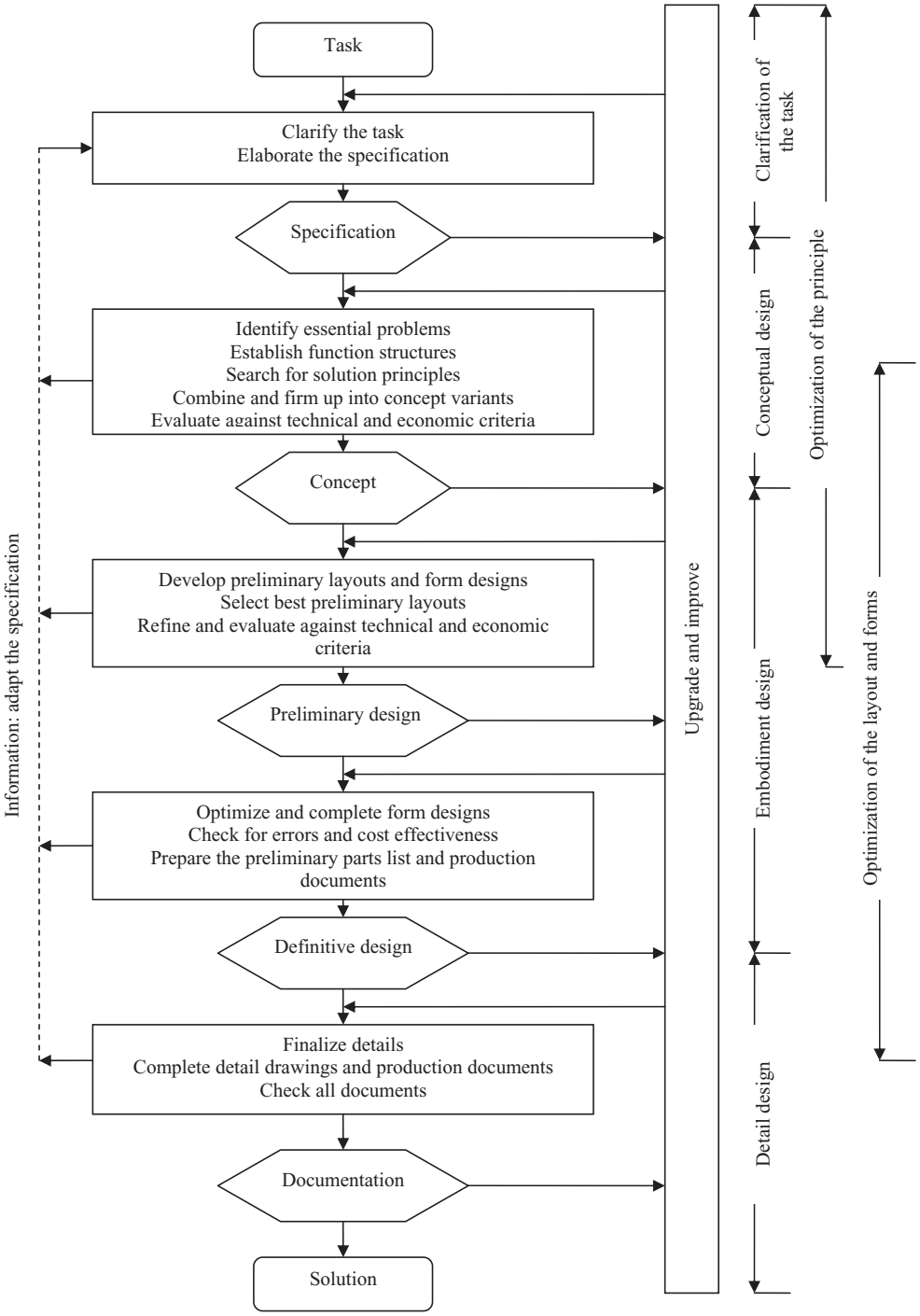


Fig. 2.1.14 The basic engineering design process according to Pahl and Beitz (1984).

higher-level objectives (an end) and the tree is redrawn from a bottom-up approach, the result is a means-end chain or cause–effect chain. This method helps clarify objectives. Figures 2.1.19 and 2.1.20 show top-down and bottom-up trees for the safety of automobile travel (Keeney, 1992).

A checklist approach may be used to develop a comprehensive list of objectives. Pugh (1990) provided a list of 24 factors that may be used in a checklist format to come up with a comprehensive list of subobjectives. These factors were summarized by Roozenburg and Eekels (1995):

1. **Performance:** Which function(s) does the product have to fulfill? By what parameters will the functional characteristics be assessed? Speed? Power? Strength? Accuracy? Capacity?

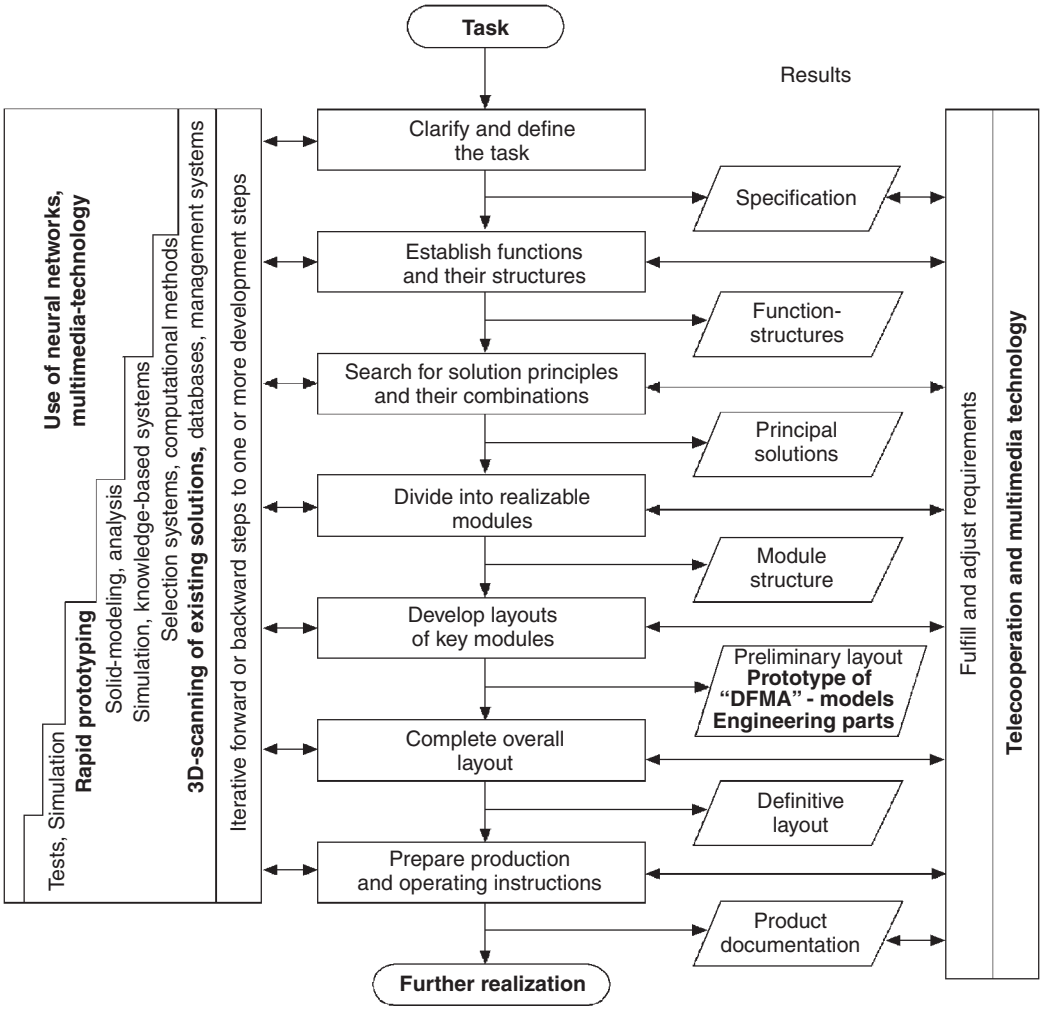


Fig. 2.1.15 A general approach to engineering design (VDI 2221, 1987).

2. **Environment:** To which environmental influences is the product subjected during manufacturing, storing, transportation, and use? Temperature? Vibration? Humidity? Which effects of the product on the environment should be avoided?
3. **Life in service:** How intensively will the product be used? How long does it have to last?
4. **Maintenance:** Is maintenance necessary and available? Which parts have to be accessible?
5. **Target product cost:** How much may the product cost, considering the price of similar products?
6. **Transportation:** What are the requirements of transport during production, and to location of use?
7. **Packaging:** Is packaging required? Against which influences should the packaging protect the products?
8. **Quantity:** What is the size of the run? Is it batch or continuous production?
9. **Manufacturing facilities:** Should the product be designed for existing facilities? Are investments in new production facilities possible? Will the production or a part of it be contracted out?
10. **Size and weight:** Do production, transport, or use put limits as to the maximum dimensions? Weight?
11. **Aesthetics, appearance, and finish:** What are the preferences of the consumers? Should the product fit in with a product line or house style?
12. **Materials:** Are special materials necessary? Are certain materials not to be used?
13. **Product life span:** How long is the product expected to be produced and marketable?
14. **Standards:** Which standards apply to the product and its production? Should standardization within the company be taken into account?
15. **Ergonomics:** Which requirements (perception, use, handling, etc.) does the product have to meet?

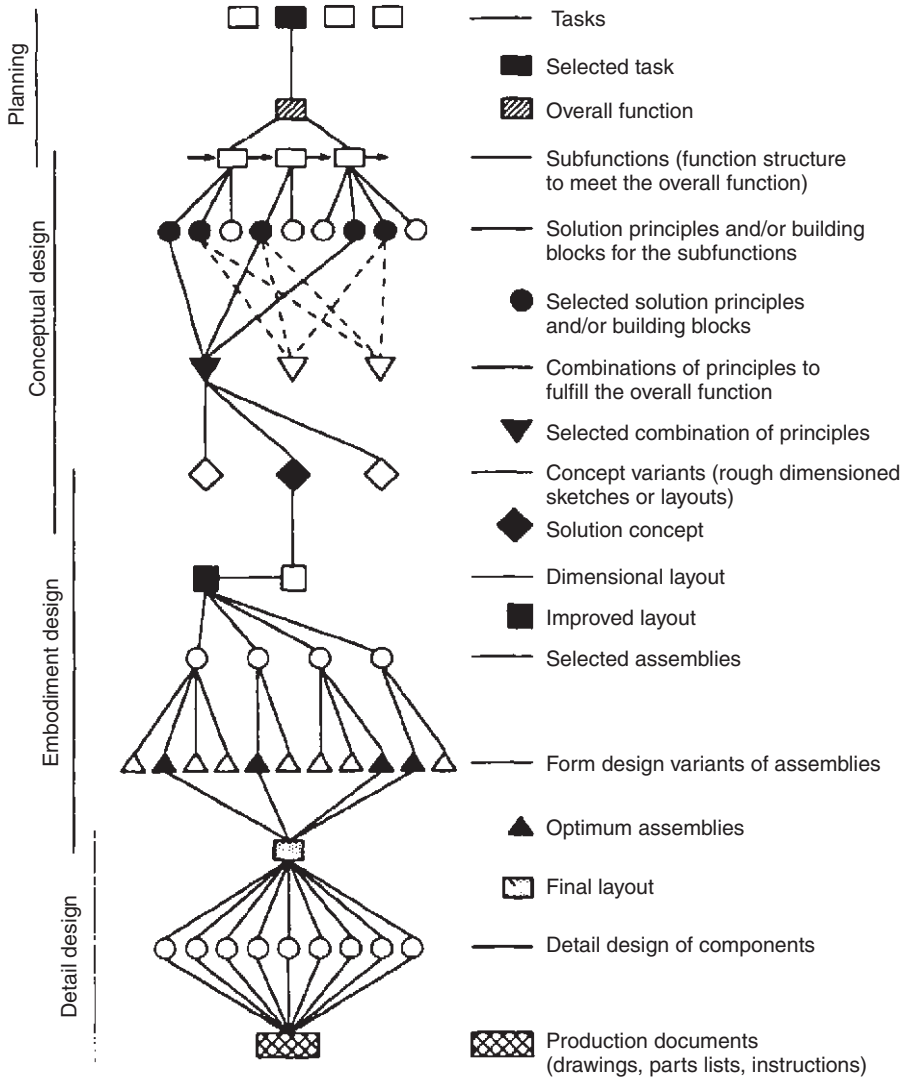


Fig. 2.1.16 Divergence and convergence in the design process (VDI 2222, 1987).

16. **Quality and reliability:** How large may mean times between failure and mean times to repair be? Which failure modes and resulting effects on functioning should not occur?
17. **Shelf life and storage:** During production, distribution, and use, are there periods of time in which product is stored? Does it require specific measures?
18. **Testing:** To which functional and quality tests is the product submitted within and outside the company?
19. **Safety:** Should any special facilities be provided for the safety of the users? Nonusers?
20. **Product policy:** Does the current and future product range impose requirements on the product?
21. **Social and political implications**: What is the public opinion with regard to the product?
22. **Product liability:** For which unintended consequences of production, operation, and use can the manufacturer be held responsible?
23. **Installation and operation:** Which requirements are set by final assembly and installation outside the factory and by learning to use and operate the product?
24. **Reuse, recycling, and disposal:** Is it possible to prolong the material cycle by reuse of materials? Parts? Can the materials and parts be separated for waste disposal?

The list of objectives should be analyzed to remove similar objectives and objectives that are not biased. Objectives should be further stated in terms of

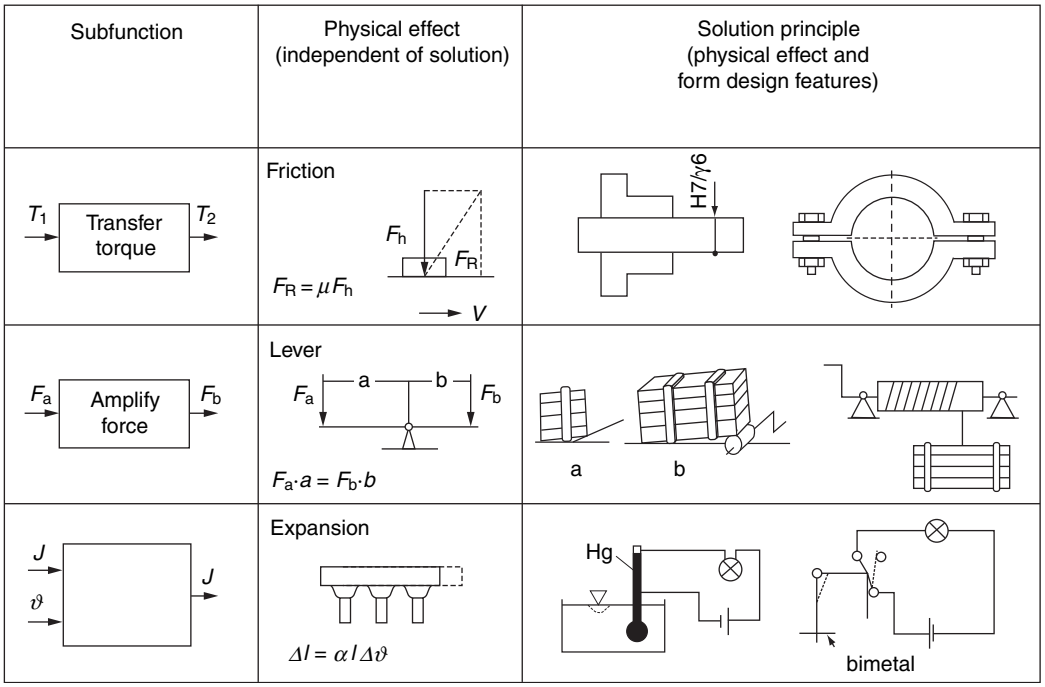


Fig. 2.1.17 Solution principles to subfunctions (examples from Pahl and Beitz, 1984).

performance and their hierarchical relationships should be verified. Requirements and standards should be checked for acceptable values and a means-end relationship should be drawn.

2.1.5.4 Establishing functions, setting requirements, and developing specifications

Once the objectives list has been reexamined, consolidated, and edited, a function analysis must be performed to determine what the product should achieve (not how). For this purpose, the product can be treated as a black box that converts specified inputs into outputs. The overall function should be as broad as possible and this is what the product design accomplishes: conversion of inputs into outputs. The designer should ask questions about the inputs and outputs (e.g., what are the inputs? where do they come from? what are the outputs for? stages of conversion? sequence of conversion?) to develop a complete list of inputs and outputs. These inputs and outputs can be flows of materials, energy, and information. For instance, a coffee bean grinder takes

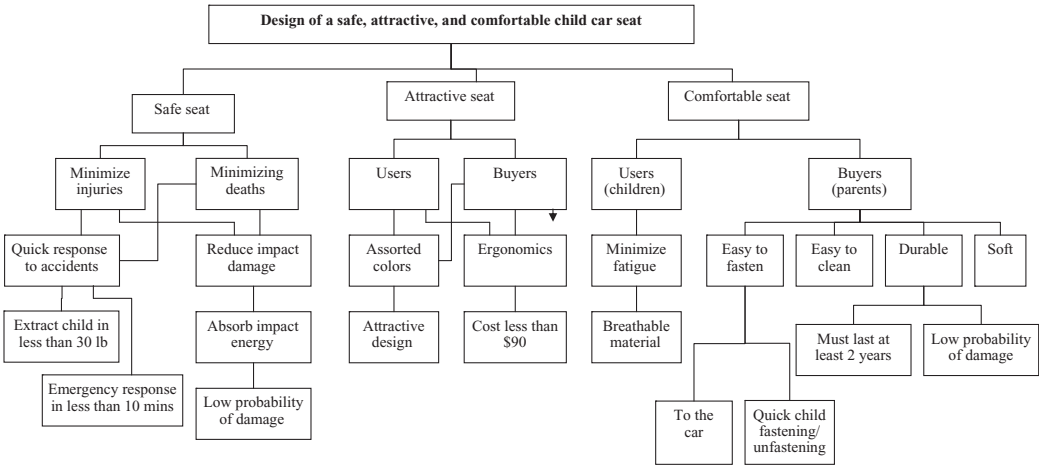


Fig. 2.1.18 An objectives tree for a child car seat.

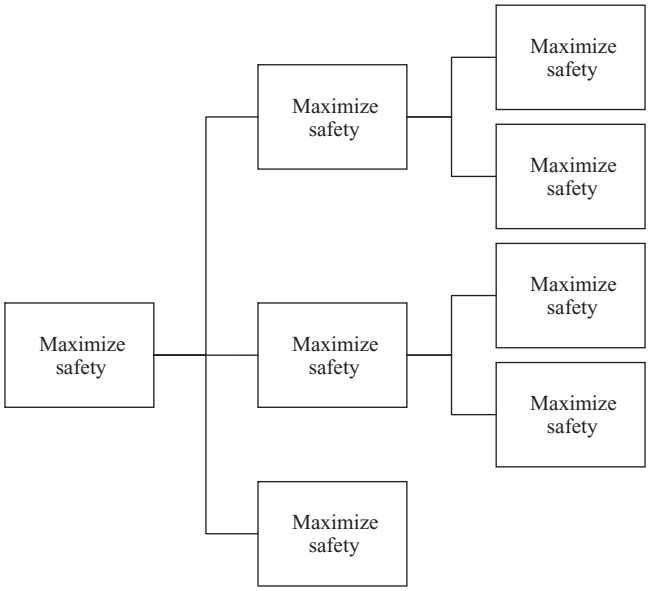


Fig. 2.1.19 An objectives tree for safe automobile travel, top-down approach (adapted from Keeney, 1992).

beans (matter), energy (electrical), and information (signal) as inputs and grinds beans. The ground beans (matter), heat (energy), and information (signal) are the outputs. In this case, the inputs and outputs are specified but not how the beans are ground.

Next, the function in the black box is replaced with a block diagram showing all subfunctions and their links with each other and with inputs and outputs. There can be no loose inputs and outputs or subfunctions. Eventually, the designer must look for proper components for accomplishing each subfunction. The component can be a person or a mechanical or an electrical device. Figure 2.1.21, modified from Hubka et al. (1988), shows the black box model and function analysis for automatic coffeemaker.

The function analysis specifies what a design must achieve. However, no limits are set. Setting precise limits on the functions is called *setting performance specifications*. This means that we must set limits on the range of acceptable solutions. Performance specifications should be neither too narrow (otherwise normally acceptable solutions are eliminated) nor too broad (which may lead to inappropriate solutions). The design objectives and functions are the source of performance specifications. The list of performance attributes contains all conditions the design must meet. These include requirements (functional as well as nonfunctional) and standards. Additionally, if one wishes, performance specifications for desired attributes may be included as well. For each attribute, the designer must determine what the product must achieve or do as far as the performance is considered. It is most desirable to express this quantitatively, if at all possible. For instance, the toothbrush cleaning cycle must be 120 seconds long or the product handle diameter should be 1.5 inches; vague or qualitative specifications, such as approximately 10 pounds or as low as possible should be avoided. In general, performance specifications

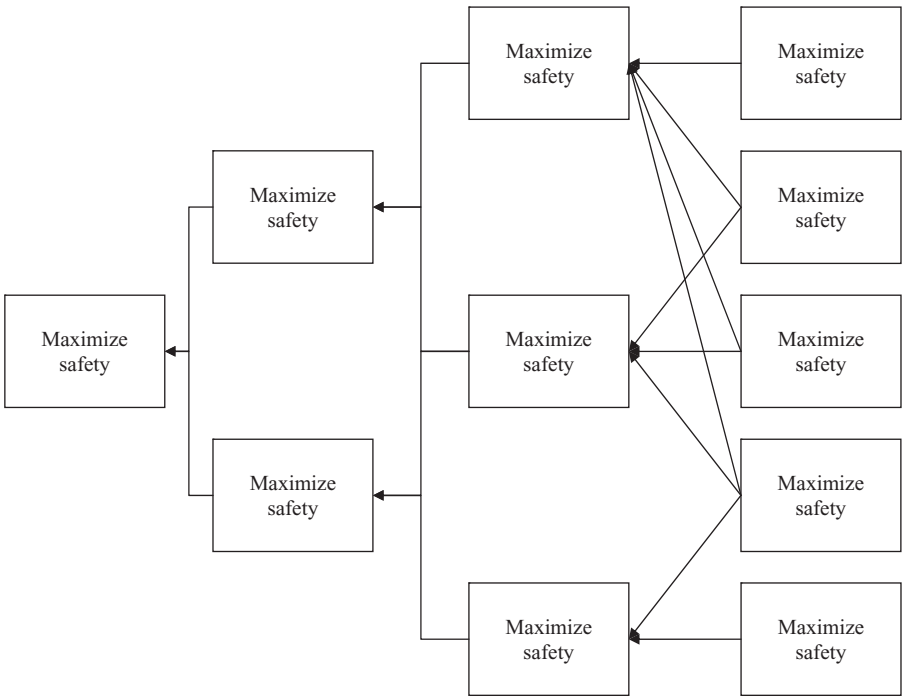


Fig. 2.1.20 A means-end tree for safe automobile travel, bottom-up approach (adapted from Keeney, 1992).

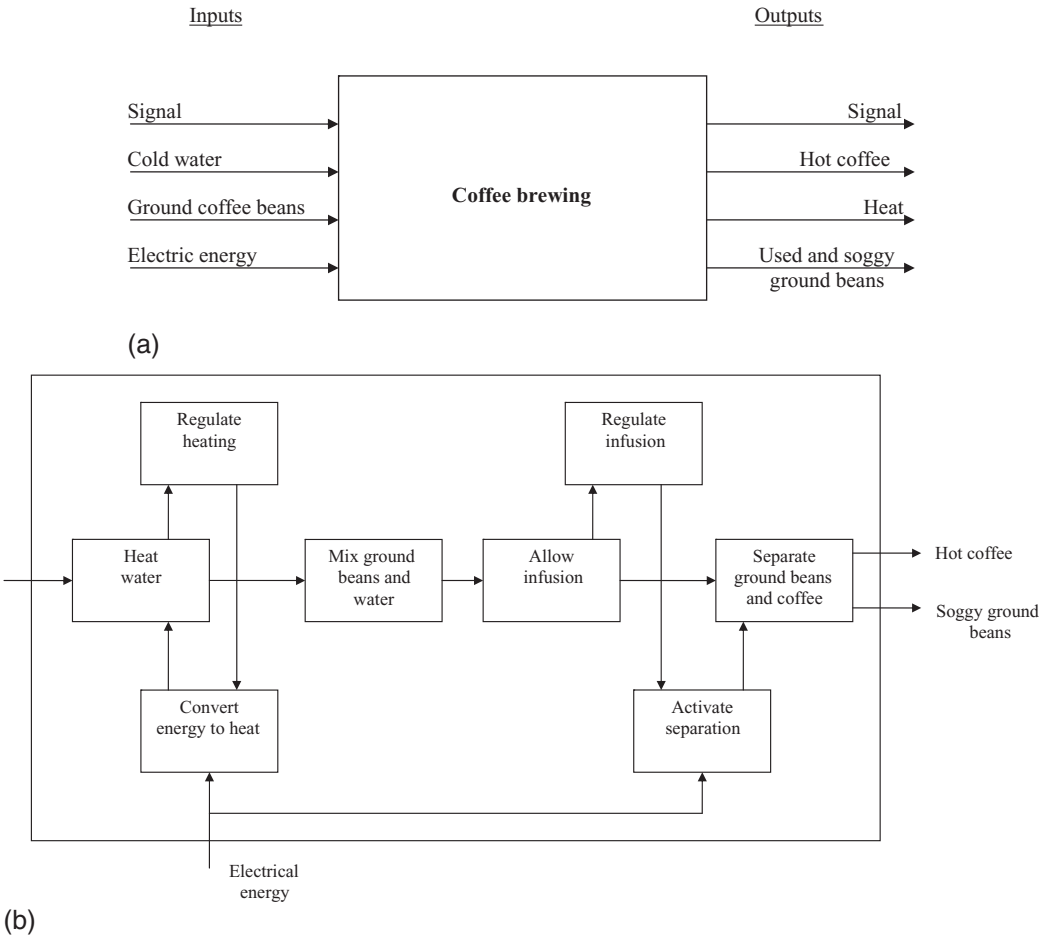


Fig. 2.1.21 (a) A black box model for coffee brewing. (b) Function analysis for automatic coffeemaker (adapted from Hubka et al., 1988).

should be independent of any solution. If a range would satisfy performance specifications, it should be specified instead of a fixed number. Table 2.1.2 provides design specifications for a small city car developed by Pighini et al. (1983).

One way to generate product specifications is to use the method based on the principles of quality function deployment. The method requires using the "house of quality" interaction matrix with product attributes as rows and technical parameters (also known as *engineering characteristics*), such as weight, volume, material, and force required, as columns. The technical parameters have to be related to the basic engineering characteristics. For instance, torque is determined by the gear ratio of the transmission train and the motor power. These, in turn, are affected by variables such as voltage and resistance. It is these basic variables that have to be related to product attributes to determine the specifications. Each attribute is assigned a weight, or relative importance, factor. Next, the relationship between the product attribute and technical parameter is determined using notations for strong positive, medium positive, neutral, medium negative, and strong negative; and a value is assigned to each of these relationships. Multiplying the weight and the relationship score produces a value for the strength of the product attribute–technical parameter relationship. The sum for each column (parameter) indicates the priority of that technical parameter in product design. At the bottom of the matrix, the measurement units of technical parameters are listed. A target value for each parameter is set on the basis of its importance and how the values of the attribute for competitors' products would compare against the target.

2.1.5.5 Developing provisional designs

The methods for developing design solutions to problems are called *design methods*. Any procedures, techniques, aids, or tools for designing are a part of the design methods. While some methods are informal, many are formal and rely on decision theory and techniques used in management science. Jones (1981) outlined as many as 35 design

Table 2.1.2 Design specifications for a small city car (adapted from Pighini et al., 1983)

Characteristics	
Required	**Desired**
1. General characteristics:	
Car for city	X
Number of seats: 2(4)	X
Number of wheels: 4	X
Space utilization: maximum	X
Range: >100 km	X
Economical	X
2. Working conditions:	
City area	X
3. Dimensions	
2.5-meter length	X
1.5-meter width	X
1.6-meter height	X
4. Weight:	
Maximum net: 400 kg	X
Loading capacity: 200–300 kg	X
Gross weight: 600–700 kg	X
5. Luggage capacity:	
Minimum volume: 150 dm^3	X
With dropped back: 359 dm^3	X
6. Speed:	
Maximum: 70 km/hour	X
7. Motor type	
8. Safety:	
As high as possible: active and passive	X
9. Pollution:	
Meet standards	X
Zero	X
10. Form and aesthetics:	
Pleasing	X
Convertible	X
11. Production:	
500 units/year	**X**
12. Price:	
$2000 to $2500	**X**

methods. In general, the design methods are classified under the following categories: association methods, creative confrontation methods, and analytic methods.

2.1.5.5.1 Brainstorming

The most widely known creative association method is brainstorming. This method, invented by Osborn (1963), is very effective in generating a large number of ideas. Even though most of the ideas originally generated are rejected, some ideas will be worth pursuing. The method requires a small group of people (four to eight) of diverse background to participate in the solution-generating activity. According to Osborn, the following four rules must be followed:

1. There should be no criticism of the ideas generated. The participants should not think about the utility or worth of the idea and certainly should not criticize them. Such a criticism stifles the idea-generation process.
2. Any idea is welcome; no matter how outlandish. Such divergent thinking is productive. The participants must feel secure in proposing any wild ideas they think relevant.
3. Adding to other ideas to improve on them is welcome.
4. The goal is to generate as many ideas as possible.

The procedure requires that the problem statement be provided to the selected group of participants in advance of the brainstorming session. Background information, some examples of solutions, information regarding the participation rules also should be provided. A preliminary meeting explaining the rules and purpose may be carried out.

During the actual session, participants provide their ideas through the moderator. Nonparticipants also can provide their ideas in writing. Once the list of ideas is complete, participants discard silly ideas and focus on those that seem relevant. The relevant ideas also may be elaborated by others, later.

Some variations of the original brainstorming method are also practiced, for instance, the brainwriting-pool method. In this variant, five to eight people write their ideas on a sheet of paper in silence, once the problem has been explained. The sheet of ideas is placed in a pool. Participants who already put their sheet in the pool pull a sheet from the pool and add ideas to it. After about 30 minutes, the ideas on the sheets are evaluated as in the original method.

Another variation of the original method is the 6-3-5 method, in which six participants write three ideas on a sheet. The sheets are passed five times among the participants, during which ideas are added. As many as 108 ideas can be generated in this manner.

2.1.5.5.2 Analogies and chance

Many innovative ideas can be attributed to chance. For instance, vulcanized rubber was invented by Charles Goodyear when he accidentally added sulfur to rubber when working in his wife's kitchen. Similarly, John Boyd Dunlop invented the pneumatic tire after observing the behavior of the garden hose when watering his garden.

Cyanoacrylate, also known as superglue or Krazy Glue, developed by Harry Coover at Eastman Kodak during World War II, is another example of chance discovery. Coover was searching for a way to make plastic gunsight lenses when he discovered cyanoacrylate, which did not solve the problem since it stuck to all apparatuses used to handle it. It was patented in 1956 and developed into Eastman 910 adhesive in 1958. The new glue was demonstrated in 1959 on the television show *I've Got a Secret*, when the host Garry Moore was lifted into the air by two steel plates held together with a drop of Eastman 910. Cyanoacrylates are now a family of adhesives based on similar chemistry.

The Post-it note was invented in 1968 by Dr. Spencer Silver, a 3M scientist who stumbled on a glue that was not sticky enough. In 1974, a colleague, Arthur Fry, who sang in a church choir, was frustrated that the bookmarks kept falling out of his hymnal, so he applied some of Silver's glue to his markers. 3M launched the product in 1977 but it failed, as consumers had not tried the product. A year later, 3M swamped Boise, Idaho, with samples. Of people who tried them, 90% said that they would buy the product. By 1980, the product was sold nationwide and a year later it was launched in Canada and Europe.

The examples cited here demonstrate the role of analogy and chance in solving problems. Bright people often can come up with a design solution when two situations are accidentally confronted. Seemingly, the two situations have nothing to do with each other or have some indirect relationship. Gordon (1961), and later Prince (1970), formalized the mechanics of the process which now is known as *synectics*.

Synectics is different from brainstorming: instead of generating ideas, the group tries to work collectively toward a particular solution. The session is much longer and much more demanding. The group is pushed to use particular type of analogies by making the strange familiar and the familiar strange. The types of analogies encouraged are:

- **Direct analogies:** A biological solution to a similar problem is sought. As an example, Velcro was designed using an analogy with plant burrs.
- **Personal analogies:** The team members put themselves in the problem being solved; for example, how would I feel if I were a transmission?
- **Symbolic analogies:** The team members use metaphors and similes, such as the "jaw" of a clamp.
- **Fantasy analogies:** The team members dream of an ideal solution to the problem; for example, a child's fantasy, such as door opening by itself when the owner reaches it.

The synectics method has the following steps:

1. State the problem (e.g., how to fly and stay in place).
2. Use a direct analogy (e.g., hummingbirds stay flying in one place).
3. Analyze the analogy (e.g., flap wings rapidly, 15–80 times per second; how to achieve this?).

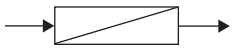
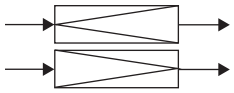
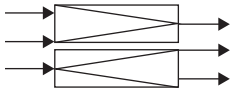
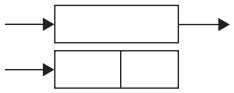
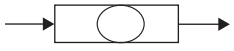
Characteristic	Generally vaild functions	Symbols	Explanations Input(I)/Output(O)
Type	Change		Type and outward form of I and O differ
Magnitude	Vary		I < O I > O
Number	Connect		Number of I > O Number of I < O
Place	Channel		Place of I ≠ O Place of I = O
Time	Store		Time of I ≠ O

Fig. 2.1.22 Symbols for general functions (adapted from Pahl and Beitz, 1984).

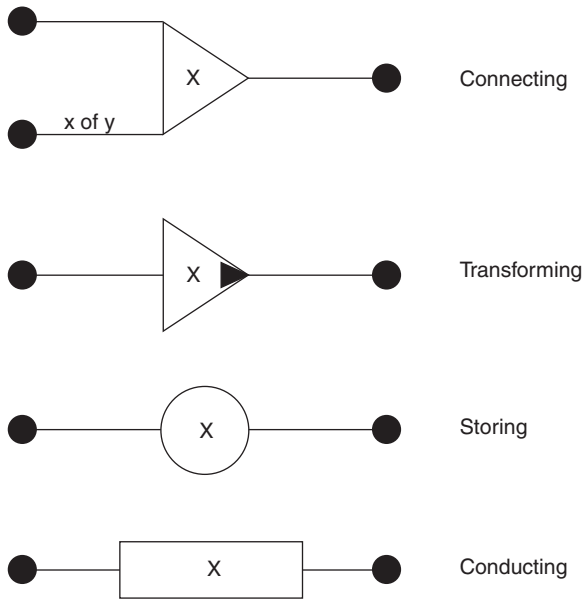


Fig. 2.1.23 Symbols for general functions (adapted from Roth, 1970).

4. Force a fit (e.g., use multiple horizontal rotors).
5. Generate ideas.
6. Develop the ideas.

In addition to synectics, other methods in this category include random stimulus, intermediate impossible, and concept challenge. In the random stimulus method, a word, object, or image, chosen at random, is linked to the original problem. In the intermediate impossible method, an ideal solution to the problem is thought up, and from it, gradually, a practical solution is developed. The concept challenge method undermines the problem statement in order to envision a new solution.

2.1.5.5.3 Analytic methods

The analytic methods are also called *systematic methods*. The method described in this section is a combination of function analysis (discussed in Section 2.1.5.4) and the morphological chart method. A morphological chart is a summary of subsolutions to subfunctions.

As mentioned in Section 2.1.5.4, the main function in the black box connecting inputs to outputs is replaced with several subfunctions, showing interrelationships among them as well as the inputs and outputs (e.g., Figure 2.1.21(b)); subfunctions can be added step by step. Next, the function structure is elaborated by changing, splitting, or combining subfunctions until the best function structure results. In the third step, the subfunctions are replaced with general function symbols. Figure 2.1.22 shows the symbols recommended by Pahl

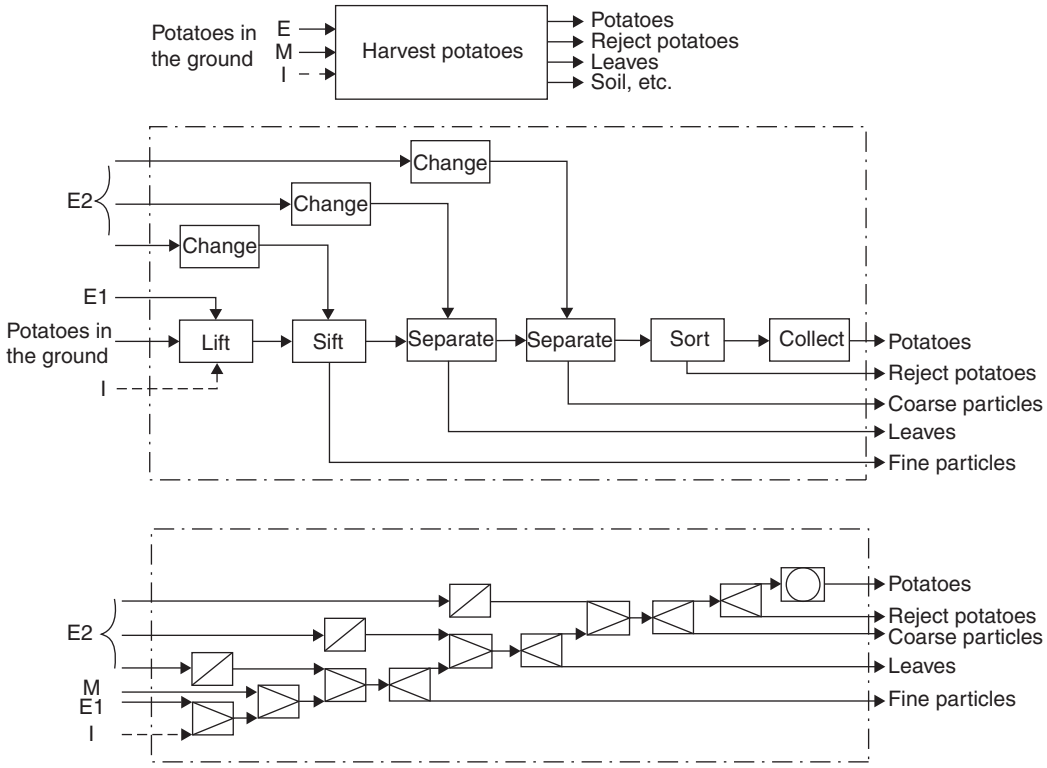


Fig. 2.1.24 Function analysis and structure for a potato harvesting machine (adapted from Pahl and Beitz, 1984).

and Beitz (1984). Roth (1970) also provides general function symbols, shown in Figure 2.1.23.

Figure 2.1.24, from Pahl and Beitz (1984), shows how the general function (harvesting potatoes) in the black box is replaced with subfunctions and these subfunctions with the general function symbols shown in Figure 2.1.22. Figure 2.1.25, from Roth (1970), using his symbols provided in Figure 2.1.23, shows how subfunctions can be added step by step.

Once the function analysis is complete, the task at hand is to develop a matrix of subfunctions as rows and possible solutions as columns. The best means to achieve each subfunction is identified. The selected solution box is highlighted. The combination of chosen solutions to subfunctions should yield the design solution to the problem. Figure 2.1.26 shows the morphological chart for potato harvesting from Pahl and Beitz (1984).

2.1.5.6 Evaluation and decision-making

Once a number of alternative designs have been developed, the designer must choose the best one. Consideration of the objectives is essential in making the final choice. Since there usually are multiple objectives, each with a different value, it is necessary to weight the objectives so that all designs may be compared objectively and across a wide range of objectives. Generally, a matrix of objectives with associated weights and the various evaluation criteria, such as cost, performance, and comfort, is prepared. Each design is evaluated against these criteria using a scale (a 5-point, 7-point, 10-point, or other). The scale graduations may be descriptive, such as excellent solution, poor comfort, and very economical design. The final score for each design is the sum of the scores for each design weighted for each criterion. The

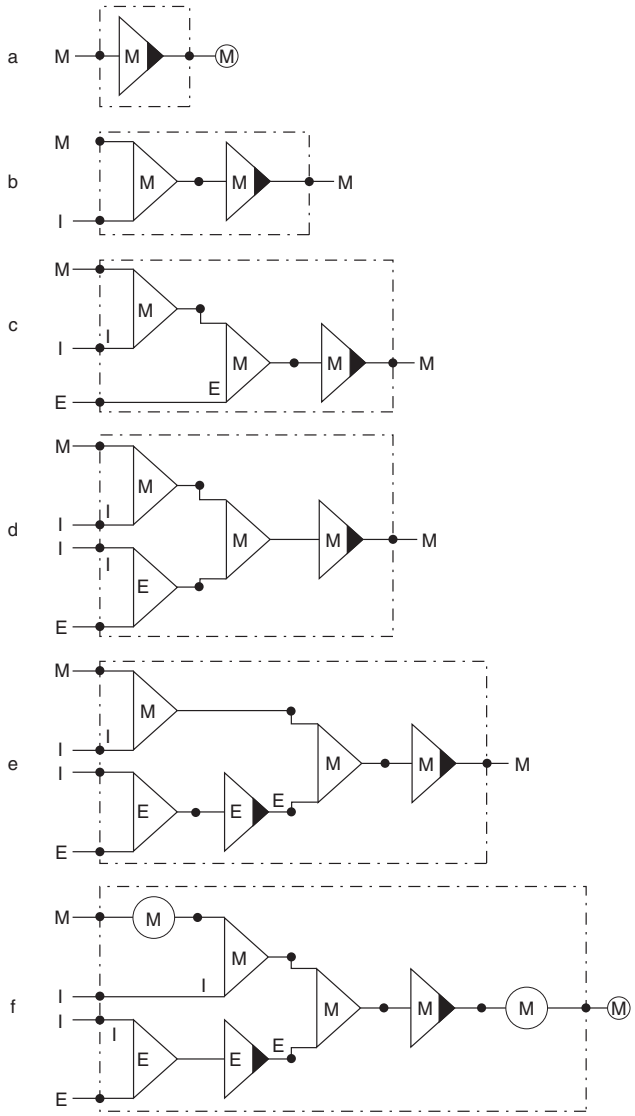


Fig. 2.1.25 Step-by-step addition of subfunctions for an electric coffee mill (adapted from Roth, 1970).

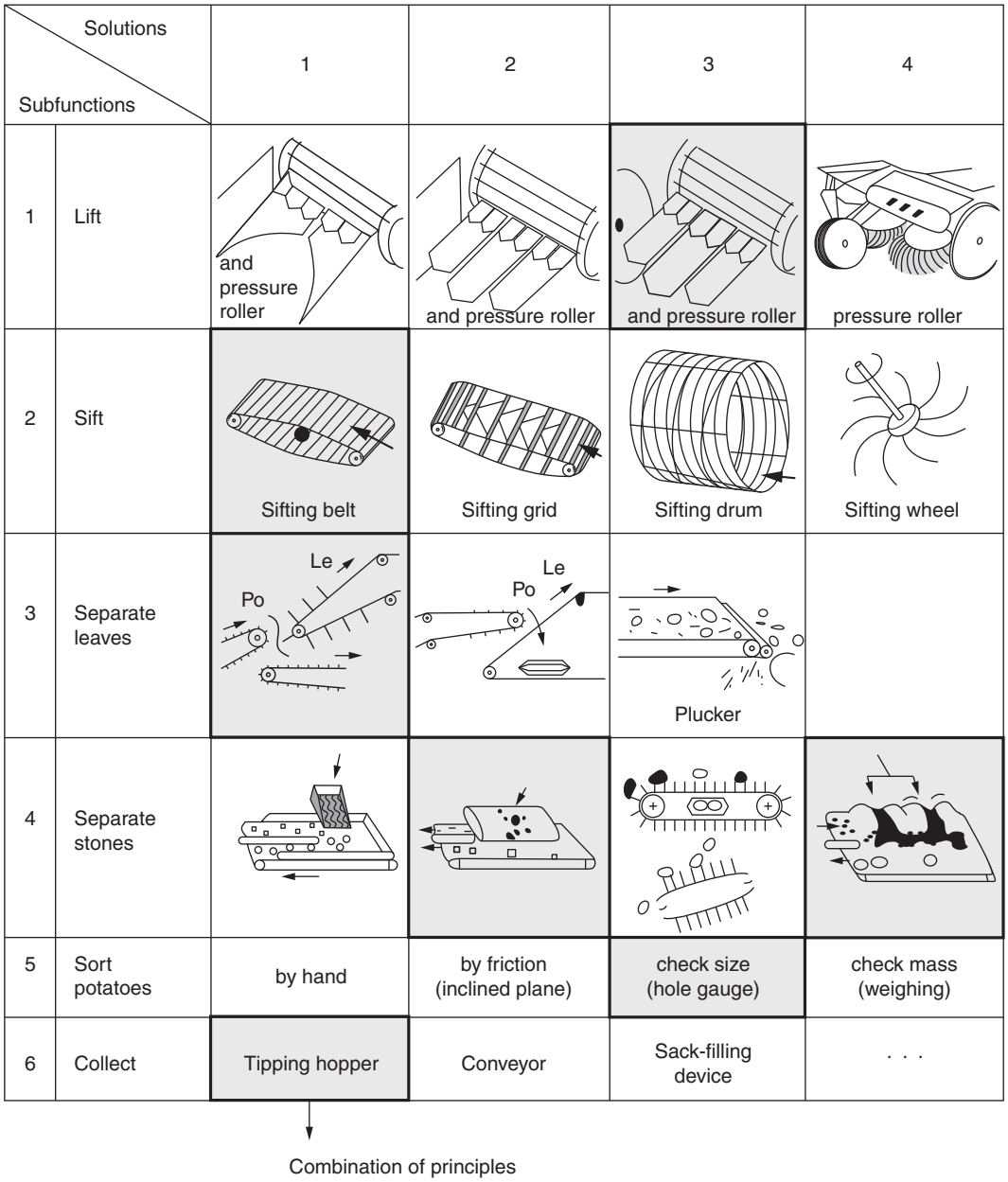


Fig. 2.1.26 Morphological chart for a potato harvesting machine (adapted from Pahl and Beitz, 1984).

Table 2.1.3 A simplified design decision matrix

		Cost			Comfort			Reliability		
Design criteria	**Total**	***W***	***S***	***T***	***W***	***S***	***T***	***W***	***S***	***T***
Design 1	0.5	9	4.5	0.3	7	2.1	0.2	8	1.6	8.2
Design 2	0.5	8	4.0	0.3	8	2.4	0.2	8	1.6	8.0
Design 3	0.5	9	4.5	0.3	9	2.7	0.2	9	1.8	9.0

W = weight; S = score; $T = W \times S$.

design scoring the highest is the winner. This, however, may not be the final design, as the designer may incorporate features that rated higher from other designs to modify the chosen design. Table 2.1.3 shows an example of automobile designs rated for various criteria. In this case, Design 3 is selected.

2.1.6 Summary

This chapter outlines the structure of the product design process. It defines *design* and briefly describes how the design process has changed over time. We also discuss some important design paradigms learned as a result of design failures. The basics of design requirements are outlined, and finally, we provide a systematic, step-by-step analytical design procedure. We realize that we condense a significant amount of information in this chapter, and the reader is encouraged to consult books devoted completely to the process of design. A number of references are provided next. Finally, we would like to the leave the reader with the impression that the end of this chapter is not the end of the product design process. Once the basic design is developed, it must be refined from the standpoint of quality, materials, processes, assembly and disassembly, maintenance, functionality, and usability. This refinement is discussed in Chapter 2.2.

References

Alexander, C. *Notes on the Synthesis of Form*. Cambridge, MA: Harvard University Press, 1969.

Caldecote, V. Investment in New Product Development. In *Product Design and Technical Innovation*, eds. R. Roy and D. Wield. Milton Keynes, UK: Open University Press, 1989.

Engineering News-Record, "A bridge to remember," McGraw-Hill, Inc. Section: Editorials; Pg. 190, January 21, 1982

French, M. J. *Engineering Design: The Conceptual Stage*. London: Heinemann, 1971.

Gordon, W. J. J. *Synectics: The Development of Creative Capacity*. New York: Harper and Row, 1961.

Hall, A. D. *A Methodology for Systems Engineering*, 6th ed. Princeton, NJ: Van Nostrand, 1968.

Hubka, V., and W. E. Eder. *Theory of Technical Systems: A Total Concept Theory for Engineering Design*. Berlin: Springer, 1988.

Hubka, V., M. M. Andreasen, and W. E. Eder. *Practical Studies in Systematic Designs*. London: Butterworth, 1988.

Jones, J. C. *Design Methods: Seeds of Human Futures*. New York: John Wiley, 1970.

Jones, J. C. *Design Methods*. Chichester, UK: John Wiley, 1981.

Keeney, R. L. *Value-Focused Thinking: A Path to Creative Decision-Making*. Cambridge, MA: Harvard University Press, 1992.

Osborn, A. F. *Applied Imagination: Principles and Procedures of Creative Problem Solving*, 3rd ed. New York: Scribner, 1963.

Pahl, G., and W. Beitz. *Engineering Design*. London: Design Council, 1984.

Petroski, H. *Design Paradigms: Case Histories of Error and Judgement in Engineering*. New York: Cambridge University Press, 1994.

Pighini, U., G. Francesco, Y. D. Zhang, S. A. Vicente, and A. Rivalta. The Determination of Optimal Dimensions for a City Car Using Methodical Design with Prior Technology Analysis. *Design Studies* 4, no. 4 (1983): 233–243.

Prince, G. M. *The Practice of Creativity: A Manual for Dynamic Group Problem Solving*. New York: Harper and Row, 1970.

Pugh, S. *Total Design; Integrated Methods for; Successful Product Engineering*. Wokingham, UK: Addison Wesley, 1990.

Pye, D. The Nature of Design. In *Product Design and Technical Innovation*, eds. R. Roy and D. Wield. Milton Keynes, UK: Open University Press, 1989.

Roozenburg, N. F. M., and J. Eekels. *Product Design: Fundamentals and Methods*. New York: John Wiley, 1995.

Roth, K. "Systematik der Machine und ihre Elemente." *Feinwerktechnik* 11 (1970): 279–286.

Schon, D. A. Design in the Light of the Year 2000. *Student Technologist* (Autumn 1969); quoted by N. Cross, The Changing Design Process. In *Product Design and Technical Innovation*, eds. R. Roy and D. Wield. Milton Keynes, UK: Open University Press, 1989.

VDI Standards 2221 and 2222. *Systematic Approach to the Design of Technical Systems and Products*. Dusseldorf, Germany: Verein Deutscher Ingenieure, 1987.

This page is intentionally left blank

Chapter 2.2

2.2

Designing to ensure quality

Anil Mital, Anoop Desai, Anand Subramanian and Aashi Mital

2.2.1 Introduction

The creativity in product design and process selection is the critical component in ensuring quality in the product and its associated processes. In previous chapters, we presented details of the design process, strategies, and tools. Now, we focus on the need to integrate quality into the product design and the design review process.

Product quality constitutes a crucial component in the product design process, directly affecting consumer loyalty and company profitability. Historically, manufacturing enterprises relied on the reactive approach of inspecting the quality of a product to ensure it conforms to design specifications. While this approach has its own advantages, its principal limitation lies in the manufacturers' implicit resignation to the fact that quality needs to be inspected in since it cannot be built into the product design at the design stage. However, there has been a gradual, yet definite, transition from a reactive to a proactive strategy to managing quality by incorporating design techniques that do away with the largely unproductive inspection process. Several leading manufacturing enterprises have been successful in entirely eliminating the need to inspect by adopting a proactive approach to product design.

Quality is defined as a physical or nonphysical characteristic that constitutes the basic nature of a thing or is one of its distinguishing features. This broad definition of *quality* can be extended to the engineering and industrial realm by defining *quality* as a characteristic or group of characteristics that distinguish one article from another or the goods of one manufacturer from those of its competitor (Radford, 1992). The two different aspects of this characteristic can be classified into objective and subjective categories. The objective nature of quality is independent of the existence of mankind, whereas the subjective nature of this characteristic is expressed in terms of the general manner in which humans perceive it. Figure 2.2.1 shows the nine dimensions of quality with their meanings. These dimensions are relatively independent of each other; hence, a product can be excellent in one dimension, average in another, and poor in still another dimension. Therefore, quality products can be determined by using a few of the dimensions of quality.

When the expression *quality* is used, we usually think in terms of an excellent product that fulfills or exceeds our expectations. The International Organization for Standardization (ISO) extends this definition to include the service industry as well. The ISO defines *quality* as "the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs" (ISO Standard 9000).

2.2.1.1 Why quality control?

The American Society for Quality Control (ASQC, 1991) commissioned a Gallup poll of over 3000 consumers in the United States, Japan, and Germany in the 1990s. The survey pointed to significant commonality in customer thinking regarding quality. Performance, price, and reputation were three of the most valuable attributes customers desired when selecting one company's product over another's. The importance of customer satisfaction and loyalty to a company's profitability and ability to compete in a global marketplace can hardly be overemphasized. In formulating manufacturing strategies,

Product Development; ISBN: 9780750683098
Copyright © 2008 Elsevier Ltd. All rights of reproduction, in any form, reserved.

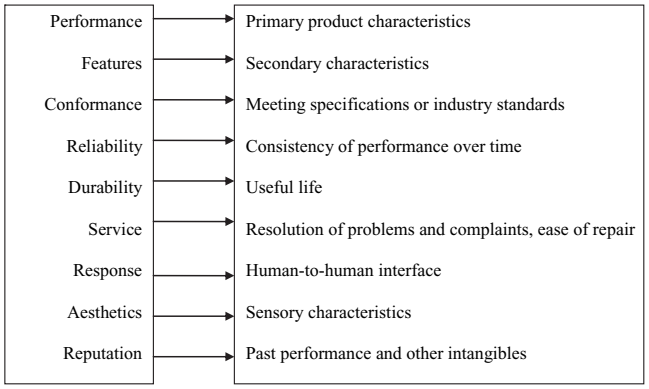


Fig. 2.2.1 Quality dimensions (adapted from Garvin, 1988)

quality is the most important factor in determining market success (Hill, 1989). "The quality of our product is excellent, but the price is too high" does not make a successful product marketing strategy. There must be a balance between quality loss and product price. The price represents the loss to the customer at the time of purchase, and poor quality represents an additional loss during the use of the product. A goal of quality control is to minimize such losses to the customer.

Given the obviously undeniable importance of incorporating quality into product or service design, it is necessary to draw a distinction between quality assurance and quality control. Quality assurance consists of all planned and systematic actions that are necessary to provide adequate confidence that a product or service will satisfy given requirements for quality (ISO Standard 9004). Quality control, on the other hand, is a system whereby quality is assured economically (Japan Industrial Standards, JIS Z Standard 8101). Generally speaking, quality is created through processes that the manufacturer develops and maintains (Kolarik, 1999). Each of these processes plays an integral role in satisfying the eight fundamental requirements for product or service conception, development, delivery, and disposal. Figure 2.2.2 depicts the creation of quality as just described by the sequential achievement of eight fundamental processes. Bear in mind that this sequence of activities has to be approached systematically if customers are to have a positive quality experience.

Creation of Quality	Experience of Quality
Qualities of definition Qualities of design Qualities of development Qualities of production Qualities of delivery Qualities of sales and service Qualities of use Qualities of disposal/recycle	Benefits and burdens most apparent to internal customers (e.g., job challenge, empowerment, recognition for work, pay, hours) Benefits and burdens most apparent to external customers (e.g., product function, form and fit, long life, cost, and delivery time)

Fig. 2.2.2 Creation and experience of quality (adapted from Kolarik, 1999).

2.2.1.2 Reactive versus proactive quality control

Quality control strategies can be distinctly classified into two categories: reactive and proactive. The majority of quality control strategies are aimed at detecting and correcting problems that already exist. In other words, the designer of a product, process, or service incorporates a system of checks and measures to isolate and catch defects as and when they occur. By their very nature, reactive quality control strategies are better suited to identify problems and resolve them and, as such, are clearly defensive in nature. Reactive strategies try to limit losses by incorporating the largely wasteful inspection process. The reactive strategy emphasizes traditional loss accounting and data-intensive statistical inferences to justify action. This course of action is comforting for a decision maker, since action can be readily justified based on historic data (Kolarik, 1999). Traditionally, most of the classical topics of quality assurance and statistical quality control currently being taught in universities are largely reactive in nature (Banks, 1989; Duncan, 1986; Grant and Leavenworth, 1988; Montgomery, 1991).

The proactive approach to quality control is based on the cause and effect relationship and, as such, does away almost completely with historic data and related statistical analyses. This approach is based on the premise that, if a manufacturer has a reasonably good understanding of the expectation of quality of its customers (relative to its competitors), then processes aimed at creating high quality can be structured in accordance with those requirements. This concept tends to reduce risk of business failure since all operations are geared toward customer satisfaction.

The advantages of adopting a proactive approach to quality control, as enlisted here, is twofold: accelerated product and process development cycles, and avoidance, not management, of losses. This leads to lower equipment downtime.

It is obvious that both advantages result in a keener competitive edge in an increasingly competitive global marketplace. Quality systems and quality-transformation-related strategies have proven their importance by improving productivity, especially when appropriate leadership, training, and managerial structures are provided (Kolarik, 1999). Product and process planning strategies have myriad forms. For instance, quality function deployment is a product planning tool used most commonly during the design stages of a new product. Quality function deployment (QFD) has been observed to render extremely positive results in generating customer satisfaction and enhancing market share as well as profits (Akao, 1990; Juran, 1988). Some examples of the emergence of proactive strategies include robust design and early off-line experimental programs that

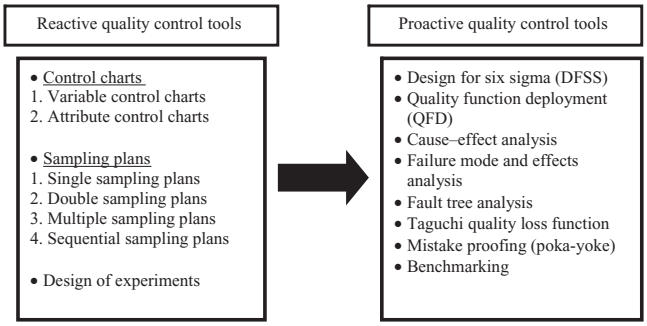


Figure 2.2.3 The transition from reactive to proactive quality assurance techniques.

support system, parameter, and tolerance design (Kolarik, 1999; Taguchi, 1986). Mistake proofing, commonly referred to as *poka-yoke* in industry, and source inspection are examples of yet more proactive strategies that stress the prevention and elimination of quality problems (Shingo, 1986).

From the preceding discussion, it is clear that proactive quality control strategies, when used strategically, enable the designer to design and build high quality into the product. This does away with time, effort, and resources wasted at the end of the production process. The proactive approach to quality control emphasizes the "do it right the first time" approach to manufacturing, quality, and ultimately customer satisfaction. Figure 2.2.3 is a graphical depiction of this transition.

2.2.2 Procedures for incorporating high quality in design stages

The reactive approach to quality control is commonly associated with suboptimal use of resources, a high amount of waste, and shrinking market share. This not only adds to the product price but also leads to an increase in the lead time from manufacturing to marketing the product, resulting in untimely release to the market and a loss of market share. To ensure high quality, it needs to be built into the product beginning at the design stage itself instead of inspecting for it at a later stage. Here, we discuss some of the frequently used proactive quality control techniques used to solve problems in industry.

2.2.2.1 Design for six sigma

The term *six sigma* has several meanings. At the most encompassing level, a corporation can define it as a philosophy, a way of thinking. Technically, six sigma is a data-driven approach to reducing the defects produced due to a variation in a product or process.

Design for six sigma (DFSS) can be described as a comprehensive approach to product development. It links business and consumer needs to critical product attributes to product functions to detailed designs to tests and verification. This is achieved by integrating a full suite of analytical methods from six sigma including quality function deployment, goal deployment, failure mode and effect analysis (FMEA), simulation, quality loss functions, and prototyping to create more reproducible, customer-driven designs faster.

Similar to six sigma, DFSS uses a structured set of steps, DMADV (define, measure, analyze, design, and validate; see Figure 2.2.4), to ensure repeatability and continuous improvement. DFSS focuses on translating customer requirements qualitatively and quantitatively to product specifications, then ensuring that proposed designs robustly meet customer-defined scorecards. The DMADV process is to fundamentally redesign a process.

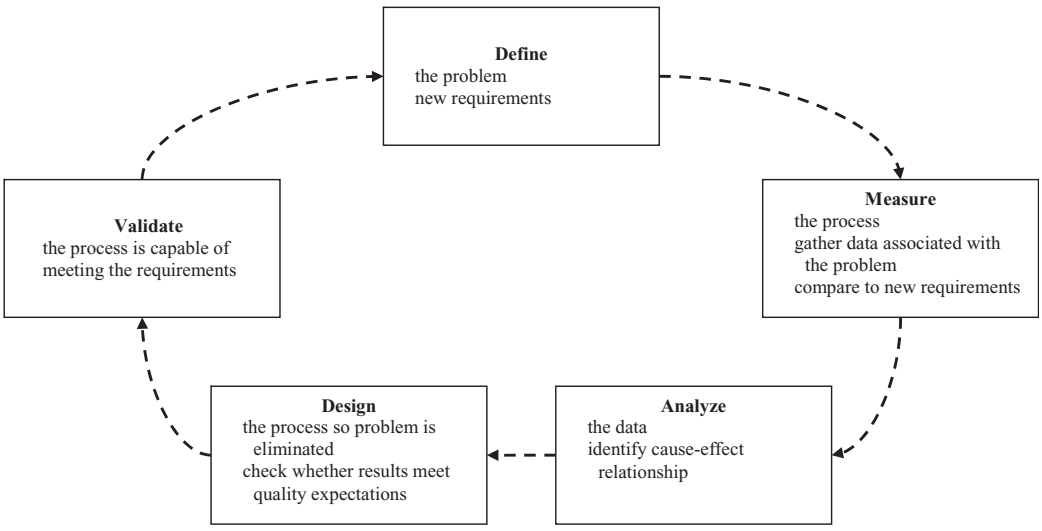


Fig. 2.2.4 The DFSS approach to product and process design, DMADV.

It also may be used to design a new process or product when requirements change. The goal of DFSS is to design a new process or product to replace the incapable existing process or product.

DFSS can be used to maximize the performance of products as well as services. DFSS can be utilized to enhance the speed and quality of the design processes within an organization. Organizations use DFSS when they have to design or redesign a process, product, service, or transaction. DFSS gets to the source of product and service imperfections by "designing in" performance from the earliest stages of research and development. It teaches people a methodical approach to involving the right people, asking the right questions, and using the right tools from the very beginning of any design project. DFSS can be applied to any industry and any product or process design methodology. It can be used to create new products and services, streamline software design and systems integration or improve existing product performance. By implementing DFSS, companies can avoid costly redesign projects by ensuring that products and processes are designed, built, and launched with greater reliability and a higher performance-to-cost ratio.

2.2.2.2 Mistake proofing (*poka-yoke*)

Mistake proofing a product's design and its manufacturing process is an important element of design for *X*. Mistake proofing also is a key element in improving product quality and reliability and an element of the DFSS concept.

The Japanese concept of *poka-yoke* (mistake proofing) seeks to find and correct problems as close to the source as possible. This is because finding and correcting defects caused by errors cost more and more as a product or item flows through a process. Over time, more emphasis has been placed on the design of the product to avoid mistakes in production. Often the benefits of mistake proofing not only help with production of the product but also contribute to correct user operation and maintenance as well as servicing the product.

The concept of mistake proofing involves finding controls or features in the product or process to prevent or mitigate the occurrence of errors and requires simple, inexpensive inspection (error detection) at the end of each operation to discover and correct defects at the source.

There are six mistake-proofing principles or methods. These are listed in order of preference or precedence in the manner in which mistakes are addressed (Belliveau, et al., 2002):

1. Elimination seeks to eliminate the possibility of error by redesigning the product or process so that the task or part is no longer necessary; for instance, product simplification or part consolidation that avoids a part defect or assembly error in the first place.
2. Replacement substitutes a more reliable process to improve consistency; for example, the use of robotics or automation that prevents a manual assembly error, automatic dispensers or applicators to ensure the correct amount of a material, such as an adhesive, is applied.
3. Prevention engineers the product or process so that it is impossible to make a mistake.
4. Facilitation employs techniques and combines steps to make work easier to perform; for example, visual controls including color coding, marking, or labeling parts to facilitate correct assembly; exaggerated asymmetry to facilitate correct orientation of parts; a staging tray that provides a visual control that all parts were assembled, locating features on parts.
5. Detection involves identifying an error before further processing occurs so that the problem can be corrected quickly; for example, sensors in the production process to identify when parts are assembled incorrectly.
6. Mitigation seeks to minimize the effects of errors; for example, simple rework procedures when an error is discovered and extra design margin or redundancy in products to compensate for the effects of errors.

Ideally, mistake proofing should be considered during the development of a new product to maximize opportunities to mistake proof through design of the product and the process (elimination, replacement, prevention, and facilitation), because over 70% of a product's life cycle costs can be attributed to its design stage. Once the product is designed and the process selected, mistake-proofing opportunities are more limited (prevention, facilitation, detection, and mitigation).

2.2.2.3 Quality function deployment

Quality function deployment was originally developed by Yoji Akao and Shigeru Mizuno in the early 1960s. They extended the original "house of quality" (HOQ) approach by deploying "hows" resulting from the top-level HOQ into lower-tier matrices addressing aspects of product development, such as cost, technology, and reliability. The basic quality function deployment methodology involves four phases that occur over the course of the product development process. During each phase, one or more matrices are constructed to plan and communicate critical product and process planning and design information. The QFD methodology flow is depicted in Figure 2.2.5.

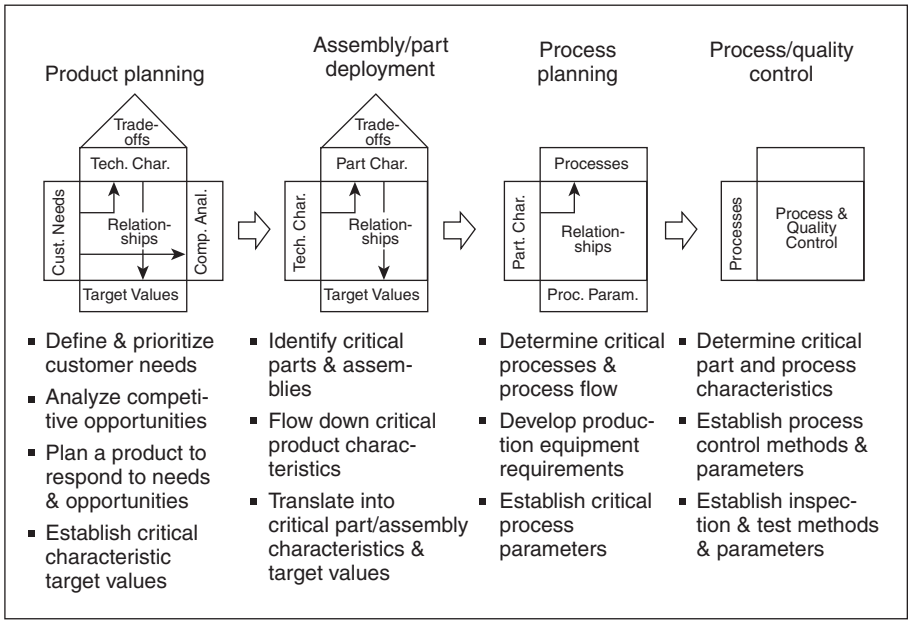


Fig. 2.2.5 The four phase QFD approach to building quality (adapted from Kolarik, 1999).

Once customer needs have been identified, preparation of the product planning matrix or "HOQ" can begin. The product planning matrix is prepared as follows:

1. State the customer needs or requirements on the left side of the matrix. These are organized by category, based on the affinity diagrams. It needs to be ensured that the customer needs or requirements reflect the desired market segment(s). If the number of needs or requirements exceeds 20–30 items, the matrix is decomposed into smaller modules or subsystems to reduce the number of requirements. For each need or requirement, state the customer's priorities using a 1–5 rating. Ranking techniques and paired comparisons are used to develop priorities.
2. Evaluate prior generation products against competitive products. Surveys, customer meetings or focus groups and clinics are used to obtain feedback. Competitor's customers are included to get a balanced perspective. Identify price points and market segments for products under evaluation. Warranty, service, reliability, and customer complaint problems are taken into account to identify areas of improvement. Based on this, a product strategy is developed. Consider the current strengths and weaknesses relative to the competition. Identify opportunities for breakthrough's to exceed competitors' capabilities, areas for improvement to equal competitors' capabilities, and areas where no improvement will be made. This strategy is important to focus development efforts where they will have the greatest payoff.
3. Establish product requirements or technical characteristics to respond to customer requirements and organize them into related categories. It is important to note that characteristics should be meaningful, measurable, and global.
4. Develop relationships between customer requirements and product requirements or technical characteristics. This involves using symbols for strong, medium, and weak relationships.
5. Develop a technical evaluation of prior generation products and competitive products. Get access to competitive products to perform product or technical benchmarking. Perform this evaluation based on the defined product requirements or technical characteristics. Obtain other relevant data, such as warranty or service repair occurrences, and costs and consider this data in the technical evaluation.
6. Develop preliminary target values for product requirements or technical characteristics.
7. Determine potential positive and negative interactions between product requirements or technical characteristics, using symbols for strong or medium, positive or negative relationships. Too many positive interactions suggest potential redundancy in the "critical few" product requirements or technical characteristics. Focus on negative interactions, consider product concepts or technology to overcome these potential trade-offs or consider the trade-offs in establishing target values.
8. Next, calculate importance ratings. This involves assigning a weighting factor to relationship symbols

and multiplying the customer importance rating by the weighting factor in each box of the matrix. Finally, all resulting products in each column are added.

9. Develop a difficulty rating (1–5 point scale, where 5 means very difficult and risky) for each product requirement or technical characteristic. Avoid too many difficult/high-risk items, as this will likely delay development and exceed budgets. Assess whether the difficult items can be accomplished within the project budget and schedule.
10. Analyze the matrix and finalize the product development strategy and product plans. Determine required actions and areas on which to focus. Finalize target values. Are target values properly set to reflect appropriate trade-offs? Do target values need to be adjusted, considering the difficulty rating? Are they realistic with respect to the price points, available technology, and the difficulty rating? Are they reasonable with respect to the importance ratings? Determine items for further QFD deployment. To maintain focus on the critical few, less-significant items may be ignored with the subsequent QFD matrices. Maintain the product planning matrix as customer requirements or conditions change.

The "HOQ" matrix often is called the *phase one matrix*. Figure 2.2.6 shows a typical HOQ and its elements. In the QFD process, a phase two matrix translates finished product specifications into attributes of design (architecture, features, materials, geometry, subassemblies, and component parts) and their appropriate specifications. Sometimes, a phase three matrix is used in attributes of design specifications into manufacturing process specifications (temperature, pressure, viscosity, rpm, etc.). Some key elements that determine the successful implementation of QFD follow:

- Management must make it clear that QFD is a high priority.
- Set clear priorities for QFD activities. Specifically, management needs to allocate resources for and insist on execution of market research and technical competitive assessment.
- Make QFD training available, preferably "just in time" to use QFD.
- Insist that decisions be based on customer requirements.
- Understand the terms used in QFD.
- Insist on cross-functional commitment and participation.
- Become leaders of QFD rather than managers.

One guideline for successful development of QFD matrices is to keep the amount of information in each matrix at a manageable level. With a more-complex product, if 50 potential needs or requirements were identified and these were translated into an equal or even greater number of product requirements or technical characteristics, there would be more than 2500 potential relationships to plan and manage. Generally speaking, an individual matrix should not address more than 20 or 30 items on each dimension of the matrix. Therefore, a larger, more-complex product should have its customers' needs further classified into lower hierarchical levels.

To conclude, a product plan is developed based on initial market research or requirements definition. If necessary, feasibility studies or research and development are undertaken to determine the feasibility of the product concept. Product requirements or technical characteristics are defined through the matrix,

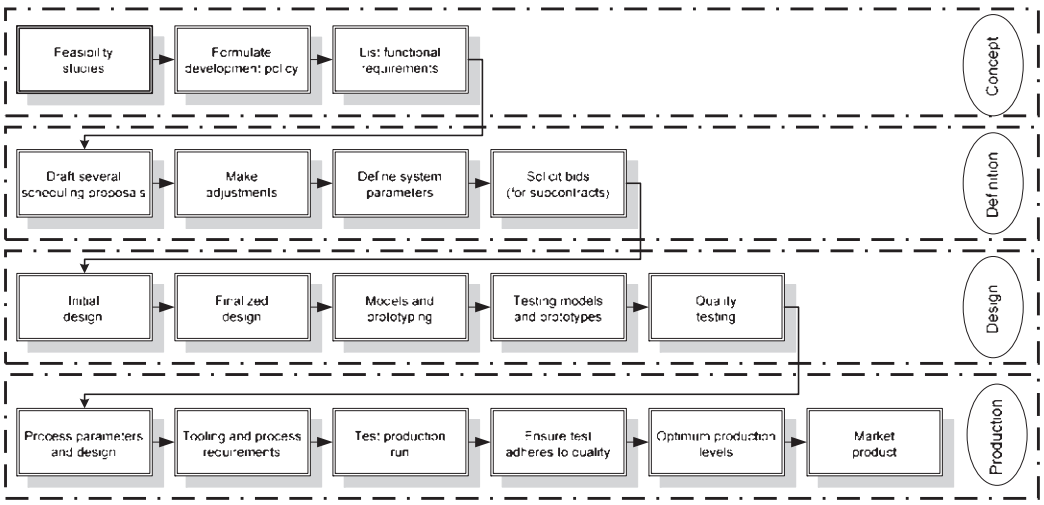


Fig. 2.2.6 Design review process and product development phases.

a business justification is prepared and approved, and product design then commences.

2.2.2.4 Design review

The purpose of a design review is to provide a systematic and thorough product-process analysis, a formal record of that analysis, and feedback to the design team for product and process improvement. According to JIS Z 8115-1981, design review is the judgment and improvement of an item at the design phase, reviewing the design in terms of function, reliability, and other characteristics, with cost and delivery as constraints and with the participation of specialists in design, inspection, and implementation. The design review process allows for an independent critique of a product and its related process at appropriate time intervals in the product life cycle. It is an important tool to identify product-process bottlenecks using the proactive quality control strategy.

Formal design reviews (FDRs) should be performed from an independent perspective. A key requirement for a reviewer in the FDR is that he or she be an expert in that field and hold no vested interest in nor be directly responsible for the product or process under review. This constraint helps the process elicit unbiased comments, concerns, and recommendations. Development of an effective design review requires a team of functional experts (the reviewers) and a product or process life cycle plan or program (which is the subject of review).

Figure 2.2.6 shows the steps in a typical design review process. In the conceptional design review stage, designers ensure that the initial design direction maps to the business goals and user needs and review the design for alignment with broader initiatives and possible integration with other product designs. In the standards check point stage, the designs are reviewed to meet appropriate standards for consistency, accessibility, usability, and the like. The user-interface design review is used to review specific interactions and provide guidance to designers on problematic issues. Finally, the creative design review is used to ensure that the visual design maps to the creative direction of the project.

Design reviews vary in both formality and structure. Reviews typically are planned into the design process so that often we see a number of reviews conducted in sequence. Each design review focuses on one or more aspects of the design or plan as it proceeds across the product life cycle, from needs assessment to disposal considerations. These individual reports are fed back to the design and planning team for action and resolution. Each stage is recorded sequentially, and actions taken are formally entered into the record. Based on these reports, formal documentation is developed and becomes a permanent part of the quality record.

Typically, there are two forms of design review: internal and external. The internal design review process looks into the manufacturability and design part of the product life cycle. It addresses the feasibility of the design with respect to function, form, and fit; the producibility of the design with respect to the production and process capabilities, sales and service capabilities, cost–volume–profit estimates; and the economic feasibility of the product. The external design review process addresses the customer, target market, and their true quality characteristic demands. This review provides input on whether the product will provide customer satisfaction and feedback on what can be done to improve the design accordingly.

Some commonly associated problems with the implementation of the design review process are

- Unevenly matched skills and knowledge among the design review team.
- Lack of communication between product developers and the related departments.
- No time to make design-review-based changes.
- Lack of design review experience.
- Each department considers design review a separate stage and not included in the initial design process.

The design review process, when utilized correctly, serves as a birth-to-death record of the product, processes, and the organization's diligence in serving its target market. A review of the records of a product shows the major concerns or warnings, their treatment and resolution records, and a record of how timely they were resolved.

Figure 2.2.7 shows the design review sequence. The key elements of the design review system are the soft-hard (SH) review and the FMEA processes.

2.2.2.4.1 SH review

The SH review addresses the need to design a product for safety, that is, in terms of the real-world conditions. The soft reviews look into the careless misuse of products by users, beyond normal wear and tear. The hard reviews look into the loss of function due to malfunctions or deterioration of each component over the estimated service life. The records generated using the SH review are an important data resource to help designers identify the various methods by which a particular product has been used throughout its life and, hence, make changes to design accordingly. Moreover, even though safety standards often exist, they do not adequately address every product use situation. SH reviews help fill in the gaps. A more important advantage of SH reviews is in determining the need for backup systems. The secondary

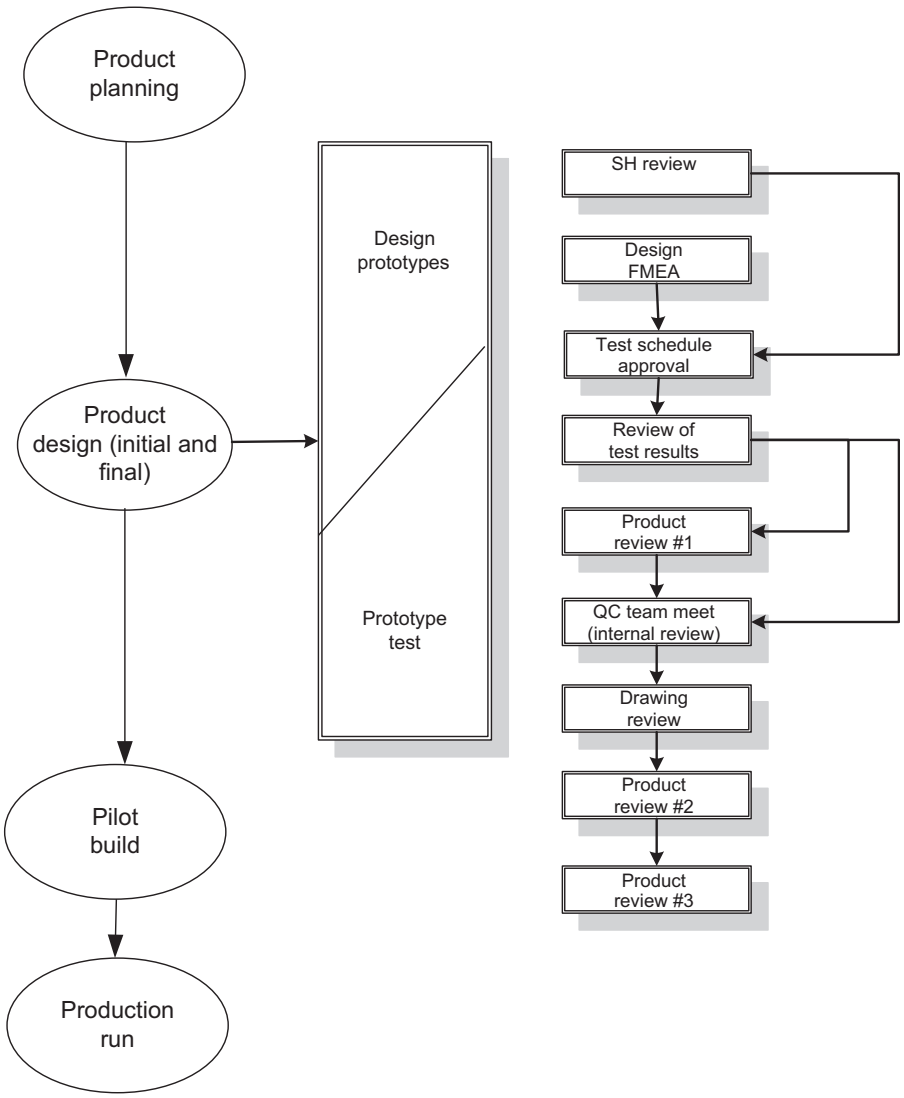


Fig. 2.2.7 The design review sequence (adapted from Ichida, 1996).

safety backup systems might be developed to prevent sudden breakdown of components.

2.2.2.4.2 Failure mode and effects analysis

FMEA examines potential failures in products or processes. FMEA helps select remedial actions that reduce cumulative impacts of life-cycle consequences (risks) from a systems failure (fault). FMEA can be explained as a group of activities intended to recognize and evaluate the potential failure of a product or a process and its effects, identify actions that could eliminate or reduce the chance of the potential failure, and document the process.

This method illustrates connections among multiple contributing causes and cumulative (life-cycle) consequences. It is used in many formal quality systems such as QS-9000 or ISO/TS 16949 (Chrysler Corporation, Ford Motor Company and General Motors, 1995).

The basic method is to describe the parts of a system and list the consequences if each part fails. In most formal systems, the consequences then are evaluated by three criteria and associated risk indices: severity (S), likelihood of occurrence (O) or probability of occurrence (P), and inability of controls to detect it (D).

Each index ranges from 1 (lowest risk) to 10 (highest risk). The overall risk of each failure is the risk priority number (RPN), which is the product of the severity, occurrence, and detection rankings: $\text{RPN} = S' O' D$.

A basic FMEA consists a set of nine columns:

1. Function, equipment, or process identification
2. Function, equipment, or process purpose
3. Interfaces

4. Failure mode
5. Failure mechanism
6. Failure detection
7. Failure compensation
8. Failure effects
9. Preventive measures

FMEA is a fundamental tool, useful in improving reliability, maintainability, safety, and survivability of products and processes. It encourages systematic evaluation of a product or process, recognition of hazards and potential failures, effects of the failures, the countermeasures to eliminate the failures or create secondary safety systems, and their documentation.

2.2.2.4.3 Experimental design

The purpose of experimental designs is to provide a systematic plan of investigation and analysis, based on established statistical principles, so that the interpretation of the observations can be defended as to technical relevance. The objective is to determine those variables in a process or product that form critical parameters and their target values. By using formal experimental techniques, the effect of many variables can be studied at one time. Changes to the process or product are introduced in a random fashion or by carefully planned, highly structured experiments. Planned experiments consist of six basic steps:

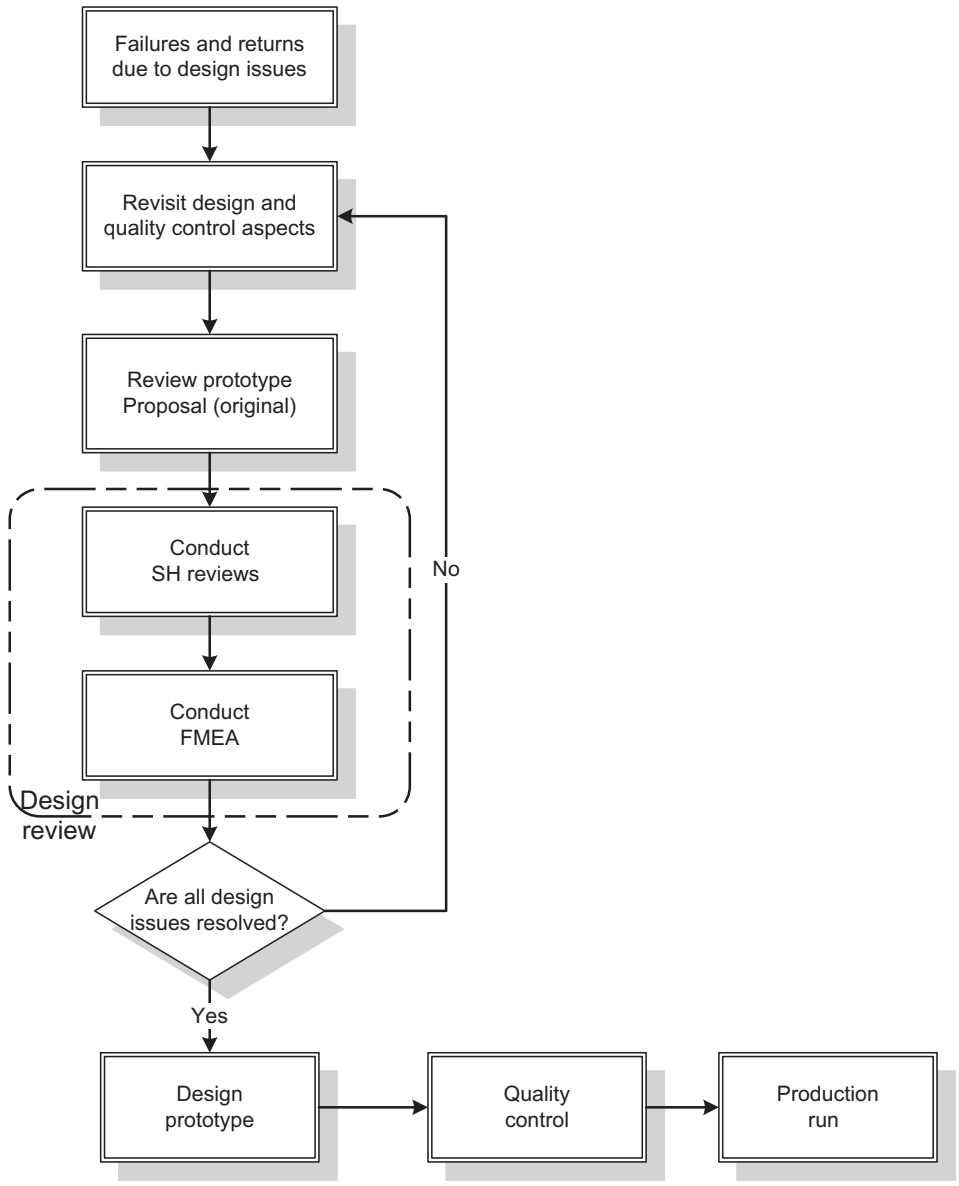


Fig. 2.2.8 Design review sequence adopted by Z-Air Systems.

1. Establish the purpose
2. Identify the variables
3. Design the experiment
4. Execute the experiment
5. Analyze the results
6. Interpret and communicate the analysis

There are three approaches to designing experiments: classical, Taguchi, and Shainin. Most common among these is the classical approach, based on the work of Sir Ronald Fischer in agriculture during the 1930s. In the classical approach, relevant variables are clearly identified and defined. An appropriate randomization and replication structure is developed within the context of the selected model. Three major analyses are used to generate the results: descriptive (which includes summary statistics and graphs), inferential (which includes formal hypothesis), and predictive (which includes the models used to predict future responses).

When designing experiments, statisticians generally begin with a process model of the black box variety, with several discrete or continuous input factors that can be controlled. Controlling variables implies varying them at will. This is coupled with one or more measured outputs or responses. The outputs or responses are assumed to be continuous. Experimental data are used to derive an empirical or approximation model linking the outputs and inputs. These empirical models generally contain what are known as *first-* and *second-order terms*. Often, the experiment has to account for a number of uncontrolled factors, which may be discrete, such as different machines or operators, or continuous, such as ambient temperature or humidity.

Taguchi's method, commonly known as *robust design*, improves the performance of a product by minimizing the effect of the causes of variation, without eliminating the causes. According to Taguchi, et al. (1989), quality is the loss a product causes to society after being shipped, other than any losses caused by its intrinsic functions. Taguchi's definition of *quality* and concept of loss minimization lead to robust design. Robust design consists of a conceptual framework of three design levels: system design, parameter design, and tolerance design. Within these design levels, quantitative methods based on loss functions are developed.

System design denotes the development of a basic prototype design that performs the desired and required functions of the product with minimum deviation from target performance values. It focuses on the relevant product or process technologies and approaches. It includes selection of materials, parts, components, and assembly systems.

Parameter design is a secondary design level, within or below the system design level. The main focus here is to ascertain the optimal levels for the parameters of each element in the system, to minimize the functional deviations of the product. The point is to meet the performance target with the least expensive materials and processes and to produce a robust product, one that is on target and insensitive to variations. It is the process of optimizing the functional design with respect to both cost and performance. Taguchi's method concentrates on designed experiments and specialized signal-to-noise ratio (SNR) measures.

Tolerance design is the process of determining the tolerance of each parameter by trading off quality loss

Table 2.2.1 S factor analysis (careless misuse)

Item	Description	Pass/fail	Reason for rejection
Performance related			
Automatic shutoff	Tip-over switch activated	0	N/A
Overheating	Motor thermal fuse	0	N/A
Continuous operation (switch left on)			N/A
User related			
Instruction not followed			N/A
Left too close to wall (or other object)	Motor thermal fuse	0	N/A
Operated in damp conditions	Motor thermal fuse	0	N/A

Table 2.2.2 H factor analysis (product safety function)

Component	Hazardous condition	Automatic response	Comment
Motor thermal fuse (improper operation)	Coil starts overheating	Motor shuts off	Replace motor
Power switch (left on)	Improper use causes overheating	Motor shuts off	Automatic shutoff function
Timer	Set at high temperature	Motor shuts off	Automatic shutoff function

Part No: 3C1749	Part Description: Space Heater Fan	FAILURE MODE EFFECT ANALYSIS			Design Lead: Analysis Period:				Michael J. Fischer Jan 2006 - March 2006
Component	**Function**	**Failure Mode**	**Effect of Failure**	**Causes of Failure**	**# of Occurrences**	**Severity Rating**	**Detection Rating**	**RPN**	**Preventive Actions taken**
Frame	Provide support, hold motor assembly	Corrosion and scale formation	Motor falls from frame	Metallic coating insufficient	4	5	2	40	Change metal coating process to _____
				Vapor condensation due to inadequate through flow of air	6	4	2	48	Confirm with condensation tests
				Improper brazing process	3	6	1	18	Use different joining process
		Frame breakage	Bends and breaks	Improper packaging	5	3	1	15	Improve packaging and cushion for products
				Material of frame not able to withstand vibrations due to inadequate hardness	8	8	3	192	Change material of the frame to _____

Figure 2.2.9 Sample design FMEA sheet.

and cost. The parameter design sets up the targets for the process parameters. The tolerance design step is a logical extension of parameter design to a point of complete specification or requirement. Obviously, this step determines the most economic tolerances, those that minimize product cost for a given tolerance deviation from target values.

2.2.3 Case studies

2.2.3.1 A design review case study

The design team of Z-Air Systems is facing a predicament. The newly launched space heater system is running into rough weather due to product failures and

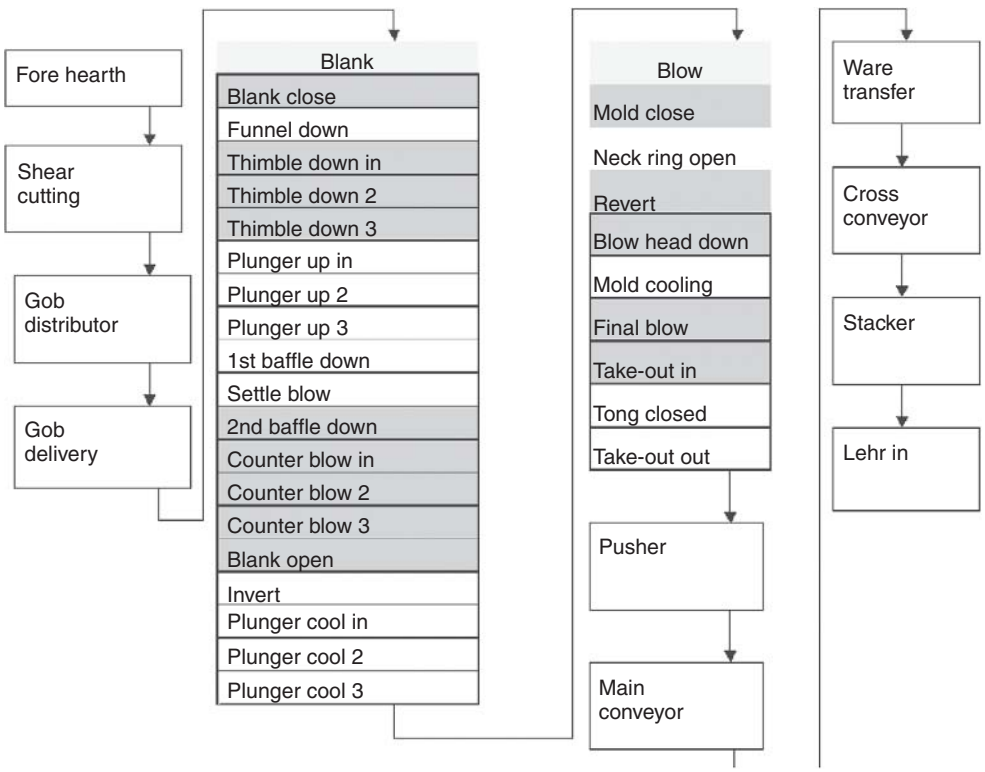


Figure 2.2.10 High-level process mapping for the bottling plant.

returns from customers. Market research indicates that about 80% of the problems are related to product performance and quality. Analysis of the problems indicate that most issues are due to design defects. Management formed the opinion that going back to the drawing board can solve the problem. The design team is under pressure to produce results, and it resorts to a comprehensive design review program. The following details the design review program adopted by Z-Air Systems and its results.

Figure 2.2.8 shows the sequence of the design review process adopted by the design team. The entire process was initiated due to rejects and returns of the space heater systems by customers. The sequence of the process includes SH review, FMEA review, analysis of test results, performing a pilot run, reviewing design and quality, then finalizing the product run.

All the information conveyed by the marketing department regarding the complaints and reasons for rejection or return are documented and fed into the company's information system. These data are utilized in developing the SH tables. The SH tables not only allow the designers to check the product for safety but also provided inputs regarding consumer complaints, quality issues, and the like. The soft analysis provides an in-depth view of the possibilities for careless misuse of the space heater beyond its normal wear and tear. The hard analysis provides the various potentials for function loss due to malfunction or deterioration over each component's estimated service life. These results are applied to improve products' design for safety aspects.

The team observes no overlaps in the S and H factors in the review. This means no possibility of any user injury from the use of the wrong part at the wrong time. Tables

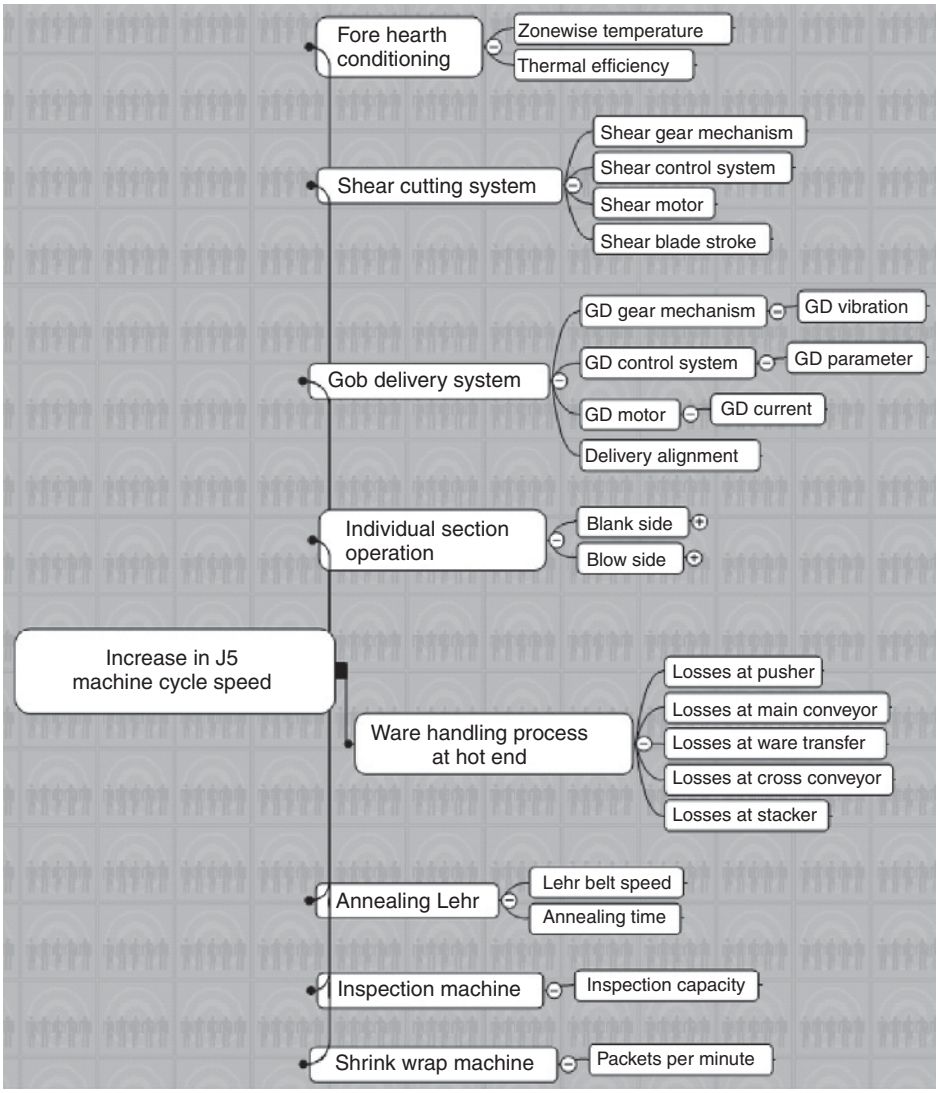


Figure 2.2.11 Tree diagram: Factors affecting cycle time, one-level.

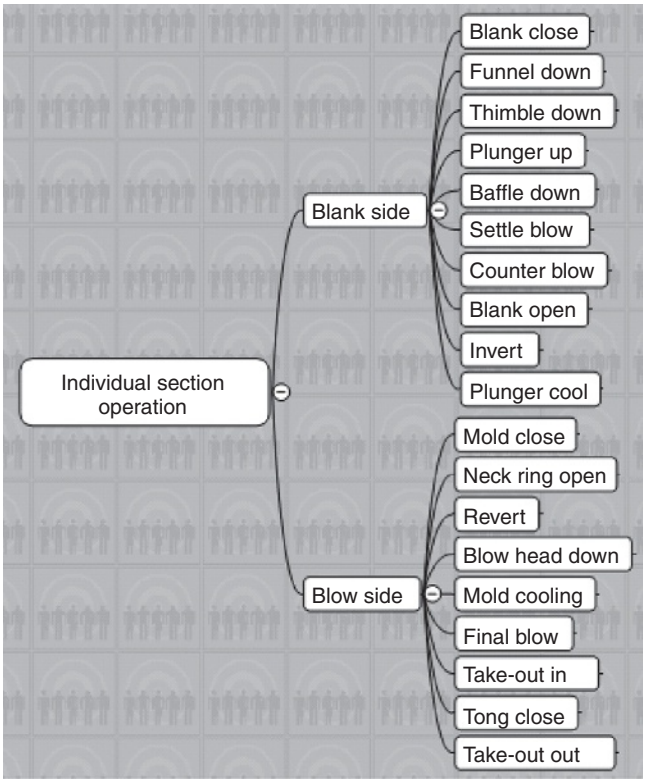


Figure 2.2.12 Tree diagram: Factors affecting cycle time, two-level.

2.2.1 and 2.2.2 show a sample of the team's SH analysis table.

The next stage, after performing the SH review, is to examine the performance of the new design. The FMEA method uses the RPN index, which is used to set priorities on the design and safety issues. Figure 2.2.9 shows a sample design FMEA for the space heater.

The implementation of the design review method results in a sharp drop in the defect rate. Market data indicate that the new design, after the review sequence, results in increased sales and revenues. Further, the rejection and return rate is reduced by over 77% compared to the older model. This measure of the sales and market performance of the product is an indication of the success of the design review process.

Observing the tremendous improvements in design and performance of the product, the management at Z-Air System implemented a plantwide requirement to apply the design review process to all products. This change in management policy has yet to be quantitatively documented at the plant, but if the past results are any indication, the company is headed toward a brighter future and could achieve a level of quality that would render its products superior to its competitors.

2.2.3.2 A six sigma case study

Classic Plastics Inc. is a leading bottling plant that supplies plastic bottles to all leading pharmaceutical companies. Because of both quality and productivity concerns, ABC needs to analyze its operations and improve its output in meeting the increased market demands. The current manufacturing plant has 35 bottling machines. Cycle time was identified as the key element. The entire process is shown in Figure 2.2.10.

The engineering team decided to analyze the process, using the DFSS strategy to improve it. After outlining the schedule for the project, the team held brainstorming sessions to determine the factors affecting the process. On close analysis, it arrived at the tree diagram shown in Figures 2.2.11 (one-level tree diagram) and 2.2.12 (two-level tree diagram).

On carefully planning and conducting the experimentation, the following actions were taken:

- Cycle time was changed from 15 to 15.5 ms.
- Gob length increased by 1.5 mm.
- All section differentials increased by two degrees.
- Cooling blower pressure increased from 700 to 740 mm WC.
- Pusher retract time decreased by three degrees.
- Number of bottles per stack row decreased from 27 to 25.
- Stacker forward length increased from 320 to 330 mm.
- Stacker forward speed increased from 50% to 60%.
- Lehr belt speed increased from 420 to 450 mm/minute.

The results of the changed cycle time are shown in Table 2.2.3. The engineers brainstorming and identifying potential problems in the process reported the following:

1. Bottle falling problem can be reduced by increasing the bearing surface.
2. Proposed action: Bar knurling instead of crescent knurling.
3. Hot blank can lead to blank seam and loading problem.
4. Proposed action 1: Cooling fins in blank to be provided.
5. Hot neck ring can lead to neck ring bulging.
6. Proposed action 1: Neck ring fins to be provided.
7. Proposed action 2: Number of balancing holes to be increased.

The changes, when implemented, resulted in an annual savings of $370,000. Further, it resulted in a horizontal plantwide deployment on all the 35 lines. The secondary changes in all the factors also resulted in improved aesthetics throughput improvements.

Table 2.2.3 Results documented for change in cycle times (current, proposed, and achieved)

Description	Actual			Proposed (15.5 cycles)			Achieved — G12 A 70 ml		
	On	Off	On time in mSec (15)	On	Off	On time in mSec (15.5)	On	Off	On time in mSec (15.5)
Blank side									
Blank Close	5	205	2222.2	4	202	2129.0	4	210	2215.0
Plunger Up in	32	92	666.7	32	92	645.2	30	92	666.7
Plunger Up 2	32	92	666.7	32	92	645.2	30	92	666.7
Plunger Up 3	32	92	666.7	32	92	645.2	30	92	666.7
Funnel Down	4	94	1000.0	3	95	989.2	6	94	946.2
1st Baffle Down	42	84	466.7	42	85	462.4	42	84	451.6
Settle Blow	42	84	466.7	42	85	462.4	48	84	387.1
Plunger Down in	100	100	0.0	100	100	0.0	100	100	0.0
Plunger Down 2	100	100	0.0	100	100	0.0	100	100	0.0
Plunger Down 3	100	100	0.0	100	100	0.0	100	100	0.0
2nd Baffle Down	118	202	933.3	118	202	903.2	118	210	989.2
Counter Blow in	147	200	588.9	145	198	569.9	148	207	634.4
Counter Blow 2	147	200	588.9	145	198	569.9	148	207	634.4
Counter Blow 3	147	200	588.9	145	198	569.9	148	207	634.4
Thimble Down in	206	30	2044.4	204	31	2010.7	210	28	1914.0
Thimble Down 2	206	30	2044.4	204	31	2010.7	210	28	1914.0
Thimble Down 3	206	30	2044.4	204	31	2010.7	210	28	1914.0
Blank Open	205	345	1555.6	200	347	1580.6	210	345	1451.6
Invert	256	312	622.2	256	312	602.1	259	312	569.9
Plunger Cool 2	347	0	144.4	347	0	3731.2	320	10	537.6
Plunger Cool 3	347	0	144.4	347	0	3731.2	320	10	537.6
Blow side									
Mould Close	280	191	3011.1	280	188	2881.7	280	188	2881.7
Neck Ring Open	315	328	144.4	315	328	139.8	318	335	182.8
Revert	323	215	2800.0	323	215	2709.7	328	220	2709.7
Blow Head Down	335	202	2522.2	335	198	2397.8	335	198	2397.8
Final Blow	60	188	1422.2	60	186	1354.8	40	186	1569.9
Take-Out In	228	260	355.6	228	260	344.1	228	260	344.1
Take-Out Out	260	135	2611.1	260	135	2526.9	260	140	2580.6
Tong Closed	254	110	2400.0	254	110	2322.6	256	130	2516.1
Mould Cooling	15	160	1611.1	15	187	1849.5	0	186	2000.0

References

Akao, Y., ed. *Quality Function Deployment*. Cambridge, MA: Productivity Press, 1990.

American Society for Quality Control. ASQC/Gallup Survey (1991). In Q — *Official Newsletter of the American Society for Quality Control* 6, no. 9 (November 1991).

Banks, J. *Principles of Quality Control*. New York: John Wiley, 1989.

Belliveau, P., Griffin, A., and Somermeyer, S. *The PDMA ToolBook 1 for New Product Development*. New York: John Wiley & Sons, 2002.

Chrysler Corporation, Ford Motor Company and General Motors, 1995, Quality System Requirements QS-9000 2nd Edition, AIAG (810) 358-3003.

Duncan, A. J. *Quality Control and Industrial Statistics*, 5th ed. Homewood, IL: Irwin, 1986.

Garvin, D. A. *Managing Quality: The Strategic and Competitive Edge*. New York: Free Press, 1988.

Grant, E. L., and E. S. Leavenworth. *Statistical Quality Control*, 6th ed. New York: McGraw-Hill, 1988.

Hill, T. *Manufacturing Strategy*. Homewood, IL: Irwin, 1989.

Juran, J. M. *Juran on Planning for Quality*. New York: The Free Press, 1988.

Kolarik, W. J. *Creating Quality: Process Design for Results*. New York: McGraw-Hill, 1999.

Montgomery, D. C. *Introduction to Statistical Quality Control*, 2nd ed. New York: John Wiley, 1991.

Radford, G. S. *The Control of Quality in Manufacturing*. New York: Ronald Press, 1992.

Shingo, S. *Zero Quality Control: Source Control and the Pokayoke System*. Cambridge, MA: Productivity Press, 1986.

Taguchi, G. *Introduction to Quality Engineering: Designing Quality into Products and Processes*. White Plains, NY: Kraus International, UNIPUB (Asian Productivity Organization), 1986.

Taguchi, G., E. A. Elsayed, and T. Hsiang. *Quality Engineering in Production Systems*. New York: McGraw-Hill, 1989.

Chapter 2.3

2.3

Solving the requirements problem

Tom Gilb

2.3.1 Introduction

2.3.1.1 To design and to engineer

The basic design process is finding 'means' for 'ends': it is finding designs that match the requirements.

What is the difference between design and design engineering? They are both essentially the same generic, and basic, process of 'finding satisfactory designs.' However, engineering disciplines are characterized, in my opinion, by the following distinctive traits:

- *quantification* of variable ideas (not 'high', but '42')
- concern for *all* necessary factors (all stakeholders, all requirements and all known design options)
- concern for more than mere 'satisfaction'; concern for *competitive* optimization – 'being the best', rather than just 'getting along'
- rational and systematic argument (for example, the use of Impact Estimation tables to discuss or present design quantitatively with respect to facts, not 'less formal' design or 'emotional' design.

Design asks, "Is this a good design?"

Design Engineering asks, "What are the totality of performance and cost attributes expected from this design *in relation to the multiple, quantified, performance and cost* requirements? *What are the risks, priorities, uncertainties, issues, relationships, dependencies and long-term lifecycle considerations, that we should responsibly consider about this design?"*

2.3.1.2 Requirement specification, design engineering, and evo are all iterative processes

Design ideas emerge, and are refined, throughout the lifetime of a system. Iteration is necessary in order to improve both the design ideas and the related requirements. Requirements and design ideas *cannot* be determined well in one single pass. Feedback from initial *design engineering* processes is necessary to get a *realistic* idea of which *design ideas are possible*, to determine how much design ideas *might cost* and to identify which tradeoffs amongst performance levels *might have to* be made. Until *the design and the requirements* are adjusted to this 'balanced level' with regard to reality, it is not possible to 'finalize' a competitive *design* for implementation.

In addition, after implementation starts, as a result of the measurable *feedback* obtained from the delivery of each of the Evo steps, even further refinement has to be considered for the design ideas, the requirements, and the *implementation plans*.

Requirement Specification, Design Engineering, and Evo are intimately linked, and some iteration linking them is necessary to get the best competitive results.

2.3.1.3 Requirements dictate and constrain the design, but *detailed* requirement specification can wait

What stakeholders perceive as 'value' drives us to state 'what stakeholders want' and 'how much stakeholders

Competitive Engineering; ISBN: 9780750665070
Copyright © 2005 Elsevier Ltd. All rights of reproduction, in any form, reserved.

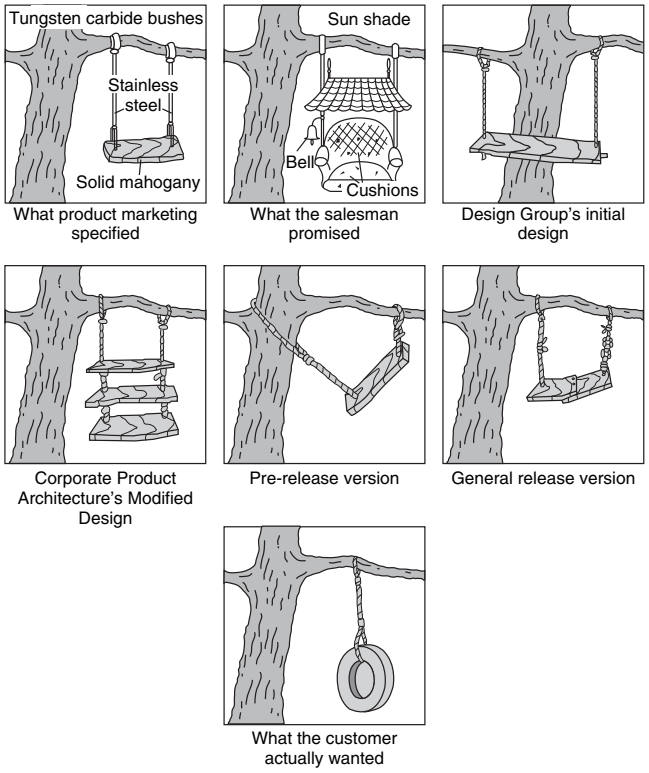


Fig. 2.3.1 The swing solutions. Source: Anon.

might be willing to pay for such change': in other words, to state the *requirements*. Requirements, which reflect values, give us a sound basis for evaluating a design idea: a basis for deciding if we might get what we will find of 'value' from a potential design idea.

> *'Interesting' results are our 'values.'*
>
> Keeney (1992)

We have to have at least a preliminary set of requirements, before we are ready to '*design.*'[1] These requirements could, even for a large project, be as simple as a statement of the handful of most critical requirements. (After all, these critical requirements are in fact usually driving the investment and the project!) The more detailed requirements can be derived gradually, as needed, during the Evolutionary Project Management (Evo) process. There is no need to try to get all the detail immediately. In fact, there is some virtue in letting the detail emerge as a function of experience and of interaction with key stakeholders.

Note: The detail is, however, ultimately important, and must be eventually specified, so we can fully understand the meaning, intent, risks, assumptions, and dependencies of all the requirements. For example, we need to understand which requirements are targeted only at specific system components.

2.3.1.4 Any design idea[2] can be considered

Any design idea that potentially contributes to the solution of the requirements, can be *suggested*. It is a question of how much a design idea contributes toward meeting the requirements, and at what costs, which determines whether a design is finally selected and implemented. It is then the design idea's *real* performance, on delivery, that will determine whether it survives, or must be replaced by another design idea.

EXAMPLE

Some Design Ideas:

- using process improvement teams
- allowing the project team an extra day's time off if a deadline is successfully met (motivation)
- buying an extra server (buy hardware)
- giving discounts to customers who field trial new products (monetary motivation)
- buying a standard component (buy hardware with known characteristics)
- contracting for a special tailored component (subcontracting and tailoring)
- building our own software component (development in-house)
- improving testing process (improving a specific development process).

This example shows a wide variation of types of design.

2.3.1.5 Design ideas can be identified during requirement specification

Even while you are initially specifying *requirements*, you should, if you feel that design ideas are flowing into your mind or the minds of colleagues, develop two separate lists of design ideas: *potential design ideas* and *design constraints*. The potential design ideas can then be kept aside for serious consideration in the *design* phase.

[1] In this chapter generally, I use the term 'to design,' but with regard Planguage processes, I actually mean 'to design *engineer*,' that is, to use rational and quantitative approaches.

[2] The term 'design idea' is used in this chapter. Solution, idea, strategy, design, means, idea and design solution are all synonyms amongst many other synonyms for 'design idea.'

2.3.1.5.1 Potential design ideas

These are either design ideas that were first assumed to be *requirements*, but then were recognized as *really* being *optional* designs, or they are simply design ideas that surfaced during requirement specification. Sometimes such design ideas are deliberately 'brainstormed' (for example, if experts in a specific area are available only during the initial requirements' gathering, then capturing their design ideas might be opportune). Here is an example of a way of keeping track of any potential design ideas; there is *no commitment to* implementing them at *this* stage.

EXAMPLE

Availability:

Type: Quality Requirement.

Scale: % Uptime.

Goal [USA, Version 1.0]: 99.90% <- Marketing Plan [April 20, This Year].

Design A [Availability = 99.90%]: Design Idea: Reuse of <high MTBF> Components <- Ed's suggestion.

Stretch [Worldwide, Version 3.0 and on]: 99.998% <- CEO Vision, "World Class."

Design B [Availability = 99.998%]: Design Idea: Triple Redundant Distinct Software <- Mike.

The two design idea specifications are local to the two different target specifications. They are not design constraints. They are clearly suggestions that need to be evaluated like any other suggestion.

Allowing systems engineers to note design ideas at an early stage is useful in several ways:

- it keeps track of potentially valuable design ideas which otherwise might get forgotten
- it helps make the distinction between the requirements and the design technology clearer ('clear ends–means separation')
- it lends credibility to the proposed goal levels (there exists some credible technology for the goal level suggested)
- it avoids the 'frustration' that some systems engineers feel when they are not allowed to be specific about the technology they have in mind
- it allows us to send a message that we have noted a systems engineer's suggestion or 'pet idea' and credited them with it – without yet officially approving it.

Some of the early design ideas may be *politically* wise to consider, due to the fact that influential stakeholders have suggested them. There is no risk of any unfairness in considering these design ideas, because they will have to compete with the later design ideas. All design ideas must win their place for implementation by being the best, in terms of numeric satisfaction of the requirements.

2.3.1.5.2 Design constraints

These are design ideas within the requirement specification, which have to be implemented at some stage. They can either specify or veto the use of specific designs. Usually, specific qualifying conditions apply.

EXAMPLE

Project Interface [Product Line = New Generation, European Market]:

Type: Design Constraint.

Description: The full Project Interface shall be implemented using the most <current version> available. It shall be updated whenever <newer versions> are available.

Rationale: Project Consortium Agreements.

This design constraint (a requirement) applies only to the Product Line of New Generation within the European Market.

2.3.1.6 The need for alternative design ideas

2.3.1.6.1 Choosing the best from the alternatives

When searching to find design ideas, it is important to look for alternative design ideas. Each individual design idea will produce different effects on a system's scalar attributes: the resource usage and performance levels. It is a question of selecting the design idea which has the best performance to cost ratio or the 'best fit to the requirements' with regard to 'delivering stakeholder value' compared to 'resources used' (value to cost ratio).

2.3.1.6.2 Choosing the best combined set from all the alternatives

Design ideas put together in different combinations will interact with each other in different ways: there could be negative side effects and/or positive 'combining' effects (synergy). By having several alternatives, it is possible to select the combination of design ideas, which has the best, estimated impact on the requirements. (Of course, the chosen combination can always be altered over time, in the light of feedback from evolutionary delivery.)

2.3.1.6.3 Reducing risk by use of alternative design

Another main reason for having alternatives is to reduce risk. If there are several candidate design ideas, then if the first choice fails there is always a backup. At an

extreme, alternative design ideas may be implemented in parallel to ensure that specific requirements are fully met.

2.3.1.7 Design optimization

When you are designing, you need to decide what type of optimization strategy you intend to use. The strategy options for Design Optimization Tradeoff include:

- **Cost Minimization:** When performance targets are met by specified designs, we can choose to continue to find alternative designs, that are at least equally well performing, with a view to reducing costs to the cost targets (if not below them!).
- **Design to Cost:** Another approach would be to design *to fully use* the all budgeted resources and to look for the designs that give maximum impact on the performance targets. In other words, the *most value* for a specific amount of limited resources. This is called 'Design to Cost.' '*Cutting your coat to suit your cloth.*'
- **Design to Performance Targets within Cost:** Another option would be to design to meet all the performance targets within cost, but to stop the process once all the planned performance levels were met. In other words, do not use additional time to reduce resource utilization further. This could be a possible approach when Time to Market is the most critical resource.
- **Design for Risk:** Another optimization concept would be to design with regard to risk. The *most pessimistic* estimates of performance impact, and of costs, would be used to determine the 'best design'. There are many other devices in Planguage that help us consider risk when designing (*see specifically, Impact Estimation*).

There are more combinations than those mentioned above. But you can see some basic choices. It is important in any project that you recognize how you are approaching the design optimization process, and that you communicate with your management about it. It could be there is some misunderstanding – maybe there *is* more financial budget available from them, as long as you show a track record of successful delivery.

If there are specific resource budgets that are critical to you, such as 'Time to Market', we recommend that you *initially* 'Design to Cost' with respect to 'calendar time' for delivery to market. Generally, you will want to design to meet the most critical constraints first, then see if you can maximize delivery of performance attributes and minimize other cost aspects in a second round of design effort.

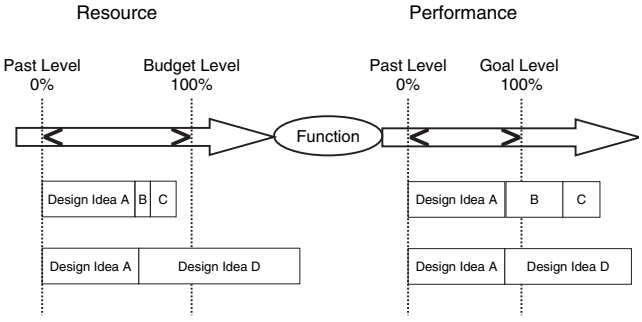


Fig. 2.3.2 To 'design' is to find design ideas, like A, B, and C, which will contribute toward planned performance and resource levels, while simultaneously respecting all constraints. Design idea D is 'good,' but costs too much.

2.3.1.8 Brief recap of Planguage methods and the design engineering process

With regards the Design Engineering Process:

- **Requirement Specification** supports the design engineering process by capturing the requirements. The requirements include specific information required for design decision-making. (For example, see Section 2.3.7.1 for further discussion.)
- **Impact Estimation (IE)** is part of the Design Engineering Process. It is the Planguage method used to *evaluate and choose* design ideas. It also incorporates risk evaluation. In addition, IE can also be used to monitor the actual progress toward meeting the requirements. The actual step measurements, obtained after each Evo step has been completed (with delivery of one or more design ideas), can be input into an IE table and compared against the original estimates. This feedback is used by the design engineering process, to understand where the gaps in design actually exist (that is, the gaps, which require additional design).
- **Evolutionary Project Management (Evo)** is used to *actually deliver* the design ideas. Evo handles risk by several means:
  - implementing design, step by step
  - demanding that we choose the design ideas *most likely* to provide *high benefit* (highest value to cost ratios, highest performance to cost ratios) for early delivery (design ideas are 'sequenced' by some chosen evaluation of their potential benefits and costs into an Evo step plan)
  - testing the reality of the design ideas 'in the field'
  - providing and using feedback data after each step. We can then realistically understand the accuracy of our estimates, concerning design ideas, and can

take appropriate measures, depending on the level of risk we perceive
- we have incrementally 'banked' *some* results and eliminated some risk, which maybe means we can afford to discuss taking some higher risk steps.

Note: Both the requirement specification process and the design engineering process are incorporated into Evo; each result cycle demands re-evaluation of the design and brief re-evaluation of the requirements (possible adjustments and tradeoffs). As stated earlier in this section, there is continuous iteration amongst these processes.

2.3.2 Practical example: beginning the design engineering process

Let us say we have specified the following requirements for a project 'Staging a Conference':

Staging a Conference: Type = Function.

========== Conference Performance Requirements ==========
Participation: Quality Requirement:
Scale: Percentage of Worldwide Membership participating.
Goal: 10%.
Representation:
Scale: Percentage of Worldwide Membership represented within defined <groups>.
Goal [Age under 25 or equating to <Student Status>]: 10%.
Information:
Scale: Percentage of Talks rated as 'good' or better (5+ on feedback sheet scale).
Goal: 50%.
Conviction:
Scale: Percentage of Participants wanting to return Next Conference.
Goal: 80%.
Influence:
Scale: Percentage of Participants who <improve as result of the Conference>.
Past: 90%.
Goal: 95%.
Fun:
Scale: Percentage of Participants rating the Conference City quality as 'good' or better (5+ on Feedback Sheet scale).
Past: 45%.
Goal: 60%.

========== Conference Budget Requirements ==========
Financial Cost: Resource Requirement [Financial]:
Scale: Average Participant Conference Cost for an individual Participant including Travel Costs.
Fail: Less than $2000.
Budget: Less than $1200.

A set of requirements for a conference, mostly performance requirements and one budget.

Now we can, driven by these relatively clear requirements, start designing.

We begin by listing any design constraints – the 'given' *design* ideas. Here there are none, the only 'given' is the main function that we are to stage a conference.

We can then list 'at least one potential design idea, for *each* of the requirements.' This is an arbitrary way of covering the requirements with 'some' design.

Design ideas

Central: Choose a location in the membership center of gravity (New York?).

Youth: Suggest and support local campaigns to finance 'sending' a young representative to conference.

Facts: Review all submitted papers on <content>.

London: Announce that the conference is to be in London the next year.

Diploma: Give diplomas for attendance, and additional diplomas for individual tutorial courses.

Table 2.3.1 Impact estimation table

Design ideas requirements	Central	Youth	Facts	London	Diploma	Events	Discounts	Sum for requirement
Performance Requirements Participation	80%	60%	0%	0%	30%	20%	30%	220%
	±50%	±70%	±50%	±50%	±50%	±50%	±50%	±370%
Representation	80%	80%	10%	0%	10%	20%	50%	250%
	±50%	±50%	±50%	±50%	±50%	±50%	±40%	±340%
Information	0%	20%	80%	0%	20%	0%	0%	120%
	±50%	±40%	±50%	±20%	±50%	±50%	±30%	±290%
Conviction	0%	20%	60%	80%	10%	80%	0%	250%
	±10%	±50%	±30%	±50%	±50%	±50%	±50%	±290%
Influence	0%	40%	60%	0%	80%	80%	0%	260%
	±50%	±40%	±50%	±50%	±50%	±5%	±50%	±340%
Fun	50%	40%	10%	0%	0%	80%	0%	180%
	±50%	±50%	±50%	±0%	±0%	±50%	±0%	±200%
Sum of performance	210%	260%	220%	80%	150%	280%	80%	
	±260%	±300%	±280%	±220%	±250%	±300%	±220%	
Resource requirements Financial cost	20%	1%	1%	1%	1%	30%	30%	111%
	±30%	±1%	±1%	±1%	±5%	±50%	±50%	±135%
Performance to cost ratio	210/20	260/1	220/1	80/1	150/1	280/30	80/30	

Events: Have entertainment activities organized every evening, such as river tours.

Discounts: Get discounts on airfare and hotels.

Now, are these design ideas going to make the conference what we want it to be, as defined by the target levels? Nobody really knows and nobody can say. Why not? Well, it depends on the interpretation of the design ideas and their *execution* in practice. Can we influence that? Yes. By specifying a more detailed design specification with precise details of what we are going to do, and exactly how it is to be done in practice (the implementation and operational design detail). In other words, we now have that first 'sketch of the building,' but we need to get down to the detailed 'blueprints' (engineering) needed by the 'bricklayers and carpenters.'

The first step is to assess what we can evaluate about the impact of our proposed design ideas on the requirements (perhaps a little exaggerated to make our point).

The Impact Estimation table is a way to 'see' what we are doing. A 100% estimate on this table is a belief (right or wrong, well founded or not) that we *will* reach the planned level on time. The plus/minus estimate is a *rough* notion of the uncertainty. Until we get better definition and justification, these numbers are of only slightly better value than words, such as good, bad, excellent. But they do give us *a systematic basis for improvement* in our planning.

(I ask the reader to be patient. All I am doing here is illustrating how one might define some requirements and design ideas, and then evaluate the impacts of each of the design ideas on all the requirements.)

The first observation I would make *here* is that we need to *redefine* the design ideas, *with more detail.* This is because of the high plus/minus uncertainties specified.

EXAMPLE

Central: Town must be cheaply accessible by most Participants. Location itself must offer reasonable priced Accommodation (like university dorms) within walking distance of the Conference Facilities. Easy access to shops, restaurants, entertainment. <Add even more, and give concrete suggestions>

The ideal would be to create a hierarchy of the components of the design idea, Central and evaluate each separately, to home in on exactly what aspects of the design idea gave most stakeholder value.

2.3.3 Language core: design idea specification

Design specification is not just writing down the bare outline of the 'design idea' itself. You have the option of including a large number of additional parameters to describe the design idea. Why bother? Well, it is a matter of how much you want to force yourself to think about your idea, how much you want to share in writing with others, and how much you want to control any risks involved with the design. You must have reasonable confidence that the design idea really will deliver the results you have estimated.

In the design specification, you should ensure that you:

- Supply more detail in the Definition parameter, as this will lead to better understanding of its specific performance and cost impacts. This can be done using structural breakdown (see the Definition parameter in 'Transport by Buses' example below). Each of the sub-design ideas can be refined, until you feel that you have enough detail in the design ideas to guarantee the result levels and result timings, which are planned across all the requirements (or you identify that you need additional design ideas).
- Clarify and limit the design ideas to the specific ones that you want. Avoid ambiguity so that other people can't misinterpret your design intent.
- Identify and specify designs that clearly deliver at least partly one required performance attribute. Any 'side-effect' impacts of each design, on the other requirements, must also be analyzed and estimated. Use the '->' Impacts parameter to explicitly declare which attributes you hope, or expect, will be impacted by specific sub-design ideas. ("Design Idea A -> Safety.")

Of course, you need to tailor your design specifications to suit the circumstances. A simple rule to guide you is 'to try the design specification parameters out at least once.' Too much description of a design idea will not hurt you, and can easily be deleted if it does not serve a useful purpose. Observe what the engineering team feels is worthwhile, and use that level of specification.

Here is an example of specifying a design idea. It was actually used in charity relief-organization work. (It shows how the main idea can be supported by sub-designs. The aim being to get better control over the results.)

EXAMPLE

Transport by Buses: Design Idea.

Description: Drive Refugees back across the border by bus.

Definition [Sub-designs]:

Village: Refugees should be selected from the same, or nearby, village -> Financial Cost.

White Paint: Buses should be painted UN white, and UN marked -> Safety <- Geneva Convention, Article 6.3.

Agreement: <Agreement with government> to allow transport and resettlement, without harassment, shall be made *before* crossing the border. Agreement papers will be onboard the bus -> Safety.

Radio: Buses shall have radio or mobile telephone contact with our headquarters *during* the transport -> Safety. "Maybe also video and tape recorder?"

Witness: UN employees, or relief agency employees, perhaps UN soldiers will accompany the buses -> Safety.

2.3.4 Rules: design specification

Tag: Rules.DS.
Version: October 7, 2004.
Owner: TG.
Status: Draft.

Note: Design specifications are either for optional design ideas (possible solutions) or required design constraints (that is, actual requirements AND consequently, pre-selected solutions).

Base: The rules for generic specification, Rules.GS apply. If the design idea is a design constraint (a requirement), the rules for requirement specification, Rules.RS also apply.

R1: **Design Separation:** Only design ideas that are intentionally 'constraints' (*Type: Design Constraint*) are specified in the *requirements*. Any other design ideas are specified separately (*Type: Design Idea*). *Note all the design ideas specified as requirements should be explicitly identified as 'Design Constraints.' (Repeat of Rules. RS.R9: Design Separation.)*

R2: **Detail:** A design specification should be specified in *enough detail* so that we know precisely what is expected, and do not, and cannot, inadvertently assume or include design elements, which are not actually intended. It should be 'foolproof.' *For complex designs, the detailed definition of its sub-designs can satisfy this need for clarity, the highest level design description does not need to hold all the detail.*

R3: **Explode:** Any design idea (Type: Complex Design Idea), whose impact on attributes can be better controlled by detailing it, should be broken down into a list of the tag names of its elementary and/or complex sub-design ideas. *Use the parameter 'Definition' for Sub-Designs.*

Example of a Design Specification

Tag: OPP Integration.
Type: Design Idea [Architectural].
========== Basic Information ==========
Version:
Status:
Quality Level:
Owner:
Expert:
Authority:
Source: System Specification Volume 1 Version 1.1, SIG, February 4 – Precise reference <to be supplied by Andy>.

Gist: The X-999 would integrate both 'Push Server' and 'Push Client' roles of the Object Push Profile (OPP).
Description: Defined X-999 software acts in accordance with the <specification> defined for both the Push Server and Push Client roles of the Object Push Profile (OPP).
Only when official certification is actually and correctly granted; has the {developer or supplier or any real integrator, whoever it really is doing the integration} completed their task correctly.
This includes correct proven interface to any other related modules specified in the specification.

Stakeholders: Phonebook, Scheduler, Testers, <Product Architect>, Product Planner, Software Engineers, User Interface Designer, Project Team Leader, Company engineers, Developers from other Company product departments which we interface with, the supplier of the TTT, CC. "Other than Owner and Expert. The people we are writing this particular requirement for."

========== Design Relationships ==========
Reuse of Other Design:
Reuse of This Design:
Design Constraints:
Sub-Designs:

========== Impacts Relationships ==========
Impacts [Functions]:
Impacts [Intended]: Interoperability.
Impacts [Side Effects]:
Impacts [Costs]:
Impacts [Other Designs]:
Interoperability: Defined As: Certified that this device can exchange information with any other device produced by this project.

========== Impact Estimation/Feedback ==========
Tag: Interoperability.
Scale:
Percentage Impact [Interoperability, Estimate]: <100% of Interoperability objective with other devices that support OPP on time is estimated to be the result>.
========== Priority and Risk Management ==========
Rationale:
Value:
Assumptions: There are some performance requirements within our certification process regarding probability of connection and transmission etc. that we do not remember <-TG.
Dependencies:
Risks:
We do not 'understand' fully (because we don't have information to hand here) our certification requirements, so we risk that our design will fail certification <-TG.
Priority:
Issues:
========== Implementation Control ==========
Not yet filled in.
========== Location of Specification ==========
Location of Master Specification: <Give the intranet web location of this master specification>.

Fig. 2.3.3 Here is a real (doctored!) example of a design specification using a version of the Design Specification Template given later in Section 2.3.7. Not all parameters are filled out yet. Notice that even the parameters which are not filled out (like Impacts [Side effects] and Issues) are asking important questions about the design – and hinting that responsible designers should answer such questions!

If you know it can be decomposed; but don't want to decompose it just now, at least explicitly indicate the potential of such a breakdown.

Use a Comment or Note parameter.

R4: **Dependencies:** Any known dependencies for successful implementation of a design idea need to be specified explicitly. Nothing should be assumed to be 'obvious.' *Use the parameter, Dependency (or Depends On), or other suitable notation such as [qualifiers].*

(For design constraints (requirements), this is a repeat of the rule, Rules. RS.R5: Dependencies.)

R5: **Impacts:** For each design idea, specify *at least one* main performance attribute impacted by it. *Use an impact arrow '->' or the Impacts parameter.*

Comment: At early stages of design specification, you are just establishing that the design idea has some relevance to meeting your requirements. Later, an IE table can be used to establish the performance to cost ratio and/or the value to cost ratio of each design idea.

EXAMPLE
Design Idea 1 -> Availability.
Design Tag 2: Design Idea.
Impacts: Performance X.

R6: **Side Effects:** Document in the design specification any side effects of the design idea (on defined requirements or other specified potential design ideas) that you expect or fear. *Do this using explicit parameters, such as Risks, Impacts [Side Effect], and Assumptions.*

Do not assume others will know, suspect, or bother to deal with risks, side effects and assumptions. Do it yourself. Understanding potential side effects is a sign of your system engineering competence and maturity. Don't be shy!

EXAMPLE
Design Idea 5: Have a <circus> -> Cost A.
Risk [Design Idea 5]: This might cost us more than justified.
Design Idea 6: Hold the conference in Acapulco.
Risk: Students might not be able to afford attendance at such a place?
Design Idea 7: Use Widget Model 2.3.
Assumption: Cost of purchasing quantities of 100 or more is 40% less due to discount. Impacts [Side Effects]: {Reliability, Usability}.

R7: **Background Information:** Capture the background information for any estimated or actual *impact* of a design idea on a performance/cost attribute. The evidence supporting the impact, the level of uncertainty (the error margins), the level of credibility of any information and the source(s) for all this information should be given as far as possible. For example, state a previous project's experience of using the design idea. *Use Evidence, Uncertainty, Credibility, and Source parameters.*

Comment: This helps 'ground' opinions on how the design ideas contribute to meeting the requirements. It is also preparation for filling out an IE table.

EXAMPLE
Design Tag 2 -> Performance X <- Source Y.

R8: **IE table:** The set of design ideas specified to meet a set of requirements should be validated at an early stage by using an Impact Estimation (IE) table.

Does the selected set of design ideas produce a good enough set of expected attributes, with respect to all requirements and any other proposed design ideas? Use an IE table as a working tool when specifying design ideas and also, when performing quality control or design reviews on design idea specifications.

Ask the 'Twelve Tough Questions' about the design ideas! (Can you quantify the impacts?).

R9: **Constraints**: No single design specification, or set of design specifications cumulatively, can violate any specified constraint. If there is *any* risk that this might occur, the system engineer will give a suitable warning signal. *Use the Risk or Issues parameters, for example.*

R10: **Rejected Designs**: A design idea may be declared 'rejected' for any number of reasons. It should be retained in the design documentation or database, with information showing that it was rejected, and also, why it was rejected and by whom.

EXAMPLE
Design Idea D: Design Idea.
Status: Rejected.
Rationale [Status]: Exceeds Operational Costs.
Authority: Mary Fine. Date [Status]: April 20, This Year.

2.3.5 Process description: the design engineering process

2.3.5.1 Process: Design engineering process

Tag: Process.DE.
Version: October 7, 2004.
Owner: TG.
Status: Draft.

Assumption: We have clearly-stated and reasonably complete requirements.

Notes:

1. *Design is an iterative process. The process given in this section should be viewed with this in mind; the procedure is written as if it were carried out in a single pass, but in practice, a much more complex pattern of cross checking, backtracking, and trade-offs would actually be carried out.*
2. *This procedure is much longer than it needs to be. You should probably use a more concise version (say, one statement for each procedure step).*

2.3.5.2 Entry conditions

E1: The Generic Entry Conditions apply. The requirement specification should ideally have exited from Specification Quality Control (SQC).

Overview of a Design Process

Design is an *intellectual* process, which is supported by problem definition, requirement specification, best-practice design process standards and analysis tools. The following fundamental questions arise in designing:

1. **Analyze the Requirements:** Which requirements are of high stakeholder value? What constraints apply? What is the priority time sequencing for delivery of requirements?
2. **Find and Specify Design Ideas:** How do we *find* and *specify* potential design solutions for our requirements?
3. **Evaluate Design Ideas:** How do we *evaluate* potential design solutions?
4. **Select Design Ideas and Produce Evo Plan:** How do we *choose* from several 'good' design alternatives? What do we do about *uncertainties*, and about the *risk* that the selected designs are not as good as we thought?

Notes:

Ansoff points out that "Simon has shown that solution of any decision problem in business, science, or art can be viewed in four steps:

1. *PERCEPTION of decision need or opportunity. Simon calls this the INTELLIGENCE phase.*
2. *FORMULATION of alternative courses of action.*
3. *EVALUATION of the alternatives for their respective contributions.*
4. *CHOICE of one or more alternatives for implementation." (Ansoff 1965)*

This supports the choice of the four main sub-processes of the Design Process!

Fig. 2.3.4 Overview of a Design Process. This applies with or without a quantified engineering approach to design.

E2: Any existing feedback, from Impact Estimation (design idea analysis), or practical trials, is made available to the design engineer.

2.3.5.3 Procedure

P1: Analyze the Requirements: You may well identify stakeholder conflicts and overlaps[3] amongst the requirements while analyzing them. These need conflict resolution: consider if the 'tougher' requirement level can be used, identify the 'owning' stakeholders for all values and negotiate with the stakeholders. It may well be worth waiting until you have some alternative design ideas before you negotiate with the stakeholders, as by then you will have a better understanding of what solutions can be delivered.

P1.1: **Establish the Stakeholder Value on Delivery of Each Requirement:** The value on delivery of each requirement to the system/ organization should be assessed. You are looking for the requirement areas where there are major benefits. (What is of value depends on the stakeholders: it might not be just financial resource.) Identify the volume of use associated with each requirement.

Ideally, the stakeholders will have already selected the highest value requirements as the *critical requirements*.

For example: resource savings for performance requirements will be relatively simple to determine. Say you wanted to bring the time of carrying out a transaction down from two minutes to one minute. The benefit to the business, assuming 200 such transactions were carried out a day, would be 200 minutes per day. If operators had to wait while each transaction went through – this could amount to freeing up over three work-hours each day to carry out additional activities. In other words, assuming 250 work-days each year, 250 multiplied by 200 minutes each year. To the business, the ability to free up staff or the cost of employing the staff in this area, is the 'value' gained on delivering the requirement.

P1.2: **Sequence the Delivery Order of the Requirements:** Sequence the requirements for attention in the order of maximum stakeholder value first. Adjust to cope with any dependencies amongst the implementation of requirements (These dependencies either prevent implementation or prevent some level of benefit being achieved).

Note also the possibility for delivering each requirement in stages either by gradually improving a performance level or cost level or by delivering into different areas at different times (that is, to divide according to qualifier conditions; for example, by geographic area, by role, and/or by timescale).

P1.3: **Establish Scope of Design Interest:** Establish and specify the scope of interest for the system design.

[3] Overlaps in requirements represent either an opportunity for additional value to be delivered or, the possibility for over-estimation of value *('double accounting')*.

This will be a set of qualifying conditions covering a specific timeframe and specific system space (locations/components/functionality). For larger systems, you might want to divide the system into major subsets – maybe, say, by functionality and/or by timescale, so that you can work on a series of smaller design areas.

P1.4: **Make a List of Requirements within the Scope defined:**

Identify any function, performance, resource or condition *constraints* specified in the requirements. Then identify the *target* requirements. Note the qualifying conditions, which apply to each one of them.

P2: **Find and Specify Design Ideas**:

P2.1: **Exploit any Earlier Notes of Design Ideas:** Check to see if there is already a list of potential design ideas developed at the same time as the requirement specification. If there is, include those design ideas for consideration.

P2.2: **Establish the Design Constraints:** Read the requirements to see if there are any design constraints specified (Type: Design Constraint). If there are, then note them and any specific conditions qualifying them [time, place, event].

P2.3: **Brainstorm Design Ideas:** Search all available sources of design information for good matches to our stated requirements. (Specifically with regard to meeting the function requirements, achieving the performance levels and, delivering within the budgets. Any constraints and conditions must also be considered.)

Identify any dependencies amongst design ideas.

Also identify any design ideas that are alternatives.

Attention needs to be focused especially on the areas of greatest benefit to the system/organization. It is *the gaps* between our current updated system design process benchmarks (how well we have satisfied requirements until now) and our specified targets, which are of interest.

We are totally dependent in our search on the following:

- Knowledge of the *existence* of good design ideas (where are they? do we have the best ones?).
- Having complete and reliable *information about the likely impacts* of the design ideas on system attributes, so we can match the best design ideas to our residual requirement gaps. Most design ideas have too little specified, or available, data about their performance and cost characteristics.
- Understanding how design ideas *mix and interact* with each other. (Maybe the mixture will conflict? A design idea, in itself, might seem satisfactory, but the effect of combination with other design ideas, already in place, or under consideration, could be a counter productive. Alcohol and driving don't mix well; though each in the right time and place might be acceptable.)

Hint: Select design ideas from available knowledge: books, periodicals, conference proceedings, past products, memory, colleagues, web searches, company experience, competitive analysis, benchmarking, and others.

Stop the search when a set of satisfactory design ideas has been found, or when you run out of time to search for more.

P2.4: **Draft Design Ideas:** Draft a set of the design ideas, which might satisfy the requirements. See the design specification template outlined in Section 2.3.9.

P2.5: **Collect Specification Detail (to support later Impact Estimation):** Add detail to the design idea specifications in order that there is sufficient information to enable estimates of the impact that each design idea will have on each performance and cost attribute. Refine the design idea specifications to the levels of detail, which reflect the 'level of uncertainty' and 'risk of deviation' from planned levels, which *you* are prepared to accept. Ensure all suspected risks, assumptions, and uncertainties are documented.

Document clearly, where any design idea has weaknesses with regard to the requirements. (For example, "Risk: Too long an implementation time." and "Risk: High risk of user dissatisfaction over usability.")

P2.6: **Consider Design Implementation:** Once you have defined the design ideas themselves, then turn *your attention to their implementation processes*. What qualifications are required for the implementers or subcontractors? What process should they follow? How will they be required to prove or measure their results? Leave nothing essential to the whims of others! Get control over your design ideas. *See 'Implementation Control' section in the design specification template in Section 2.3.9.*

P2.7: **Consider System Capacity and Growth:** Even when things work well in practice initially, there is no guarantee they will *continue* to do so. Success breeds volume. Volume breeds *capacity* problems. The 'good' system is no longer good enough. So we must be prepared to undertake a *continuous* responsibility for modifying the system design to meet *changed* circumstances. Sometimes 'gradual adjustment' is all that is necessary. Sometimes major new architecture is necessary. You need to be explicit about system capacity and give, if relevant, an outline of your plans of how to cater for system growth.

P3: **Evaluate Design Ideas:** We must evaluate the effects of a design on the system to which it will be added, and with regard to how it will mix with 'design changes yet to be implemented,' or even 'yet to be imagined,' by considering our long-range requirements, and architecture, for adaptability. We must ensure that we evaluate design ideas for their incremental effects on *all* of our required attributes – not just the requirements we initially designed them to primarily impact. Consider side effects – good and bad, intended and unintended.

P3.1: **Filter for Violation of any Constraint:** A design idea must not violate any applicable constraint(s). Check each design idea against each constraint. Mark the status of any design idea that violates, or potentially might violate, any constraint as 'Rejected' giving the reason in a 'Rationale' parameter. *(This is a more systematic check than might have been carried out when brainstorming during P2.3.)*

P3.2: **Estimate all Impacts:** Using an Impact Estimation (IE) table, estimate the impact of each design idea (or set of design ideas), on each performance target and each cost budget.

Cite evidence, plus/minus uncertainty and source(s) for each impact estimate. Determine the credibility of each estimate (using the credibility ratings scale from 0.0 to 1.0).

P3.3: **Consider Side Effects:** It is not just a case of checking that a design idea delivers the *required* benefits, we must also consider whether a design idea has any *unintended* negative side effects, which are *unacceptable (some negative effects may be tolerable overall).*

P3.4: **Consider Safety Margins:** You also need to assess whether the safety factors are met. Maybe a factor of 2 times 'over-design' is required?

If needed, return to P2 to look for further design ideas or, consider if the requirements need modifying.

P4: **Select Design Ideas and Produce Evo Plan:**

P4.1: **Initial Sequencing of Design Ideas:** We are then faced with decisions about which design ideas to select *for implementation* and which to reject. We will often be faced with several 'sufficient' alternatives; any one of which would be adequate. So how do we choose? Usually, no one dimension (for example, 'cost') is decisive.

In general, the selection decision must be made based on the *many* dimensions of measurable performance and cost. Any conditions (such as those specified in qualifiers) also impact selection.

Selection means *prioritization.* We need to determine which requirements we intend to satisfy first. *(We have an initial selection of critical requirements from P1.)* Evolutionary project management (Evo) will have the final say in determining the actual implementation sequence, but during the design engineering process we must attempt an initial sequencing of the design ideas to meet the requirements' priorities.

Establish which design ideas impact each of the constraints.

Establish which design ideas impact each target requirement (function targets, performance targets, and budget targets). The impact estimation table will provide the information for the performance and budget targets. The design specifications will also hold some information depending on how much detail has been captured. *Remember the performance and budget attributes do not 'hang in mid-air'– they will be attached to some functionality.*

Next, identify any design idea dependencies.

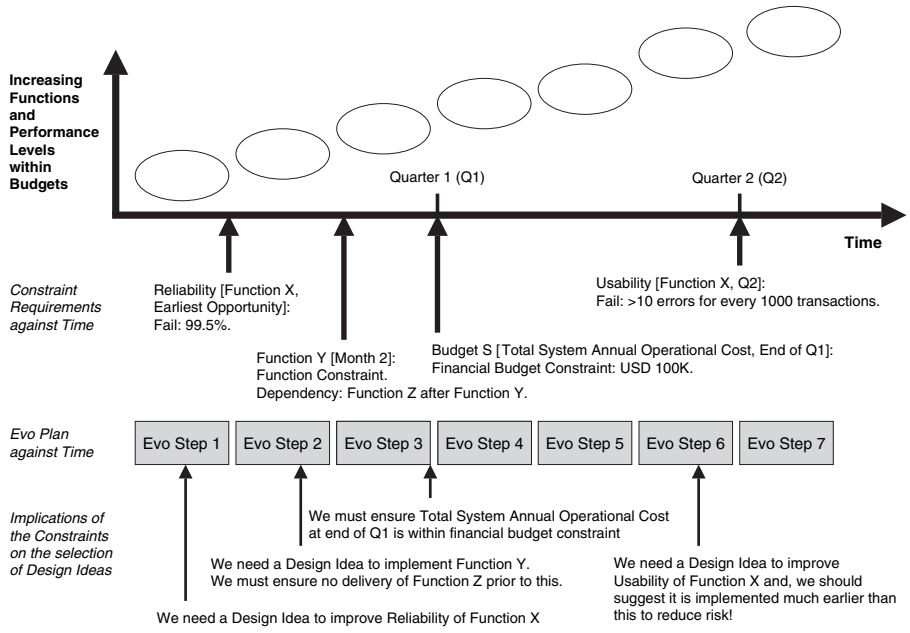


Fig. 2.3.5 Diagram shows the path for a system improving over time as Evo steps are delivered. The points marked on the time axis are the times when specified constraints have to be delivered. The sequenced Evo steps are attempting to deliver the requirements on time. The design ideas making up the content of the Evo steps have, as the first priority, to try to satisfy any constraints. Not shown in this diagram – the second priority is that they deliver the required target functions, and to the target levels for the performance and cost attributes.

At this stage, there should be a sequence of design ideas dictated by the (system scope) conditions – especially by the required timescales.

Where there are alternative design ideas, the performance to cost ratios from an IE table can be used to determine which designs contribute most efficiently toward meeting the requirements.

P4.2: **Ensure adequate Safety Margins to address Risks:** Ensure the sum of the impact estimates for each performance requirement covers the required safety margins for both performance and cost targets.

P4.3: **Consider Design Ideas with regard to the Risks:** We must also consider the uncertainties in our evaluations. The initial, purely intellectual, design process is inherently at risk of giving us false conclusions because *our* system is always somehow different from all *others*. Past design idea experience might not be valid. Our information on design effects could also be too general, or even downright wrong.

We can make allowances to cope with risk by 'over-design.' We may deliberately choose more design ideas than we strictly need in order to have safety margins. For example, we may choose two solutions that back each other up, rather than one.

We must also carefully *validate* our design choices in reality, and be prepared to *re-design* whenever practical experience shows this is necessary.

P4.4: **Consider Optimization:** Once we have an initial set of design ideas, which provide a satisfactory solution, then we can try to *optimize* using a declared optimization strategy (which might be a requirement of an engineering policy statement).

For example:

- Look for the least-cost set of design ideas, which *fully* meets the requirements.
- Select design ideas with the highest performance to cost ratios. This is *generally* good competitive practice.

But there are all kinds of variations on optimization strategies depending on your priorities regarding performance targets, resource budgets and risk (see 'Design optimization' within Section 2.3.1).

P4.5: **Re-define Design Definitions:** Re-define design ideas, if you can improve your impact estimates, and get better control over desired results by doing so. Re-define them so that they have substantially different performance and cost impacts, in the direction you need them to be.

P4.6: **Consider Pilot and/or Trial:** Plan to try out design ideas, which *seem* high risk, in pilots and/or trials, or specify them for implementation in early evolutionary result cycles. Feedback any results into *this* process (at P3.2: Estimate all Impacts.). Refine your estimates.

Design knowledge, from past uses of a design, gives us some idea of how we can expect things to work. But, new systems contain many new elements of technology, inputs and people. Thus, only practical use of a new design idea, in the real environment, will assure us that we were correct in our estimates of design effect or will convince us that we were to some degree wrong. So, the design engineering process must somehow be linked with a practical process of trying things out, well before large-scale irreversible commitment is made to design ideas.

2.3.5.4 Exit conditions

X1: The Generic Exit Conditions apply. The set of design specifications should have exited SQC with no more than one estimated remaining major defect/page. *(Expectation: If you don't demand such a low exit level, the specification will have 20 or more major defects/page.)*

X2: A safety factor of <*four*> is required for all performance and cost attributes.

(*Note: 4 times over-design, is NOT 4 times more cost!*)

Note: Process 'Exit' means we can *then* use the design specification for planning *trials to get feedback* (that is, using the design in Evo steps). It does not mean the design specification is a 'final' specification.

Gap Analysis by Igor H. Ansoff

The procedure within each step of the cascade is similar.

(1) *A set of objectives is established.*
(2) *The difference (the 'gap') between the current position of the firm and the objectives is estimated.*
(3) *One or more courses of action (strategy) is proposed.*
(4) *These are tested for their 'gap-reducing properties.'*

A course is accepted if it substantially closes the gaps; if it does not, new alternatives are tried.

IgorH. Ansoff, *Corporate Strategy* (Ansoff 1965 Pages 25-26).

Also quoted in Mintzberg (1994).

2.3.6 Principles: the design engineering process

1. **The Principle of ' Design Ideas are only as Good as the Requirements Satisfied '**
   Design ideas cannot be correctly judged or validated *except* with respect to *all* the performance and cost requirements they must satisfy.
2. **The Principle of 'The Best Chess Move'**
   You should try with each increment of design specification or design implementation, to get the best possible satisfaction of your unsatisfied performance requirements, from your unused cost budgets.

3. **The Principle of 'Results Beat Theory'**
   Design ideas are only as good as their *real results*, not their *intent*.
4. **The Principle of 'Early Surprises'**
   You never *know* how it works, until you have actually tried out a design idea in practice. Get surprised as early as possible!
5. **The Principle of 'It's Not Just What You Do, It's How You Do It'**
   Design ideas must try to exercise control over both design *content* and design *implementation*. The devil is in the details!
6. **The Principle of 'Good is Not Always Good Enough'**
   A 'good' design idea might not be good enough to meet all *your* targets on *time*.
7. **The Principle of 'Designs should have Good Return on their Investment'**
   'Good' design ideas might cost *too much*, sooner or later.
8. **The Principle of 'Sneaky Gremlins'**
   Apparently 'good' design ideas might have subtly hidden nasty side effects. Estimate them, know when you don't know them, measure them, and don't assume they won't hurt you! They will show *you* no sympathy!
9. **The Principle of 'Design Beats Test'**
   Design performance 'in', and design 'to control' costs: You cannot *test* quality *into* a badly designed system.
10. **The Principle of 'Eternal Vigilance for the Butterfly Effect'**
    You never *finally* know about a design idea's effects; tomorrow's slightest change might ruin your whole project. *Even initially successful designs might have to be adjusted for growth and change*.

2.3.7 Additional ideas

2.3.7.1 Priority determination

Systems engineering can be viewed as a constant stream of priority evaluations. So priority determination is a key concern. However, the conventional means of deciding priority are frequently inadequate: a subjective weighting approach is, unfortunately, often adopted. Ideally priority determination for implementing requirements should be:

- a 'performance to costs impact' and 'resource-focused' process. It should consider value to cost ratios, return on investment (ROI) and take into account resource availability.
- an information-based process, which makes full use of the available factual information, and is able to reuse this information. Not a weight-based process.
- a dynamic process, which uses feedback from the ongoing implementation; and is open to instigating, and catering for, change in requirements and in design ideas.

Ideally the ultimate values to the stakeholders, which are the results of the system performance characteristics, would be evaluated and used to determine priority. In practice it might be difficult for a systems engineer to access the stakeholder domain data needed to calculate the value that the stakeholder would expect to experience. Even the stakeholder might have difficulty estimating the 'value delivered' accurately. So, we might choose to fall back on a more immediate notion of stakeholder value – meeting the required targets.

2.3.7.2 What is wrong with the subjective weighting approach for determining priority?

In the priority weighting (or priority ranking) process, each element of a set of elements in a decision-making model, is subjectively assigned a numeric value indicating its priority (for example, a value on a scale of 1 – 10 or, a percentage weight).

The degree of subjectivity[4] is determined by such factors as the actual people asked (the number of people, their roles and their expertise) and how they arrive at their decisions (their decision *processes*; including such things as their influences). In many cases, people are asked on a one-off basis during a group meeting to assign numerous comparative weightings 'off-the-top-of-their-heads.' Inadequate documentation of who, when, and why (experience and/or fact) is widespread. The reasons why such a process is weak, when determining the priority for implementing requirements, include:

- Information overload: too many things have to be taken into account at once for subjective assessment to work well.
- Lack of specific information: often there are gaps in the information available: evidence and source data are usually missing.

[4] Note, I am objecting to subjective weightings, not to stakeholders proposing their own subjective requirements.

- 'One-off' weightings: weightings tend to be 'frozen', they are not reassessed frequently.
- Lack of consideration of resources: resources are simply not taken into account.
- An individual stakeholder's viewpoint is limited. (A person's subjective judgment depends on many things. For example, experience and access to information.)
- Typically people can only participate in supplying their requirements and committing their resources. They are unlikely to be able to make a globally optimal priority decision, on behalf of the entire stakeholder community.
- In a group meeting, factors such as authority, office politics and personality interfere with the outcome.

2.3.7.3 The role of resource in determining priority

Planguage defines priority as follows:

> *A 'priority' is the determination of a relative claim on limited resources. Priority is the relative right of a competing requirement to the budgeted resources.*[5]

If resources were unlimited, there would be no need to prioritize things. You could have it all.

Many approaches to priority oversimplify or even eliminate consideration of resources. Yet ROI is ultimately the key driver when deciding priority. What value shall be obtained in relation to the 'resources needed' (the costs)?

Resource availability can also be a factor in determining implementation priority. Selection of a priority solution might be:

- influenced by a lack of some resources
- affected by the ability to substitute one resource with another.

2.3.7.4 Planguage information supports priority determination

Planguage captures a wide range of reusable information that supports priority determination. It quantifies all scalar requirements and caters for individual deadlines at a detailed level, and as a result gives you a greater level of priority control. Some key Planguage specification parameters assisting priority determination are as follows:

- Value
- Stakeholder
- Constraint
- Target
- Dependencies
- Qualifiers [Time, Place, Event]
- Authority
- Source.

The source, authority, and stakeholder information establishes the stakeholders affected by a requirement, and their level of responsibility. When determining priority, meeting any constraints is the first priority. The next priority is to meet the targets. The qualifiers narrow the requirements down to the specific conditions: time qualifiers specify the timescales, place qualifiers limit the system space and event qualifiers state any specific circumstances, which apply.

Planguage might capture this information, but it requires evaluation to establish the priorities. There may be priority conflicts needing negotiation.

2.3.7.5 Priority strategy

One piece of information vital for priority determination is the strategy for priority. There are several different strategies that could be chosen. See discussion on design optimization within Section 2.3.1.

2.3.7.6 Dynamic priority evaluation

Planguage adopts a dynamic, numeric idea of priority. Priority is defined as the claim on resources to develop or operate a system. It is the currently unfulfilled requirements, (the gaps) which have priority. Our highest priorities, at any moment in time, are the unfulfilled requirements that are due next, date-wise.

There are no artificial weighting factors needed in Planguage. We use only direct natural statement of the qualities and costs we want, together with when we want them. We compare 'what we want' with 'what we have' at the moment. The larger the gap between 'wants' and accomplishments, the higher the current priority in that area to do more design work or to do more implementation work.

Using Evo, priority control becomes early, frequent and continuous, throughout the project design and implementation phases. Priorities change as they are satisfied (just as appetite changes as food satisfies it).

Also, the basic requirements can change at any moment of a project. It would be convenient if they didn't, but the real world is not that co-operative!

[5] I mean all types of resource including time to deadline, human effort, money, and space.

Continuous re-assessment of priority, allows any changes in the requirements to be incorporated into the system design process.

2.3.8 Further example/case study: design specifications masquerading as requirements

This is a sample of some real design ideas, which were *found in a requirement specification.* (Certain details are changed for confidentiality.)

They are extremely early outline drafts and still need a lot of work! We certainly had not yet enhanced the specifications to the level required by the Planguage template in this chapter. However, the drafts do give some practical insights into simple Planguage formatting. The most important steps we took were as follows:

- to refuse to treat them as requirements
- to identify the performance attributes they were *intended* to impact *(see 'Impacts' parameters below)* and
- to define the impacted performance attributes properly.

EXAMPLE

Adaptive Channel Allocation: ACA: Design:

Assumption [ACA]: New Product must automatically yield to Macro cellular system, and to re-tuning of the Macro radio network.

Impacts [Co-existence]: Slow or Fast? <- Marketing Specification 3.11.

Note: This is one design idea, not a constraint.

Automatic Roaming Designs: "A rough collection of design ideas." Impacts: Automatic Roaming.

Note: these may be design constraints! <- New Product Team 4 March.

IS-41 signaling link to the public network <- Marketing Specification 5.2.1.

Signaling, data and messaging interfaces <- Marketing Specification 5.2.

The New Product must support the protocol of Cellular Messaging Teleservice over its signaling link over both public and cellular network <- Marketing Specification 5.2.2.

The New Product must support the receipt and acknowledgment protocol for voice-message-waiting indication <- Marketing Specification 5.2.3.

Cell Plan Minimization: Impacts: {Installability, Maintainability}.

Cooling Fans [Radio Heads]: to be avoided to avoid noise, but quiet ones, as defined by Quietness quality requirement, acceptable.

Impacts: Quietness <- Marketing Specification 4.1.5.

Product Evolution: Design Idea. "These design ideas are a rough collection from the Marketing Specification".

Impacts: Evolution.

New Product shall have a modular structure <- Marketing Specification 3.2.

Modular, future proof <- Marketing Specification 4.3.3.

New Product shall be easy to upgrade <- Marketing Specification 3.3.

The switch must support remote SW loading <- Marketing Specification 4.3.4.

'Plug & Play' <- Marketing Specification 4.5.2.

Remote Software Upgrade: for both correction and upgrading proposes <- Marketing Specification 4.5.14.

Software changes shall not require manual physical access to Radio Heads <- Marketing Specification 4.5.15.

New software - upgrades, patches, new releases, etc. should require a minimum of scheduled downtime for New Product <- Marketing Specification 4.5.16.

Home Location Register: HLR:

HLR is part of the Macro cellular system?? <- Marketing Specification 4.3.7. Impacts: <unspecified>.

Low Power Consumption [Radio Heads]:

Low Power consumption will be designed.

Impacts: Quietness "in order to avoid fans and consequent noise" <- Marketing Specification 4.1.5.

Low RF Power Output [Radio Heads]:

Impacts: {<avoiding interference>, Availability, Co-existing, Per User Cost, Robustness, others} <- Marketing Specification 4.2.1.8.

Remote SW Loading:

The switch must support remote SW loading <- Marketing Specification 4.3.4.

Impacts: Maintenance.

Single Cabinet [Central Equipment]:

The Central Equipment must fit into a single cabinet including power, but not batteries <- Marketing Specification 4.5.4.

Comment: This is really a way to achieve Volume of 36 liters as estimated by TW.

RH1: Assumption: RH assumed to be single cabinet.

Impacts: <unspecified>.

Basically, what we did was to identify these design specifications as design ideas, not requirements (design constraints), and to structure them so we could see their Source and their Impact intents.

2.3.9 Diagrams/icons: the design engineering process

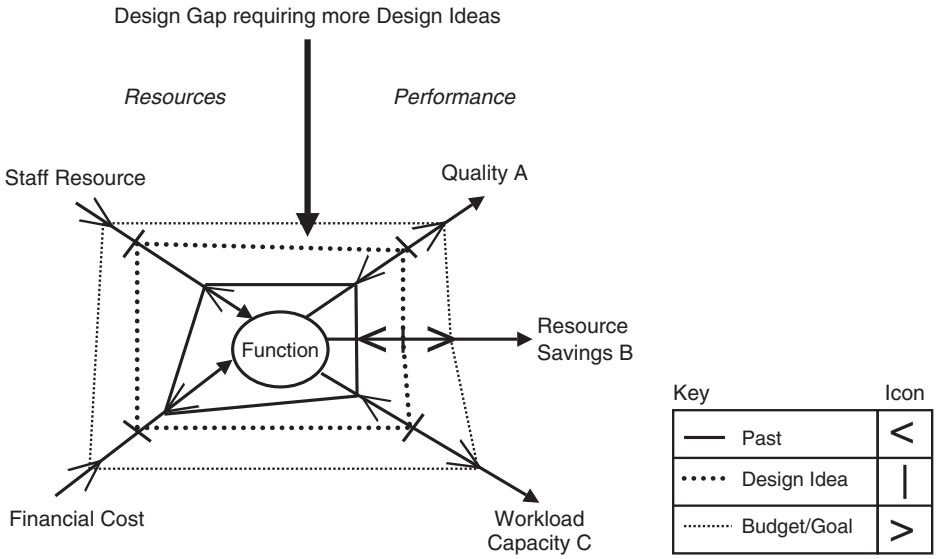


Fig. 2.3.6 Diagram showing the gap between the Past and the Goal/Budget levels and the contribution that a Design Idea makes toward filling the gap.

Design Specification Template <with Hints>

Tag: <Tag name for the design idea>.
Type: {Design Idea, Design Constraint}.

========== Basic Information ==========
Version: <Date or version number>.
Status: <{Draft, SQC Exited, Approved, Rejected}>.
Quality Level: <Maximum remaining major defects/page, sample size, date>.
Owner: < Role/e-mail/name of person responsible for changes and updates>.
Expert: < Name and contact information for a technical expert, in our organization or otherwise available to us, on this design idea>.
Authority: <Name and contact information for the leading authorities, in our organization or elsewhere, on this technology or strategy. This can include references to papers, books and websites>.
Source: <Source references for the information in this specification. Could include people>.
Gist: <Brief description>.
Description: <Describe the design idea in sufficient detail to support the estimated impacts and costs given below>.
<Term Tag here>: Definition: <Use this to define specific terms used anywhere in the specification>. "Repeat this for as many definitions as you need"
Stakeholders: <Prime stakeholders concerned with this design>.

========== Design Relationships ==========
Reuse of Other Design: <If a currently available component or design is specified, then give its tag or reference code here to indicate that a known component is being reused>.
Reuse of This Design: <If this design is used elsewhere in another system or used several times in this system, then capture the information here>.
Design Constraints: <If this design is a reflection of attempting to adhere to any known design constraints, then that should be noted here with reference one or more of the constraint tags or identities>.
Sub-Designs: <Name tags of any designs, which are subsets of this one, if any>.

Fig. 2.3.7 Design Specification Template. This is a form to fill out, with <hints in fuzzy brackets>.

========== Impacts Relationships ==========
Impacts [Functions]: <List of functions and subsystems which this design impacts attributes of>.
Impacts [Intended]: <Give a list of the performance requirements that this design idea will positively impact in a major way. The positive impacts are the main justification for the existence of the design idea!>.
Impacts [Side Effects]: <Give a list of the performance requirements that this design idea will impact in a more minor way, good or bad>.
Impacts [Costs]: <Give a list of the budgets that this design idea will impact in a major way>.
Impacts [Other Designs]: <Does this design have any consequences with respect to other designs? Name them at least>.

========== Impact Estimation/Feedback ==========
For each Scalar Requirement in Impacts [Intended] (see above):
Tag: <Tag name of a scalar requirement listed in Impacts [Intended]>.
Scale: <Scale of measure for the scalar requirement>.
Scale Impact: <Give estimated or real impact, when implemented, using the defined Scale. That is, given current baseline numeric value, what numeric value will implementing this design idea achieve or what numeric value has been achieved?>.
Scale Uncertainty: <Give estimated optimistic/pessimistic or real ± error margins>.
Percentage Impact: <Convert Scale Impact to Percentage Impact. That is, what percentage of the way to the planned target, relative to the baseline and the planned target will implementing this design idea achieve or, has been achieved? 100% means meeting the defined Goal/Budget level on time>.
Percentage Uncertainty: <Convert Scale Uncertainty to Percentage Uncertainty ± deviations>.
Evidence: <Give the observed numeric values, dates, places and other relevant information where you have data about previous experience of using this design idea>.
Source: <Give the person or written source of your evidence>.
Credibility: <Credibility 0.0 low to 1.0 high. Rate the credibility of your estimates, based on the evidence and its source>.
========== Priority and Risk Management ==========
Rationale: <Justify why this design idea exists>.
Value: <Name [stakeholder, scalar impacts and other related conditions]: Describe or quantify the knock-on value for stakeholders of the design impacts>.
Assumptions: <Any assumptions that have been made>.
Dependencies: <State any dependencies for this design idea>.
Risks: <Name or refer to tags of any factors, which could threaten your estimated impacts>.
Priority: <List the tag names of any design ideas that must be implemented before or after this design idea>.
Issues: <Unresolved concerns or problems in the specification or the system>.
========== Implementation Control ==========
Supplier: < Name actual supplier or list supplier requirements>
Responsible: <Who in your organization is responsible for managing the supplier relation?>
Contract: <Refer to the contract if any, or the contract template>
Test Plan: <Refer to specific test plan for this design>
Implementation Process: <Name any special needs during implementation>
========== Location of Specification ==========
Location of Master Specification: <Give the intranet web location of this master specification>.

Fig. 2.3.7—Cont'd

2.3.10 Summary: the design engineering process

The 'design engineering process' is a systematic, rational process of finding design specifications, which when implemented will satisfy a balanced set of requirements on time.

The term 'design engineering' means a design process based on multidimensional quantified requirements and multiple quantified design attributes. It requires concurrent use of an implementation process, like Evo, based on quantified measurement of performance and costs at frequent evolutionary cycles, and of necessary analysis and correction to maintain progress toward (potentially adjusted or traded off) formal and quantified targets.

The selection of design ideas is determined by the need to deliver a set of specified stakeholder target levels within a set of specified constraint levels.

The design engineering process is really concerned with identifying optional design ideas and evaluating the alternative possibilities to find a satisfactory architecture (that is, the sum of all design ideas), which provides:

- the best fit with the requirements,
- early delivery of key results (with high stakeholder values),
- best value to cost ratios and performance to cost ratio, and
- acceptable risks.

The design engineering process may also involve identifying the best reaction (redesign) to any feedback (good or bad feedback from actually implementing design ideas in the real system).

The design engineering process cannot usually be done, competitively, in a single pass. The effects of even a single design idea are too complex to understand without 'experience analysis' from past use of the idea, and especially without actual use on our new system. So it must normally be expected that the 'final' and 'correct' design specification can only be evolved toward (never perfectly or ideally reached) as a result of multiple feedback-and-change cycles.

Refinement of design can be done in parallel with actual use (by at least some early stakeholders) of a version of the product. The practical feedback from this early delivery can be used to improve the design; probably faster and more correctly than by staying in the 'design phase' longer.

A further complication is that as time goes on, both the 'design requirements' and 'potential and selected design technology' will 'expectedly' change, thus requiring yet another set of cycles of learning how to satisfy these new, changed requirements. Never perfect, continuously better, is the watchword.

In terms of 'Competitive Engineering' you can *always* refine the design to be more competitive. However, there is a point where the cost and time of refining the design exceeds any competitive benefit, and it is time to stop designing and to get the product out of the door, this time around.

Design Policy

Design ideas are only really finally validated when they display satisfactory attributes in a real system (that is, after successful delivery in an evolutionary step). Don't kid yourself that they are 'final' before that.

A suggested mental attitude toward design specifications. Don't believe any estimates of performance and cost, only reality as measured!

This page is intentionally left blank

Chapter 2.4

2.4

Design specification

Tom Gilb

2.4.1 Introduction: specification quality control

Specification quality control (SQC) is the name I shall use to refer to this method in this text. Within the software community, the term 'inspection' is used. However, it is a poor choice for engineering communities, which already use 'inspection' in another sense during final production line quality control (QC). SQC is remote from such assembly-line inspections, as it takes place from the earliest stages of idea specification and has different organizational impacts (for example, team building and assisting in 'on the job' training).

The primary purpose of SQC is systems engineering process control through sampling measurement of specification quality. Through SQC, we can improve systems engineering processes, save project time, and increase systems engineering productivity.

2.4.1.1 Improving process

Control of projects, designs, strategies, marketing, selling and buying, management planning, and programming, all have one thing in common at least – they rely on ideas *specified* by people, and *read* by people. If those ideas are misunderstood by the reader, incomplete, wrongly written, or out of date, then we are doomed to lose control and be less competitive, no matter how well we design, plan, and implement!

For *software*, studies have long since shown that a considerable percentage (44% at Bellcore and 62%) of all bugs in computer programs were not due to faulty programming. They were due to faulty requirements and design being handed to the programmers and the testers. In many cases, the testers, unwittingly, checked that an erroneous specification was 'correctly' programmed! Testing, in this situation, does not solve the problem: it confirms it. However, SQC can address such problems.

In *aircraft design* at Douglas Aircraft (now Boeing), 'engineering order' faults cost $2965 each to correct and 30% of engineering orders needed correction. After SQC was applied in 1987–88, the percentage of faulty engineering orders fell to 0.5% (Personal Experience). We achieved similar results in 1989 at Boeing, Renton on all aspects of aircraft design.

The tendency to commit some kind of error, when communicating complex ideas in writing to other people, is much worse than most people realize. My own experience in industrial measurement of defects suggests that technical documents, initially and routinely, contain at least 20–60, and often far more, 'major' engineering specification defects in each 'logical' page (300 non-commentary words). Through systematic use of feedback from SQC to specification writers, this level can be brought down to well under one remaining major defect/page (British Aerospace, Eurofighter Project, Wharton, achieved this in 18 months (Personal Communication)).

Competitive Engineering; ISBN: 9780750665070
Copyright © 2005 Elsevier Ltd. All rights of reproduction, in any form, reserved.

2.4.1.2 Saving time

Without SQC, a major defect left in a technical specification can cost an average of 9.3 work-hours to deal with.[1] Use of SQC at an early stage (during writing the specification) would cost only 1 work–hour to remove it. "A stitch in time saves nine" (or an SQC hour saves nine-point-three to be exact!).

2.4.1.3 Increasing productivity

The reduction of defects (as a result of using SQC) saves 'rework', which is otherwise about half of all effort in software projects. Raytheon found that software engineering productivity for about a thousand programmers increased by a factor of 2.7 over a few years of using SQC (Inspection and Defect Prevention Process (DPP)).

One major reason for defect reduction is the 'training effect' of SQC on individuals. The number of defects injected by a systems engineer reduces by about 50% each time they go through an SQC process (Personal Experience since 1988). Systems engineers rapidly learn to take the rules seriously. They see that their peers expect them to comply with the rules and that their work cannot exit, and be 'finished', until they reach at least the exit level for the estimated number of remaining major defects. I have found that this is as true in software as it is in hardware engineering.

2.4.1.4 Industrial usage

The methods needed for QC of specifications originated in the early 1970s within IBM, when they were used under the name of 'Design and Code Inspections'.[2] Since then significant changes have occurred, resulting in the SQC method. The most notable change was the introduction, again within IBM, of the DPP. The other major change is the shift to 'sampling', rather than 100% checking and trying to clean up defects.

The first large-scale *hardware* engineering uses of SQC took place at Douglas Aircraft (1988) and Boeing (1989) under this author's guidance. In recent years, Siemens, Alcatel and Ericsson have also successfully used the method on a large scale (hundreds trained) for total product development purposes. Hewlett Packard has reported estimated savings due to SQC (some use within hardware product planning) of $21.5 million and $34 million in 1993 and 1994, respectively.

The use of SQC *outside* of the software area is, as yet, little understood or appreciated, except by the few corporations who have tried it out such as Ericsson, Douglas, and Boeing. It is time that this industrial experience was more widespread knowledge. There is little difference in the specification of software engineering, management planning, or hardware engineering with regard to human specification errors, their causes, and their consequences.

2.4.2 Practical example: SQC

Take the simple performance requirement statement:

'The objective is to get higher adaptability using modular structure.'

Do you see any problems with it? Is it similar to statements you see everyday? You notice that some

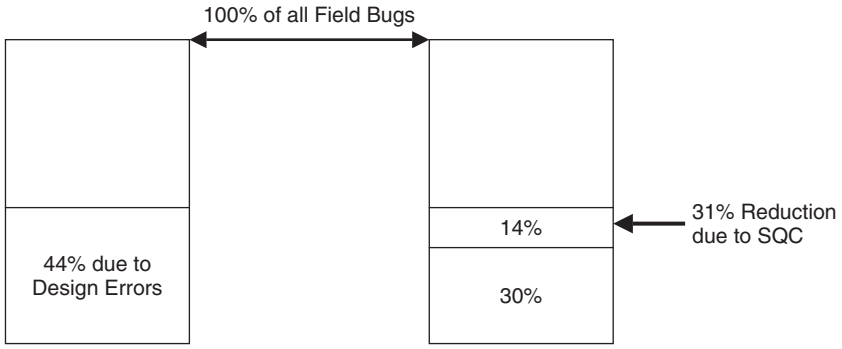


Fig. 2.4.1 Due to use of SQC during development of telecommunications software, a 31% reduction in design errors that caused bugs in the field was measured after 2 releases (Pence and Hon 1993).

[1] As measured on a 1000 defect sample by (then) Thorn EMI (electronics industry) in 1990. See Section 2.4.8.
[2] Fagan, M. E. 1976. Design and code inspections. *IBM Systems Journal*. Volume 15. Number 3. Pages 182–211. Reprinted 1999. *IBM Systems Journal*. Volume 38. Numbers 2 and 3. Pages 259–287. See http://www.research.ibm.com/journals/sj/

rules we have suggested have been violated. Of course, there is nothing wrong with it, unless we agree that these rules are in force. For some purposes they should be in force, for others not.

SQC works by using the formal rules that are in force: a 'defect' is a rule violation. SQC discovers whether people have applied the agreed rules or not. A specification writer must always know the rules that apply (and have agreed to them in advance). The specification writer should welcome any help to follow them. Rules, after all, should be 'best practice' rules.

Let us now (for the sake of this example) introduce a few short rules, which apply to the quality requirement statement above.

Rules For Performance Requirements

Tag: Rules.OBJ.

Clear: They must be unambiguously clear to the intended readers (not to 'anyone', just the relevant people).

Detail: They must detail complex concepts as a set of elementary measurable concepts.

Scale: They must specify a scale of measure to define the concept (all performance attributes are quantifiable).

Quantify: They must specify at least two points of reference on the defined Scale to define 'relative' terms, such as "higher." These are called the benchmark and target specifications. Qualify: Targets must specify exactly 'when' a performance level is to be available. Other qualifier notions, such as 'where' and 'if' should also be made explicit, if the target is not elsewhere specified.

Ends: They must not put 'designs' in the specification of 'performance requirements'. Specify the Ends, not the Means.

Source: The source statements for each requirement must be precisely referenced (for example, <- the contract and marketing documentation).

Fuzzy: Fuzzy unclear concepts shall be marked with <fuzzy/angle brackets> to indicate there is room for improvement.

A checker (a person assigned to check a specification and its selected source documents against these rules) would be obliged to report, for the performance requirement statement about 'higher adaptability', that all the above rules were violated.

There are, therefore, at least eight defects in the requirement statement. If these defects *might* have much higher costs later in a project (if not fixed at specification time), they should be classed as 'major' defects. Majors are the defects it pays to fix now, at a tenth of the cost we would otherwise suffer later. (Fixing majors early is useful, but preventing their injection is even more profitable.)

Checkers are friendly, confidential personal advisors to the specification writer. The checker's first job is to point out potential problems for correction before a specification is released to other engineers, or to customers. Checking is a service the writer will likely perform, in return, when their former peer checkers specify something themselves, and want SQC help. The responsible engineer will take a list of the checker's suggested advice regarding 'potential defects' (issues) and consider correcting them. They should address similar defects, outside the sample checked, as far as necessary, according to the applicable rules, procedures and source documents. However, it may pay off to totally rewrite the specification. The specification document 'Exit Level' is based on a general calculation of what is the best project time-saver. We don't exit, if cleaning up the specification now saves the most time, in the long run. **The following are the expected results** of **a *single pass* of SQC:**

(Note: Multiple passes should be rare.)

1. Based on defects found and corrected and, on an assumed SQC effectiveness at spotting defects of 50%, a calculation will be made about the (probable) remaining major defects in the specification (which is about as many defects as we found – since we cannot expect to be much better than 50% effective in finding defects). If these are more than permitted by the exit conditions, the specification will not be released. This is because the estimated unfound remaining majors would cause more loss of time than savings to be gained, if we let them exit downstream; that is, if we released the specification immediately.
2. he specification writer will learn about current agreed rules and their peers' interpretations of these rules. As a result they are likely, by my industrial experience, to learn to produce a specification with half the number of defects next time. (Ultimately, after several SQC experiences for the writer, about 100 times cleaner – using major defect reduction as the measure – specification is usually achieved!)
3. The checkers themselves will learn best practice rules and their peers' attitudes toward those practices. This will influence the checkers' specification work quality.
4. The 'users' of the specification will learn to expect (in terms of their entry condition) a minimum specification quality level (such as no more than one remaining major defect/page).

5. The SQC team will continuously suggest process improvements to reduce future major defects. (Poor working processes, training, tools and the working environment 'force' defects on the workforce according to Deming).
6. Project productivity will at least *double*, mainly due to fighting fewer defects later (Dion has reported productivity increasing by a factor of 2.7).

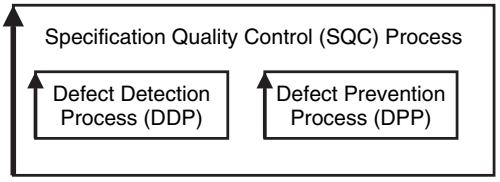
As a result of SQC we will have data to decide if it pays off to release the specification to another engineering process, or fight the defects now.

2.4.3 Language core: SQC

2.4.3.1 Basic definitions

2.4.3.1.1 Specification quality control

SQC consists of two main processes: the Defect Detection Process (DDP) and the DPP.



2.4.3.1.2 Defect detection process

The DDP is concerned with document quality, mainly with identifying defects in the documentation and using this information to make decisions about how best to proceed with the main document under SQC – the main specification.

Ideally, though sometimes not done due to the economics of the situation, a known defect must be removed as soon as possible after the error has been committed. This is to avoid the high cost of late removal (at test or in field) of the defect, or to avoid the high cost of its consequences. *"A stitch in time saves nine."*

2.4.3.1.3 Defect prevention process

The DPP is concerned with learning from the defects found and suggesting ways of improving processes to prevent them reoccurring in future. The process improvement suggestions are routed on to the relevant process owner for further consideration. *"An ounce of prevention is worth a pound of cure."*

Here are some other basic SQC concepts.

2.4.3.1.4 Issue

An issue is a *perceived* defect in a document. It is a non-confrontational way for a checker to say, "I think I may have identified a defect."

2.4.3.1.5 Defect

A failure to observe a formal, written, required rule. It is not a personal opinion or personal taste. It is failure to observe a group norm, or required best practice.

2.4.3.1.6 Major defect

A major defect is a defect (rule violation) which, if not fixed at the requirements or design stage of specification, will possibly grow approximately an order of magnitude or larger in 'cost-to-find-and-fix' and/or damage potential. It is often intentionally written with a capital 'M'. Minor defects tend not to be economic to identify or fix *(but you sometimes have to identify them to determine that they are indeed minor and not, major)*.

2.4.3.1.7 Page

A logical page, as opposed to a physical page, is defined as a specific number of non-commentary words. If no other definition is given then use '300 non-commentary words' for each logical page (default 'volume' definition). This ensures measurements of checking rates and defect densities are consistent.

2.4.3.1.8 Checking rate

The checking rate is the average speed with which an individual checker searches a specification for defects, allowing time for checking it against rules, sources, kin documents, and checklists. This is a critical factor to control for effective checking. You have to go surprisingly slowly to raise your checking effectiveness from 5% to 50%. (For example: one page an hour!)

2.4.3.1.9 Optimum checking rate

The optimum checking rate is the rate, which gives the highest checking productivity (effectiveness in finding majors). It is the checking speed that in fact works best on a given document type for an individual checker to do their assigned tasks. It is found by establishing the most effective average historical checking rate in terms of finding major defects. The optimum checking rate is usually in the range of 300 non-commentary words/hour (plus 300/minus 270). This is used as a guide for team planning. Individuals need to tune in to their personal

optimum rate, which varies from this average. The major trick to going at this 'slow' rate is to *sample*, not to attempt 100% checking of all pages and consequent 'defect clean up'.

2.4.3.1.10 Remaining major defects

The remaining major defects are the *estimated* remaining major defects 'volume' measure (which could be a page, a technical drawing or an entire specification) given for a sample or an entire specification. It is estimated based on 'total found' and 'known % effectiveness'.

2.4.3.1.11 Checklist

A 'checklist' is a list of questions, which can be asked about a document's contents by a checker, with a view to improving the effectiveness of that checker in finding major defects. Checklist questions are always directly derived from individual official rules. They are not allowed to be the rules, or to change the rules, just to *interpret* them.

2.4.3.1.12 Rule

A rule is a standard for the production of a *written* process output. A rule can be used to judge the objective quality ('defect-freeness' according to current rules) of a written process output. Violations of rules are defined as 'defects'.

Rules are often grouped into sets according to the type of standard, which they are setting (for example, 'specification clarity' or 'specification consistency').

2.4.3.1.13 Main specification

The main specification is one of potentially *many* documents involved in a single SQC. However, it distinguishes itself as the one we are trying to get formal 'exit' for. Exit (acceptable exit level) is based primarily on the specification's quality with respect to the official systems engineering standards (rules) for writing it.

2.4.3.1.14 Source documents

The source documents are the 'parents' used to produce a specific main specification. For example, contracts are typical sources for requirements. Requirements are a source for design. Requirements and design are sources for Impact Estimation. Design is source for planning, estimating, and construction or programming. Change requests are sources for an updated specification.

2.4.3.1.15 Kin documents

'Offspring of the same 'parent' (source) documents are 'kin'. *For example, test plans, source code, and user handbooks could all be derived from the same requirements or the same design*. The use of kin documents is that they can serve as information to perform defect checking in SQC.

2.4.4 Standards: SQC

Rules are standards, and are central to the SQC process; specifications must be checked against their agreed specification rules. However, the rules to be used depend on the specification type, so we won't attempt to list them here. The rules given else where in this text are suitable examples of such rules (*but they are by no means a complete list*).

Here is a list of guidelines for assessing whether your overall SQC process is functioning correctly.

2.4.4.1 Guidelines for assessing functioning of overall SQC

Economic: SQC must always make economic sense. If SQC is not saving in the order of 10 hours for every hour spent on SQC, then your SQC process should probably be modified or abandoned.

SQC Champion: There must be an SQC champion within the organization. (At the very least, a nominated person responsible for SQC; an SQC process owner.)

Team Leaders: There must be a list of current SQC team leaders. It should show that there is a sufficient number of team leaders within the organization and also that the team leaders are trained, tested, and 'certified' to ensure they know what they are doing.

Statistics: The SQC statistics must be up-to-date on the SQC database.

Meetings: All meetings must be of maximum length of 2 hours (tiredness reasons). If more time is needed, schedule a set of such meetings (*but do consider the possibility of using sampling*).

Checkers: Unless you are training novices, the number of checkers at a meeting should be five or less. Two or three people is typically most cost-effective, four to five is generally more 'effective.'[3]

[3] The original evidence for this came from research performed by Søren Nielsen in the Danish electricity industry (Danish Technical Institute), and was confirmed by further research at Jet Propulsion Labs by John Kelly. Optimum effectiveness (number of unique issues per checker) was achieved with teams of 4–6 people, optimum efficiency (cost per unique issue found) with teams of 2–4 people. The recommended team size of 4–5 people achieves the best compromise between these factors. It was Edward Weller, analysing data from more than 6000 inspection meetings conducted at Bull HN, who reported that "four-person teams were twice as effective … as three person teams." Also included in the work by Wheeler, Brykczynski, and Meeson in 1996.

Example of a Filled-In Master Plan

SQC Team Master Plan
SQC ID 57 Team Leader Lucy Jones Mail/tel. code 3322
Writer(s) Sam Murry Mail/Tel. Code 3321 Date SQC was requested 20 Jun 2000
Spec. Title Penn Marketing Plan Total physical pages 4 Version 0.1
Intended purposes of this SQC QC
Entry Conditions which apply (tags) Penn Objectives INSPT'D [✓] EC
(Generic Entry Condition SI pg. 64–66)

Current Entry States (met, waived) Met Why?
Exit Conditions which will be applied (tags) Edit/CR/RemDefects [ ] XC
(Generic Exit Conditions SI pg. 202)

Meetings
Kickoff Date 3 Jul 2000 Location Room 4 Start Time 10.30 End time 11.00
Spec. Date 10 Jul 2000 Location Room 4 Start Time 10.00 End time 12.00
Process: Date 10 Jul 2000 Location Room 4 Start Time 12.15 End time 12.45

Documents (specified parts to be used by checkers) **Master Plan**
samples or check(s)
Specification(s)
Penn Marketing Plan (all 4 pages)
Potts Marketing Plan (Pages 2,3)
Penn Objectives (Pages 2–5)

Rules: Generic [ ] SI 424-5 or in hours Rules.GR
Specific Rules.MP

Checklists: For Spec. Check.MP
For other Documents Check.OBJ

Team Setup:

Team Member Name	Tel. Ext.	SQC Role Soft. Insp. Page 362-73, e.g. Editor, Checker	Specification Part (The Specific section or pages of the document)	Source Documents and Sections you are responsible for	Rules & Checklists	Checking Procedure & other tactic	Checking Effort in hr.
Jenny Claire	2626	Checker	Penn Pages 2–3	Potts-MP	Rules.GR, Rules.MP, Check.MP	PCK PCC PCL PCB	2
Tom Franks	2533	Checker	Penn Pages 2–3	Potts-MP Section 2	Rules.MP Check.MP	PCK PCC PCL PCB	2
Harry Matthews	2522	Checker	Penn Page 1,4	Penn-OBJ Potts-MP	Rules.MP Check.MP	PCK PCC PSL PCB	2
Sam Murry	3321	Writer Checker, Editor	Penn Page 1,4	Penn-OBJ Potts-MP	Rules.GR Check.OBJ	PCK PCC POL PCB	2
Lucy Jones	3322	Team Leader	–	–	–	PLK PLC	–

Recommended Average Team Checking-Rates, SQC Goal and Strategy
Numeric SQC Goal, set during kind off
Strategy to meet SQC Goal

Optimum Checking Rate, for this type of specification is 1 pages per hour, of non-commentary text.

Spec meeting checking-rate: 2 page(s) (300 words, Non-Commentary) per hour (optimum rate of checking during the Spec meeting)

Lucy Jones this is the end of the Master Plan.

© Gilb

Individual Checker Data Collection
(filled in by each checker, after checking and before the Spec. meeting)

Actual work-hours (tenths) spent: No. of (300w NC) Pages checked at optimum rate:
Major issues [incl. Exxx-Majors (project threat)], minor issues
Process improvement suggestions ?s of intent (to author)
My Checking Rate was: Pages/hour.
How does this deviate from your planned rate? Why?
and of Individual Data Collection

Fig. 2.4.2 Filled-in master plan.

Example of a Filled-In Editor Advice Log

Date 10 Jul 2000 Start time 1000. End time 1200. Editor Advice Log

SQC ID 57 Page 1 of 12

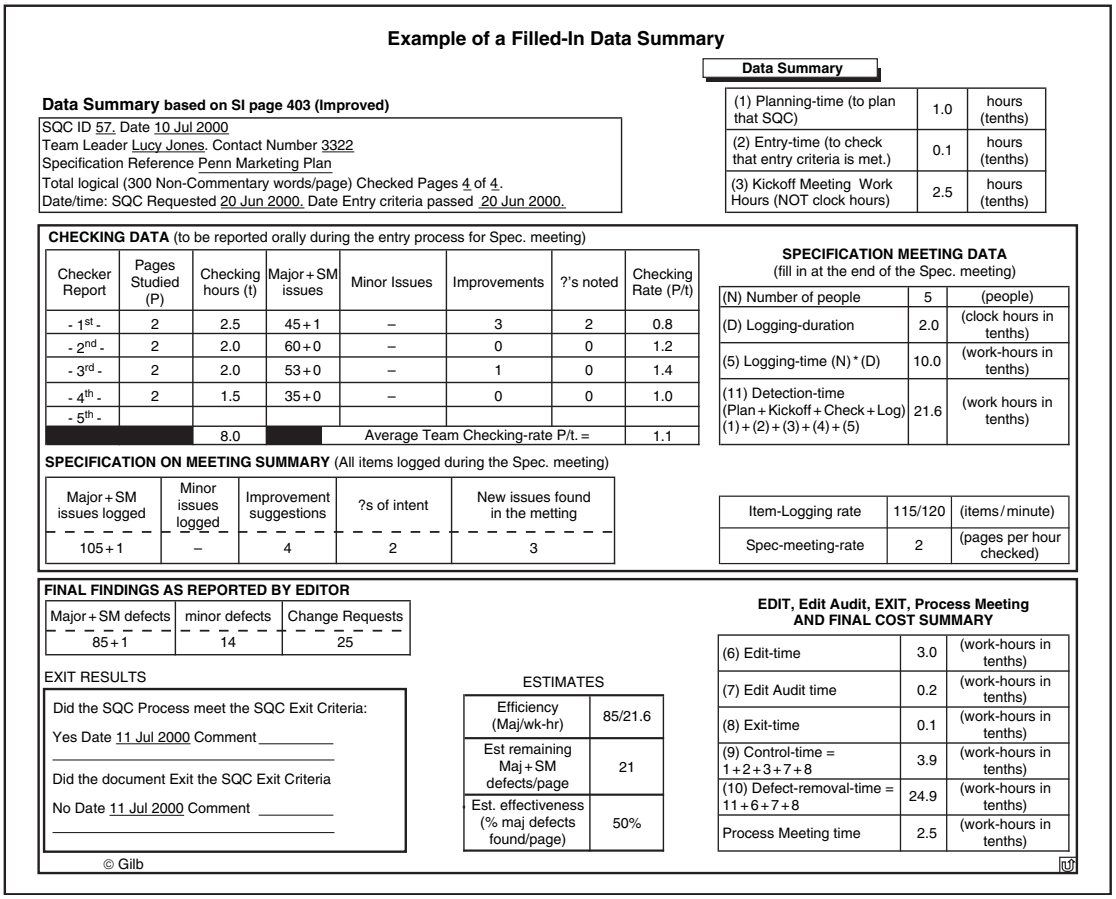
Item No.	Document Reference Tag	Page	Line or Tag	Exact Location	Type of Item	Checklist or Rule Tag	Source Inconsistency and/or Necessary Description	Occurs	Editor Action (during editing)
1	Penn	1	4	–	Major Minor ? Imp. New	SIMPLE GR2	No breakdown for figures	7	Breakdown of figures from notes entered.
2	Penn	1	5	–	Major Minor ? Imp. New	SOURCE GR4	Lack of Source Info	3	Sources quoted.
3	Penn	1	5-6	–	Major Minor ? Imp. New	CLEAR GR	Not Understood		Rewritten
4	Penn	1	-	Diag I	Major Minor ? Imp. New	INCONSISTENT GR 11	Penn-OBJ shows additional inputs		Penn-OBJ correct. Corrected.
5	Penn-OBJ	3	15	After "("	Major Minor ? Imp. New	CLEAR GR1	Not Understood		Change Request on Penn-OBJ
6	Penn	1	30	–	Major Minor ? Imp. New		Not clear if Scotland was considered?		Corrected. Scotland had been ruled out in Phase I
7	Penn	2	16	–	Major Minor ? Imp. New	MP1	Product Ref for fuller desc. is missing		Ref. to Product Desc. added.
8	Penn	2	22	–	Major Minor ? Imp. New	MP4	No promotion info		Ref. to Promotion Plan added.
9	Penn	2	22	–	Major Minor ? Imp. New	MP5	Project mgr Authority		Rejected more senior mgt decide.
10	Penn-OBJ	2	29	–	Major Minor ? Imp. New	CHECK OBJ.3	Qualifiers for USABILITY inadequate		Change Request on Penn-OBJ

Subtotals:
New Items found during the Spec. Meeting 10.
Major issues logged 9. Minor issues logged –.

Improvement suggestions logged 0. ? Questions of Intent logged 1.

© Gilb

Fig. 2.4.3 Filled-in editor advice log.



Example of a Filled-In Data Summary

Data Summary

(1) Planning-time (to plan that SQC)	1.0	hours (tenths)
(2) Entry-time (to check that entry criteria is met.)	0.1	hours (tenths)
(3) Kickoff Meeting Work Hours (NOT clock hours)	2.5	hours (tenths)

Data Summary based on SI page 403 (Improved)

SQC ID 57. Date 10 Jul 2000
Team Leader Lucy Jones. Contact Number 3322
Specification Reference Penn Marketing Plan
Total logical (300 Non-Commentary words/page) Checked Pages 4 of 4.
Date/time: SQC Requested 20 Jun 2000. Date Entry criteria passed 20 Jun 2000.

CHECKING DATA (to be reported orally during the entry process for Spec. meeting)

Checker Report	Pages Studied (P)	Checking hours (t)	Major + SM issues	Minor Issues	Improvements	?'s noted	Checking Rate (P/t)
- 1st -	2	2.5	45 + 1	–	3	2	0.8
- 2nd -	2	2.0	60 + 0	–	0	0	1.2
- 3rd -	2	2.0	53 + 0	–	1	0	1.4
- 4th -	2	1.5	35 + 0	–	0	0	1.0
- 5th -							
		8.0		Average Team Checking-rate P/t. =			1.1

SPECIFICATION MEETING DATA
(fill in at the end of the Spec. meeting)

(N) Number of people	5	(people)
(D) Logging-duration	2.0	(clock hours in tenths)
(5) Logging-time (N) * (D)	10.0	(work-hours in tenths)
(11) Detection-time (Plan + Kickoff + Check + Log) (1) + (2) + (3) + (4) + (5)	21.6	(work hours in tenths)

SPECIFICATION ON MEETING SUMMARY (All items logged during the Spec. meeting)

Major + SM issues logged	Minor issues logged	Improvement suggestions	?s of intent	New issues found in the metting
105 + 1	–	4	2	3

Item-Logging rate	115/120	(items / minute)
Spec-meeting-rate	2	(pages per hour checked)

FINAL FINDINGS AS REPORTED BY EDITOR

Major + SM defects	minor defects	Change Requests
85 + 1	14	25

EXIT RESULTS

Did the SQC Process meet the SQC Exit Criteria:
Yes Date 11 Jul 2000 Comment _____
_____
Did the document Exit the SQC Exit Criteria
No Date 11 Jul 2000 Comment _____
_____

ESTIMATES

Efficiency (Maj/wk-hr)	85/21.6
Est remaining Maj + SM defects/page	21
Est. effectiveness (% maj defects found/page)	50%

EDIT, Edit Audit, EXIT, Process Meeting AND FINAL COST SUMMARY

(6) Edit-time	3.0	(work-hours in tenths)
(7) Edit Audit time	0.2	(work-hours in tenths)
(8) Exit-time	0.1	(work-hours in tenths)
(9) Control-time = 1 + 2 + 3 + 7 + 8	3.9	(work-hours in tenths)
(10) Defect-removal-time = 11 + 6 + 7 + 8	24.9	(work-hours in tenths)
Process Meeting time	2.5	(work-hours in tenths)

© Gilb

Fig. 2.4.4 Filled-in data summary.

Checking Rate: All checking must be carried out near the relevant optimum checking rate. This rate will vary by document type and organization. It is about 1 page/hour.
Condition: Entry and exit conditions must be taken seriously. They are there to save you wasting time. The number of remaining major defects/page for successful exit must be ultimately less than one (major defect/page).
Standards: There must be an up-to-date (intranet) 'library' of official rules, checklists, and forms.
Upstream Pollution: The number of major issues identified by your team in source specifications, which have just previously - exited SQC, should be approximately 15% of the total number of logged issues. Otherwise, this is a sign that your team is not taking the 'second-round' opportunity to find source defects, seriously.

2.4.4.2 Forms:

SQC uses four main forms: the Master Plan, the Editor Advice Log, the Data Summary and the Process Meeting Log. There are examples of these forms filled in, in Figures 2.4.2 – 2.4.5. Blank forms are given in Section 2.4.9

Note forms are a 'procedure' (in the format of the form) for gathering data. Most of our clients have their own local variation of the forms and automate them (usually on an intranet web site).

2.4.4.3 SQC process roles and responsibilities

An **efficient team** *(most major defects/work-hour) uses 2 or 3 people in total. An* **effective team** *(most major defects/page) uses a maximum of 3 – 5 people in total.*

2.4.4.3.1 Team leader

A team leader is responsible for managing an SQC process. The team leader is responsible for knowing SQC thoroughly and helping the team members to perform. They follow the 'best-practice' SQC processes. An SQC team leader is normally trained for about a week, and is then formally approved to practice by virtue of their practical ability and continued correct

Process Meeting Log

Team Leader Lucy Jones Date 10 Jul 2000 SQC ID 57 Page 1 of 1

Item	Issue Reference	Cause Class (tick1)	Root Cause Ideas	Improvement Ideas
1	2 →	Communication Oversight Transcription Education	Lack of importance attached to such information	Have a header page insisting on such info
2	15 →	Communication Oversight Transcription Education	New Legislation has not been published	Send out e-mail to all managers in Division
3	33 →	Communication Oversight Transcription Education	Only partially transferred	Insist on use of one master.
4		Communication Oversight Transcription Education		
5		Communication Oversight Transcription Education		
6		Communication Oversight Transcription Education		
7		Communication Oversight Transcription Education		
8		Communication Oversight Transcription Education		
9		Communication Oversight Transcription Education		
10		Communication Oversight Transcription Education		

© Gilb

Start Time	Stop Time	Duration	No. People	Total Cost
1215	1245	30 Mins.	5 People	2.5 Workhours

Fig. 2.4.5 Filled-in process meeting log.

practice. Inadequate SQC team leader training leads to failure of the SQC process.[4]

2.4.4.3.2 Checker

Checkers are primarily 'consultants to the writer' and their detailed knowledge of the defectiveness of the writer's work is confidential. Almost all engineering team members work as checkers on occasion, including the writer and probably the team leader. (The team leader might choose to be a 'non-playing captain' of the team. They would not check in order to focus their time on the team leader responsibility or because they were not technically capable in the specification 'language'.)

Checkers are SQC team members who actively check a set of documents: the main specification, its source specifications, kin specifications, the rules, checklists, and procedures. They focus on using the

[4] Grady reported that HP failed to achieve results from 1983 to 1988 until they properly trained their team leaders on a week-long course (designed and held, as cited there, by this author). This article is reproduced in the work by Wheeler, Brykczynski and Meeson in 1996.

checklists and rules to find major defects. Exactly which documents a specific checker uses, and what they check for, is determined by the role or roles assigned to them by the team leader.

Checkers are also invited to submit specific comment on possible improvements to the process and the process standards (procedures, rules, entry conditions, exit conditions, and forms). They will, hopefully, get some insights during their checking work (for example, about the need for better rules).

2.4.4.3.3 Writer: also known as author

The writer is the person currently responsible for writing or updating a specification. The SQC process serves the writer primarily: in confidence. SQC serves the organization secondarily.

2.4.4.3.4 Editor

The editor is usually the same person as the writer. The editor is the person, who takes over the issues in the Editor Advice Log, decides (based on standards) what action is required and carries it out. Some issues will be defects and need fixing. Some issues will require clarification. Some issues will be rejected and others will require change requests to other documents to be raised. The Editor Advice Log can be updated with the editor's decisions.[5]

In extreme cases, but unfortunately all too frequently, the defect density found (for example, 90 majors in a page) will effectively spell out the fact that 'burning' the work and completely rewriting it will be more cost-effective.

2.4.4.3.5 Scribe

The scribe writes up the Editor Advice Log or other team notes at an SQC meeting. This can be any one of the team members. By default, the team leader will scribe. 'Who scribes' is not a critical decision.

2.4.4.3.6 Others

In a larger organizational setting, there are 'players' outside the team who support the SQC process. These include quality management, SQC process champions, process owners (for both SQC processes and the work processes, for example, 'Requirement process owner'), senior SQC team leaders, SQC process trainers and engineering data analysts (perhaps specialized in SQC data statistics). When the SQC process is applied to perform specification content reviews, the participants will be senior staff expected to use judgment and to take responsibility for the consequences of their approval.

2.4.5 Process description: SQC

2.4.5.1 Process: SQC

Tag: Process.SQC.
Version: October 7, 2004.
Owner: TG.
Status: Draft.

Note: Reader is referred to the 1993 work of Gilb and Graham for more detail on the sub-processes. All sub-processes are DDP unless marked as DPP.

2.4.5.2 Entry conditions

E1: The specification writer must have requested the SQC hoping to get help and exit validation for the specification.

E2: A team leader for the SQC is found from amongst the '*approved* team leaders' group.

E3: All relevant documents (main specification, kin documents, source documents, rules, checklists, and forms) are available and ideally have successfully exited SQC – apart from the main specification!

2.4.5.3 SQC sub-processes

2.4.5.3.1 Entry

The team leader ensures that the SQC *entry conditions* are met. This includes obtaining the relevant documents and checking their status. Entry conditions are evaluated *during* the Planning phase.

2.4.5.3.2 Planning

The team leader produces the master plan for the SQC (about 1 hour's work). This involves deciding what material within the specification is to be sampled, what documents are to be included, what rules must be used, who is going to be on the team and what their roles are. The optimum checking rate is determined based on history.

2.4.5.3.3 SQC strategy

The team leader decides the purpose(s) of this SQC and ensures a suitable overall SQC strategy. Again this is evaluated *during* Planning.

[5] There are many ways to report what editing action has been undertaken and any suitable method is fine.

2.4.5.3.4 Kickoff

The team meets at a Kickoff meeting, where the team leader makes sure every member knows what they *need to know* about the SQC process and the project documents, and that they are committed to working the plan *as a team*. The team may approve a suggested 'quantified goal' and 'appropriate strategy' to meet it (a DPP component).

2.4.5.3.5 Checking

The team members individually carry out their assigned defect-search roles at their self-adjusted optimum checking-rate, looking for major defects. They collect data about the cost and result of their personal checking activity. This process will typically, for a sample of about two logical pages, take 2 hours for each person. Checkers will ask the team leader for help if necessary. They will also report to the team leader any unusual or serious problems they discover that might impact the future course of the SQC process, for example, that the number of issues (potential defects) discovered is sufficiently large to consider abandoning the SQC.

2.4.5.3.6 Specification meeting

This is a team meeting (real or virtual) of up to 2 hours duration. The duration and meeting content depend on data collected from Checking.

- *Checkers' report*: At the beginning of the specification meeting, checkers *report* their data from Checking. The team leader evaluates this data and makes decisions about how the meeting and the rest of this SQC process should proceed. The meeting may be cancelled or modified in content and duration.
- *Issue logging*: The checkers report their issues, mainly potential majors, which can be in any of the participating documents. A scribe logs issues in the 'Editor Advice Log'. There should be no discussion concerning the issues discovered, just unconditional logging of the issue (the rule violation and its location in the specification). Checkers may also make process improvement suggestions *(Note: This is part of the DPP process)*, and log technical 'questions of intent' to the writer. Issue logging within a specification meeting takes up to 30 minutes.
- *Double checking*: If it is desired that additional defects are found, then double checking at the experienced specification meeting[6] optimum checking rate will be carried out during the meeting. This identifies about 15% additional major defects and adds about 1.5 hours to the meeting. This extra checking is only useful in 'cleanup mode', not when sampling and measuring to determine exit (normal mode).

2.4.5.3.7 Process meeting

After the specification meeting and a short break, the team optionally may spend up to 30 minutes, analyzing up to 10 logged potential major defects. For each chosen potential defect, 1 minute is spent describing the issue, 1 minute is spent brainstorming possible root causes, and 1 minute is spent brainstorming preventative cures. This data will later be recorded in a quality assurance (QA) database as inputs ('grass root insights': suggestions, hints, ideas) to the organization's more-systematic and formal process improvement specialists. (*Note: A Process Meeting is part of the DPP process*.)

2.4.5.3.8 Edit

The editor (usually the specification writer) takes over the 'Editor Advice Log', which consists of the issues (that could warrant correction or action) logged at the specification meeting. The editor examines the logged issues, determines how to resolve them and then at least fixes the issues considered to be major defects. The editor may discover additional defects and should make corrections to any majors identified outside the sample checked. Other reasonable action is taken, such as sending out change requests to owners of other documents. An extreme edit is a full rewrite according to all rules.

2.4.5.3.9 Edit audit

A process carried out by the team leader to verify that a reasonable and complete editing job has been done. Consequently, the editor takes formal responsibility for the editing. This can be done in minutes.

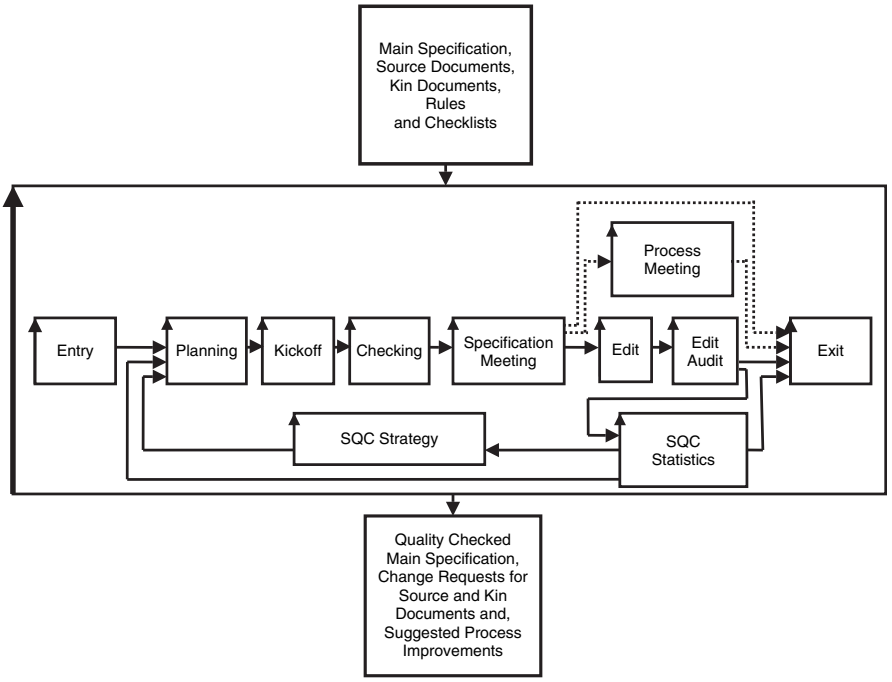
2.4.5.3.10 SQC statistics

The team leader will ensure that all the required statistics from the SQC are captured in the SQC database. This assumes a process control use of SQC data.

2.4.5.3.11 Exit

The team leader evaluates the formal SQC exit conditions to see if the specification may be released 'economically' for normal use. The estimated number of remaining major defects in the specification is especially important. If the main specification is not released, the team leader must

[6] This rate is similar but may vary by about 30% from the optimum rate average found for individual checking activity. In addition, it is a group activity rate and is not directly tunable to single individuals. Of course, single individuals will exploit the given time more or less effectively, depending on their personal ability and motivation.



Note: The 'Process Meeting' sub-process is exclusively a part of Defect Preventio Process (DPP). All the rest is Defect Detection Process (DDP), although there may be a small component of DPP within some of these sub-processes.

Fig. 2.4.6 Diagram of the SQC process showing the sub-processes.

work toward acceptable exit-levels of quality, usually in cooperation with the specification writer.

2.4.5.4 Exit conditions

X1: The main specification must have fewer remaining major defects/page than the agreed exit standard (a maximum of 'one remaining' is a reasonable ambition level, initially).

2.4.5.5 A simplified SQC process

SQC as described in the procedure above is the full-scale version. There are situations when a simplified SQC process is more appropriate (for example, to obtain a rapid assessment of the specification quality of a contract or to demonstrate to management some of the power of SQC to get their 'buy-in').

A 'Simplified SQC Process' is presented below.

Note: There are several limitations to this simplified process:

- it is only a small sample so the accuracy is not as good as a full or larger sample,
- the team will not have time or experience to get up to speed on the rules and the concept of major defect,
- a small team of two people does not have the known effectiveness of three or four people
- you will not have the basis for making corrections to the entire specification
- the checking will not have been carried out against all the possible source documents. (Usually in the simplified SQC process, no source documents are used and memory is relied on. While this means that the checking is not nearly as accurate, it does considerably speed up the process.)

However, if the sample turns up a defects density estimation of 50 – 150 major defects/page (which is quite normal), that is more than sufficient to convince the people participating, and their managers, that they have a serious problem.

The immediate solution to the problem of high defect density is not to remove the defects from the document. The most effective practical solution is to make sure each individual specification writer takes the defect density criteria (and its 'no exit' consequence) seriously. They will then learn to follow the rules and, as a result, will reduce their personal defect injection rate. On average, a personal defect injection rate should fall by about 50% after each experience of using the SQC process. Widespread use of SQC will result in large numbers of engineers learning to follow the rules.

To get to the next level of quality improvement, the next step is to improve the rules themselves.

Simplified SQC Process

Tag: Simplified SQC.
Version: October 7, 2004.
Owner: Tom@Gilb.com.
Status: Draft.

Entry Conditions

- A group of two, or more, suitable people* to carry out Simplified SQC is assembled in a meeting.
- These people have sufficient time to complete a Simplified SQC. Total elapsed time: 30 to 60 minutes.
- There is a trained SQC team leader at the meeting to manage the process.

Procedure

P1: Identify Checkers: Two people, maybe more, should be identified to carry out the checking.
P2: Select Rules: The group identifies about three rules to use for checking the specification. (My favorites are clarity ('clear enough to test'), unambiguous ('to the intended readership') and completeness ('compared to sources'). For requirements, I also use 'no design'.)
P3: Choose Sample(s): The group then selects sample(s) of about one page in length (300 non-commentary words). Choosing a page at random can add credibility – so long as it is representative of the content subject to quality control. The group should decide whether all the checkers should use the same sample or whether different samples are more appropriate.
P4: Instruct Checkers: The SQC team leader briefly instructs the checkers about the rules, the checking rate, and how to document any issues and determine if they are major defects (majors).
P5: Check Sample: The checkers use between 10 and 30 minutes to check their sample against the selected rules. Each checker should 'mark up' their copy of the document as they check (underlining issues and classifying them as 'major' or not). At the end of checking, each checker should count the number of 'possible majors' they have found in their page.
P6: Report Results: The checkers each report to the group their number of 'possible majors.' The SQC team leader leads a discussion to determine how many of the 'possible majors' are actually likely to be majors. Each checker determines their number of majors and reports it.
P7: Analyze Results: The SQC team leader extrapolates from the findings the number of majors in a single page (about 6 times** the most majors found by a single person, or alternatively 3 times the unique majors found by a 2 to 4 person team). This gives the major defect density. If using more than one sample, average the densities found by the group in different pages. The SQC team leader then multiplies this average major defects/page density by the total number of pages to get the total number of major defects in the specification (for dramatic effect!).
P8: Decide Action: If the number of majors/page found is a large one (ten majors or more), then there is little point in the group doing anything, except determining how they are going to get someone to write the specification properly. There is no economic point in looking at the other pages to find 'all the defects', or correcting the majors already found. There are too many majors not found.
P9: Suggest Cause: Choose any major defect and think for a minute why it happened. Then give a short sentence, or better still a few words, to capture your verdict.

Exit Conditions

- Exit if less than 5 majors/page extrapolated total density, or if an action plan to 'rewrite' has been agreed.

Notes:

* A suitable person is anyone, who can correctly interpret the rules and the concept of 'major.'

** Concerning the factor of multiplying by '6': We have found by experience (Gilb and Graham 1993: Bernard) that the total unique defects found by a team is approximately twice that of the number found by the person who finds the most defects in the team. We also find that inexperienced teams using Simplified SQC seem to have about one third effectiveness in identifying the major defects that are actually there. So $2 \times 3 = 6$ is the factor we use (or $3 \times$ the number of unique majors found by the team).

Simplified Specification Quality Control Form

SQC Date: **May 29, 200X**. SQC Start Time: _____
SQC Leader: **Tom**.
Author: **Tino**.
Other Checkers: **Artur**.
Specification Reference: **Test Plan**.
Specification Date and/or Version: **V 2** Total Physical Pages: **10**.
Sample Reference within Specification: **Page 3**.
Sample Size (Non commentary words): **approx. 300**.
Rules used for Checking: **Generic Rules, Test Plan Rules**.
Planned Exit Level (Majors/logical page): _____ or less.
Checking Time Planned (Minutes): **30**. Actual: **25**.
Checking Rate Planned (Non commentary pages/hour): **2**.
(*Note this rate should be less than 2 logical pages/hour*)
Actual Checking Rate (Non commentary words/minute): _____
Number of Defects Identified by each Checker:
Majors: **6, 8, 3**. Total Majors Identified in Sample: **17**.
Minors: **10, 15, 30**.
Estimated Unique Majors Found by Team: $\mathbf{16 \pm 5}$.
(Note 2 $\times$ highest number of Majors found by an individual checker)
Estimated Average Majors/Logical Page: $\mathbf{\sim 16 \pm 3 = 48}$.
(A Logical Page = 300 Non commentary words)
Majors in Relation to Exit Level: **48/1 (47 too many)**.
Estimated Total Majors in entire Specification: $\mathbf{48 \pm 10 = 480}$.
Recommendation for Specification (Exit/Rework/Rewrite): **No exit, redo and resubmit**.

_____

Suggested Causes (of defect level): **Author not familiar with rules**.

_____

Actions suggested to mitigate Causes: **Author studies rules, All authors given training in rules**.

_____

Person responsible for Action: **Project Manager**.
SQC End Time: **18:08**. Total Time taken for SQC: _____

Version: August 15, 2004. Owner: Tom@Gilb.com

2.4.6 Principles: SQC

1. **The Principle of 'Illegality'**
   'Defects' are *objective* violation of accepted written rules.
2. **The Principle of 'Majors are the pay off'**
   Major defects are the only *economically interesting* defects.
3. **The Principle of 'Keen to be seen clean'**
   The main purpose of SQC is to measure that the specification is clean enough: *not* to clean up a specification that *isn't*.
4. **The Principle of 'Cleanup your own mess'**
   Specification cleanup is the writer's responsibility, *before* SQC.
5. **The Principle of 'Prevention is better than cure'**
   There are many effects of SQC, but the most useful are learning to avoid defects caused by bad process, and committed by the writer.
6. **The Principle of '50% effectiveness'**
   History shows that you can only expect to find and fix about *half* the defects that are there.
7. **John Craven's Principle (within Hewlett Packard)**
   The team is there to make the "writer look like a *hero*."

8. **The Principle of 'Magnificent profitability'**
   The expected return on investment for SQC is at least 'ten to one'.
9. **The Principle of 'Client-server'**
   The writer is the *client* and the checkers serve as advisors.
10. **The Principle of 'The pilot in command'**
    The team leader is responsible for the SQC process.

Good execution of a badly executed specification will tend to execute you!

2.4.7 Additional ideas: SQC

There are some *central* ideas of SQC, which are worth looking at in more depth.

2.4.7.1 Economics of using SQC

The cost of finding and fixing defects has to be balanced against the benefit of removing the defects. The cost of fixing a defect escalates the longer it is left unfixed. In general, as we move from requirements/design-stages to test-stages, the total system-wide costs of *removing* major defects increase by an order-of-magnitude. As we move into design implementation, manufacturing, installation, servicing, and distribution, yet *another* additional order-of-magnitude of cost is generally our penalty for dealing with major defects later.

The cost of finding defects also varies. There are several QC nets that specifications pass through as a product is developed. Also, just because there is a defect doesn't mean that it will cause damage. This is where sampling and understanding your document quality level is essential.

If there is more than approximately one remaining major defect/page, it will tend to pay off to fight the defects immediately, using SQC, rather than downstream. With less than that, it probably pays off to let that major defect (exact location unknown) slip through *this particular* QC net, and hope it is still caught in some other QC net in the systems engineering process.

Unfortunately most real engineering environment 'approved' documents are at least one order-of-magnitude *worse* quality than one major defect/page: 10 – 20 or far more major defects/page is common, according to my frequent measurements. But without SQC, to measure for us, we don't 'know' this.

2.4.7.2 Effectiveness of SQC

If SQC is consistently carried out according to official guidelines (critically including the 'checking rate' being at the optimum level), then experience in IBM Rochester Labs, MN shows that the defect-finding *effectiveness is relatively stable*. Thirty percent effectiveness is a beginner's level (my experience). For a mature SQC process, effectiveness, for a single-pass attempt, tends to be in the range 60 – 90%, depending on the type of specification being checked. By systematic SQC process improvement, the effectiveness can slowly be improved to its maximum potential. Cumulative SQC effectiveness, for multiple passes, has been shown to reach a maximum of 95% (IBM UK and Sema UK Case).

2.4.7.3 Determining effectiveness and estimating remaining defects

SQC can be used directly to *measure* defects found and, indirectly to *estimate* the defects not found. Providing that 'effectiveness' (% of 100%) at finding defects is known and is relatively stable, it can be used to estimate the number of *unfound* defects. Effectiveness (of your teams on your specifications at finding defects) can be determined in two ways:

1. For a specific specification, we can measure the percentage of the 'available' defects, which a given SQC process was able to find. We can work out this percentage if we know the number of defects found by the SQC and the total number of defects found at later stages by other QC processes, testing, and field use. IBM has practiced this for decades in software engineering.
2. Another method, faster and cheaper, and more credible locally, is to repeat SQC on the same sample. If we find 30 defects on first attempt, fix them all, and hypothesize that we are 30% effective at finding them, this means we should have about 70 defects not found in our sample, right? If, after fixing all the 30 we found, the second SQC, done on the same sample, consistently finds about 21 defects (± about 6), it would confirm our prediction of 30% (21 of the 70 remaining from the first SQC). The '% of available defects found' is the effectiveness of the given SQC process. We use this method regularly on our training courses, and it works! It will also work for *any* test or QC process.

Once you have determined your effectiveness, you can estimate the remaining defects in a specification. We use the number of 'estimated probable remaining defects' to decide if a specification can exit (a typical exit condition is 'no more than one probable remaining major defect/page'). See Figure 2.4.7 for the calculation of 'estimated probable remaining defects'. We use effectiveness to determine the number of defects unidentified, and then

Estimating the Remaining Major Defect Density

Assumptions:
A logical page (page) is 300 non-commentary words.

- 30 major defects/page have been found during SQC.
- Your SQC effectiveness is 60% and your SQC is a statistically stable process.
- One sixth of your attempts to fix defects fail (One sixth is average failure to fix).
- New defects are injected during your attempts to fix defects at 5%.
- The uncertainty factor in the estimation of remaining defects is ±30%.

Probable remaining major defects/page = 'Probable unidentified majors' + 'Bad fix majors' + 'Majors injected'

Let E = Effectiveness expressed as a percentage (%) = 60%

Probable unidentified majors = major defects acknowledged-by-editor for each page at Edit × (100 − E) / E = 30 major defects/page found × (100 − 60) / 60 = **20 major defects/page**.

Bad fix majors = One sixth of fixed majors = So, of 30 attempted fixes, **5 major defects in each page** are not fixed.

Majors injected = 5% of majors attempted to be fixed = **1.5 major defects/page**.

Probable remaining major defects/page = 20 + 5 + 1.5 = 26.5 remaining major defects/page.

Taking into account the uncertainty factor of ±30% and rounding down to the nearest whole number gives **26 ± 7 Remaining Major Defects/Page**

(Minimum = 19, Maximum = 33 remaining major defects/page).

Fig. 2.4.7 Estimating remaining major defect density: the main specification exit condition.

we improve the accuracy by considering the effects of the editing. One sixth of fix attempts during editing fail (M. Fagan 1986:[7] IBM experience), unless an SQC for each fix is done to reduce fix failure. In addition, defect injection occurs during editing as a side effect of the fixes; the defect injection rate is sometimes 2 – 5% – but is highly variable and uncertain.

The final consideration is the uncertainty in the estimate. I have found that this remaining defect estimate is reasonably correct, and even in poor circumstances is ±30% uncertain, which is good enough for most purposes.

A specification can have 'too high a density' of major defects (equals serious engineering-cost rule violations) to be acceptable for use (to be allowed 'entry' or 'exit'). *'We will find it in test'* is a dangerous delay. Delaying *action* on your specifications major defects threatens not only cost (thus profit), but time-to-market and competitive quality. It pays off to deal with most major defects *early*.

2.4.7.4 SQC and rules

SQC is completely dependent on the rules that are applied. Just because you exit from an SQC process does not mean that all quality checking has been completed. It simply means that checking has been completed against the rules actually used, and has demonstrated an acceptable defect level.

By using different rules, different types of quality checking can be achieved. It is not simply a case of using the relevant rule set to match the type of specification. You need to consider what type of defects you are checking for and their potential cost if not detected.

2.4.7.5 Extending from SQC into specification review

There is no reason why the SQC method shouldn't be used to prepare for management reviews. You might have checked the content of a specification for consistency, completeness and clarity (Specification Rules).[8] But maybe you have not yet checked for the relevance to current business or technical demands. For example, maybe a certain level of ROI is demanded? Maybe a specific safety margin must be shown to be present? By creating a different set of rules, called Specification Review Rules, the SQC process can also be used to carry out pre-review quality control, in advance of a review meeting. This will probably result in better quality

[7] Fagan. M. E, 'Advances in Software Inspections', IEEE Transactions on Software Engineering. Volume SE-12, Number 7, Pages 744–751, July 1986.
[8] Note: This chapter mainly discusses and illustrates SQC from the viewpoint of checking for specification clarity, completeness, and consistency.

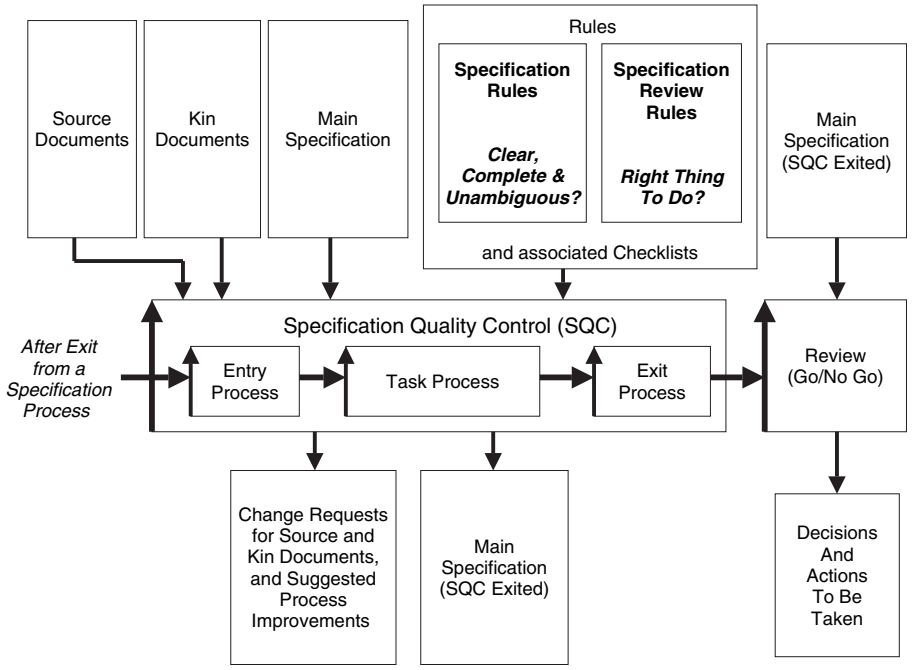


Fig. 2.4.8 Overview of the SQC process showing how Specification Review Rules fit alongside Specification Rules.

control than would be carried out in a hurried executive review meeting.

2.4.8 Further example/case study: a stitch in time saves nine

Trevor Reeve made use of SQC (at that time he called it 'Statistical Quality Control applied to Software and Documentation' or 'Documentation Quality Improvement') in all industrial aspects of a 1500 person defense electronics manufacturer (later a part of Racal). He documented four years of experience after this author ran a course on-site.

Reeve carried out an analysis of the first 1000 major defects logged by the SQC process to investigate the cost savings of using SQC. Test and field staff were asked when these defects would have been found in their test or field reports. They were also asked to indicate what the cost of finding and fixing them would be. The frequency curve in Figure 2.4.9 was drawn based on their answers. The mean time to correct these defects downstream would have been 9.3 hours. The mean time to find and fix them using SQC was about 1 hour. The defects would *otherwise* have been found by test and by customers. The result of this was that it was acknowledged by top management that removing a major defect using SQC gave a net saving of about 8 hours, or a 9.3 to 1 ratio of engineering hours 'return' on investment in SQC.

Compare this to the Raytheon return of 7.7 to 1. Six hundred inspections had been done at Thorn EMI by 1992, and over a thousand by early 1993. Savings were conservatively estimated at £500,000 each year, after one-time startup costs of £50,000. External consultants are said to have estimated real savings at double this figure. "Quality increased and development time has been reduced significantly."

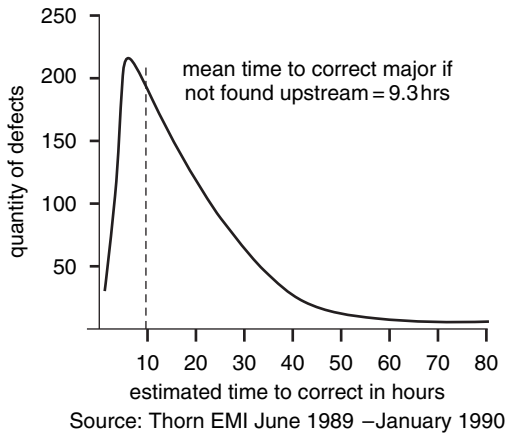


Fig. 2.4.9 MEL/Thorn EMI (later RACAL) UK, Factory and Lab-wide SQC gave order-of-magnitude savings. About 1000 major defects found by using SQC with multi-disciplinary teams were analyzed. The alternative cost to fix majors, if caught downstream, was nine times greater than if caught upstream by SQC. This is a frequency chart for the 1000 defects.

2.4.8.1 Use of SQC on many different types of document

SQC experience at Raytheon was limited to software, but at MEL/Thorn EMI, the documents

> *"ranged from system, hardware and software design documents to software code and change notes … test specifications, proposals, program management documents (for example, configuration and program plans), contracts and purchase specifications, printed circuit board design and test specifications, and procedures and standards.*
>
> *Further application (1993) all contractual documents, drawings and internal specifications (for example, information technology and financial requirements)… with all contracts using it to a lesser or greater degree by end 1992….*
>
> *The product appraisal process was revised to incorporate the technique into the normal activities performed by the organization on all types of document from contract to lowest level design and test specification including drawings." (Gilb and Graham 1993: P310)*

Note: The use of SQC on the upstream documents will produce the greatest benefits because defects will be caught earlier, and do less damage.

2.4.8.2 The organization must be supportive and SQC needs a champion

Since 1993, Trevor has confirmed many times within different organizations and various parts of the world, that two of the main factors for SQC to succeed are as follows:

1. An organization really needs to be willing to change.
2. The continuous presence of a totally committed champion of the method is necessary, for many years after the initial introduction of the method, to help the necessary culture change to take place. (This was also the experience in the same period of another client, Hewlett-Packard.)

2.4.9 Diagrams/icons: SQC

This section shows the SQC forms as follows:

- Figure 2.4.10: Master plan.
- Figure 2.4.11: Editor advice log.
- Figure 2.4.12: Data summary.
- Figure 2.4.13: Process meeting log.
- Figure 2.4.14: Simplified SQC form.

These are the blank versions of the filled-in forms given earlier in this chapter.

2.4.10 Summary: SQC

The basic ideas of SQC are simple:

- "A stitch in time saves nine": fix defects at *early* design stages (DDP), before they cause damage and/or require a costly 'defect removal' process, during test or operation.
- "An ounce of prevention is worth a pound of cure": learn from defects, which have common underlying causes, and continuously improve your work processes (DPP).

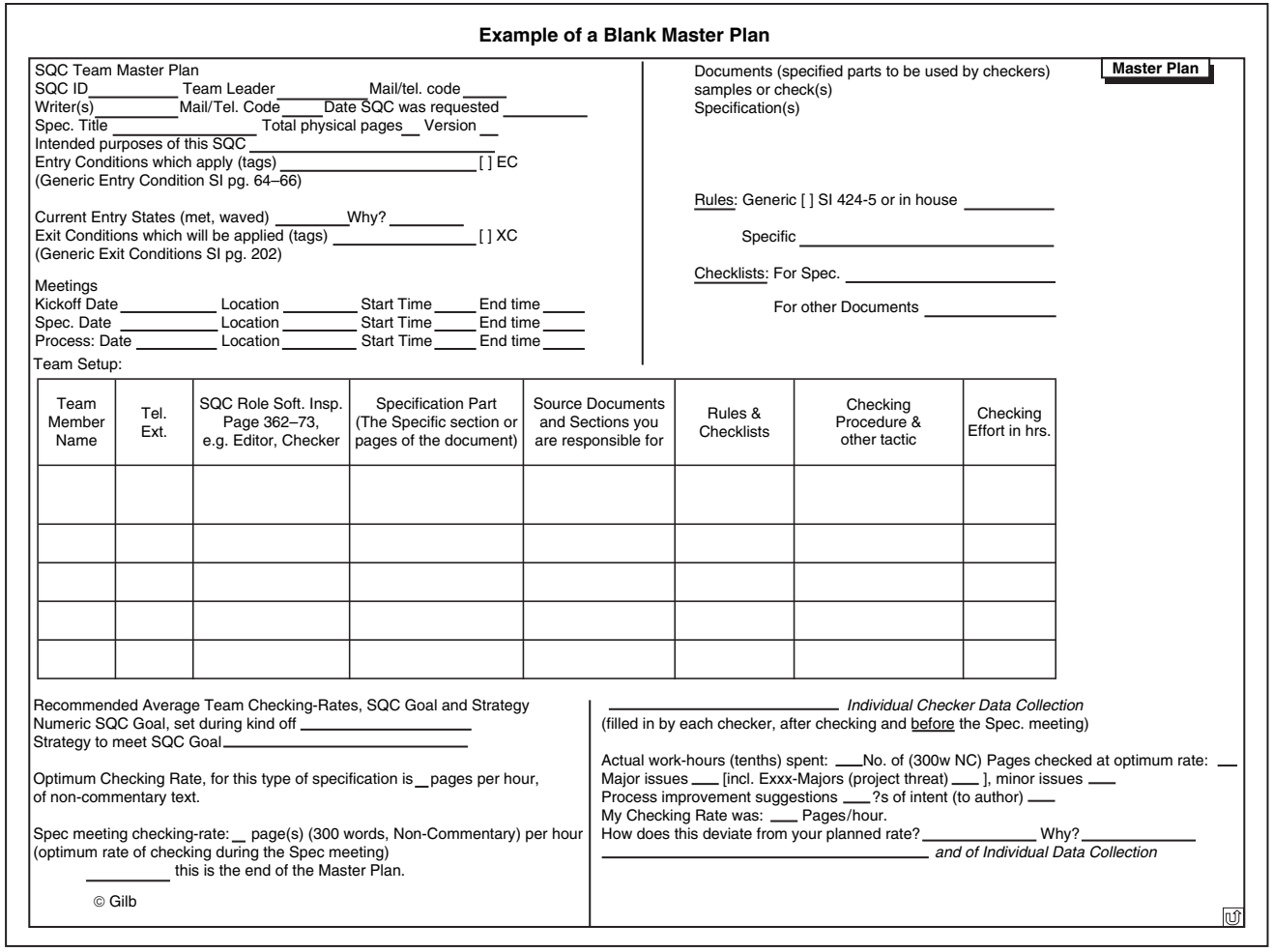
2.4.10.1 Finding defects

The DDP is more powerful than similar processes, such as 'checking'[9] (of engineering drawings, as proven at Douglas Aircraft 1988, Boeing 1989, and Thorn EMI 1990 on), 'reviews', 'walkthroughs', meetings and management approval. This is mostly due to a series of tactics, the most critical of which are probably the use of a proven *optimum defect-searching time* (optimum checking rate) and, the total focus *on finding and fixing 'major' defects* (which saves time downstream).

2.4.10.2 Understanding document quality

One of the most important opportunities using SQC is to be able to measure the degree to which systems engineering and management documents of all types really do correspond to the required standards of

[9] Do not confuse with the SQC 'Checking' sub-process! The aircraft factory traditionally used the term 'checking' for a process done by a group of people who specialized in this, called 'checkers'. The process checked engineering drawings against the official engineering drawing specification rules, which were in a large handbook – so large that copies were not given to inform individual engineers what the rules were! In 1988 we proved, with hard data on a large scale, for the engineering directors, that the SQC process was far more effective at finding interesting engineering defects than the traditional 'checking' process. We ended up within the first year with sixty times better quality in terms of rejected and reworked drawings (0.5% versus earlier about 30% reworked).



Example of a Blank Master Plan

SQC Team Master Plan
SQC ID_____ Team Leader _____Mail/tel. code _____
Writer(s)_____Mail/Tel. Code_____Date SQC was requested _____
Spec. Title _____ Total physical pages__ Version __
Intended purposes of this SQC _____
Entry Conditions which apply (tags) _____[ ] EC
(Generic Entry Condition SI pg. 64–66)

Current Entry States (met, waved) _____Why? _____
Exit Conditions which will be applied (tags) _____[ ] XC
(Generic Exit Conditions SI pg. 202)

Meetings
Kickoff Date _____ Location _____ Start Time _____ End time _____
Spec. Date _____ Location _____ Start Time _____ End time _____
Process: Date _____ Location _____ Start Time _____ End time _____

Documents (specified parts to be used by checkers)
samples or check(s)
Specification(s)

Master Plan

Rules: Generic [ ] SI 424-5 or in house _____

Specific _____

Checklists: For Spec. _____

For other Documents _____

Team Setup:

Team Member Name	Tel. Ext.	SQC Role Soft. Insp. Page 362–73, e.g. Editor, Checker	Specification Part (The Specific section or pages of the document)	Source Documents and Sections you are responsible for	Rules & Checklists	Checking Procedure & other tactic	Checking Effort in hrs.

Recommended Average Team Checking-Rates, SQC Goal and Strategy
Numeric SQC Goal, set during kind off _____
Strategy to meet SQC Goal _____

Optimum Checking Rate, for this type of specification is __ pages per hour, of non-commentary text.

Spec meeting checking-rate: __ page(s) (300 words, Non-Commentary) per hour (optimum rate of checking during the Spec meeting)

_____ this is the end of the Master Plan.

© Gilb

_____ *Individual Checker Data Collection*
(filled in by each checker, after checking and before the Spec. meeting)

Actual work-hours (tenths) spent: ___No. of (300w NC) Pages checked at optimum rate: ___
Major issues ___ [incl. Exxx-Majors (project threat) ___], minor issues ___
Process improvement suggestions ___?s of intent (to author) ___
My Checking Rate was: ___ Pages/hour.
How does this deviate from your planned rate?_____ Why?_____
_____ *and of Individual Data Collection*

Fig. 2.4.10 Blank master plan.

Example of a Blank Editor Advice Log

Date _____ Start time _____ End time _____ Editor Advice Log

SQC ID_____ Page __of ____

Item No.	Document Reference Tag	Page	Line or Tag	Exact Location	Type of Item	Checklist or Rule Tag	Source Inconsistency and/or Necessary Description	Occurs	Editor Action (during editing)
1					Major Minor ? Imp. New				
2					Major Minor ? Imp. New				
3					Major Minor ? Imp. New				
4					Major Minor ? Imp. New				
5					Major Minor ? Imp. New				
6					Major Minor ? Imp. New				
7					Major Minor ? Imp. New				
8					Major Minor ? Imp. New				
9					Major Minor ? Imp. New				
10					Major Minor ? Imp. New				

Subtotals:
New Items found during the Spec. Meeting ___
Major issues logged. Minor issues logged _____ Improvement suggestions logged___. ? Questions of Intent logged ___.

© Gilb

Fig. 2.4.11 Blank editor advice log.

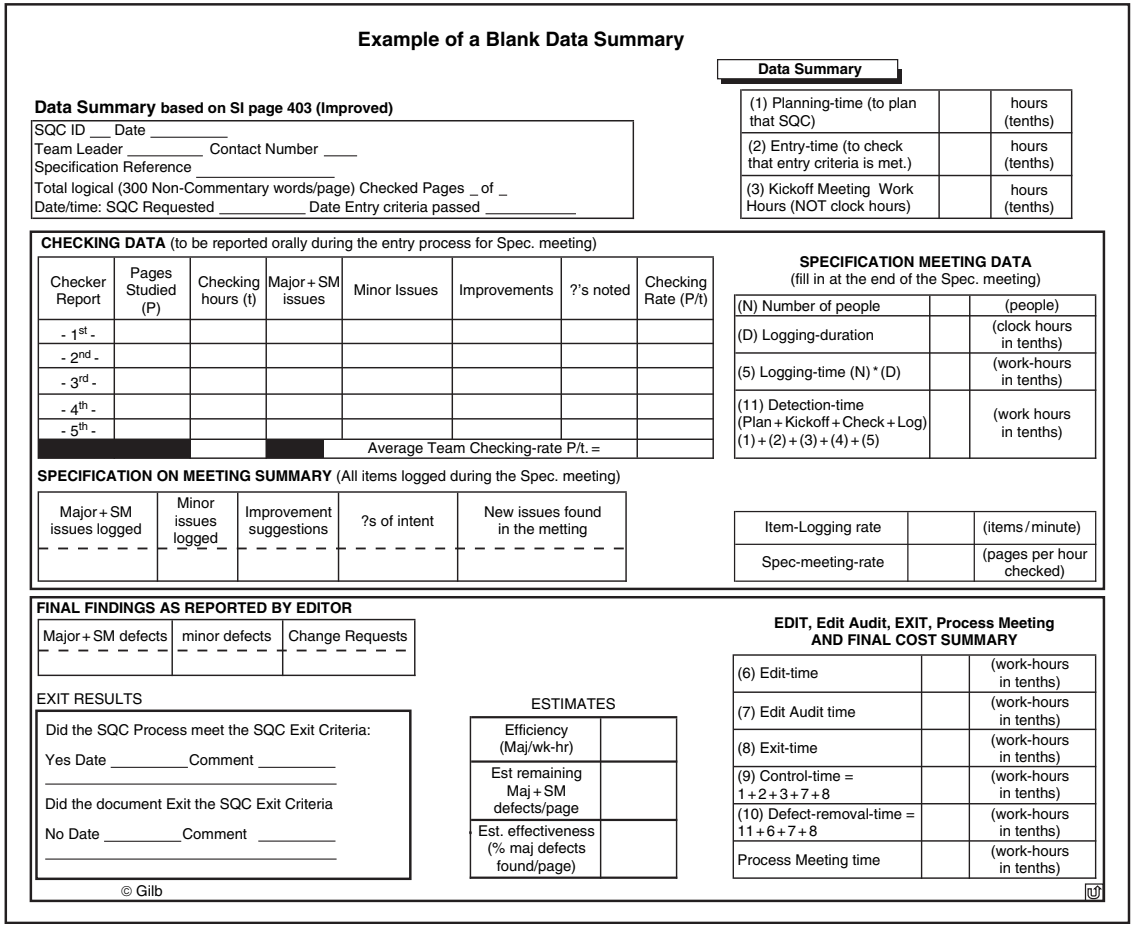


Fig. 2.4.12 Blank data summary.

Process Meeting Log

Team Leader _____ Date _____ SQC ID __ Page _ of _

Item	Issue Reference	Cause Class (tick 1)	Root Cause Ideas	Improvement Ideas
1		Communication Oversight Transcription Education		
2		Communication Oversight Transcription Education		
3		Communication Oversight Transcription Education		
4		Communication Oversight Transcription Education		
5		Communication Oversight Transcription Education		
6		Communication Oversight Transcription Education		
7		Communication Oversight Transcription Education		
8		Communication Oversight Transcription Education		
9		Communication Oversight Transcription Education		
10		Communication Oversight Transcription Education		

© Gilb

Start Time	Stop Time	Duration	No. People	Total Cost

Fig. 2.4.13 Blank process meeting log.

quality. The concepts of 'entry' to, and 'exit' from, all systems engineering and management processes are enabled by our ability to measure 'probable remaining major defects/page' and to decide by estimation if a specification is economic enough to release downstream ('exit'), or economic enough to start work on (allow 'entry'). If necessary, sampling of large documents is an economic way to measure quality levels before making decisions of consequence concerning those documents.

The fact that the SQC process is universally applicable to any readable specification (in any intellectual, administrative, project management, planning, systems engineering, software, or user specification task), means that any group of people can use it wherever they want control over quality-in-

Simplified Specification Quality Control (SQC) Form

SQC Date: _____ SQC Start Time: _____
SQC Leader: _____
Author: _____
Other Checkers: _____

Specification Reference: _____
Specification Date and/or Version: _____ Total Physical Pages: _____

Sample Reference within Specification: _____
Sample Size (Non commentary words): _____

Rules used for Checking: _____
Planned Exit Level (Majors/logical page): _____ or less.

Checking Time Planned (Minutes): _____ Actual: _____
Checking Rate Planned (Non commentary words/minute): _____
(Note this rate should be less than 2 logical pages/hour)

Actual Checking Rate (Non commentary words/minute): _____
Number of Defects Identified by each Checker:
Majors: _____ Total Majors Identified in Sample: _____
Minors: _____

Estimated Unique Majors Found by Team: _____ ± _____
(Note 2 × highest number of Majors found by an individual checker)

Estimated Average Majors/Logical Page: _____ (*A Logical Page = 300 Non commentary words*)
Majors in Relation to Exit Level: _____
Estimated Total Majors in entire Specification: _____
Recommendation for Specification (*Exit/Rework/Rewrite*):
_____

Suggested Causes (of defect level):
_____

Actions suggested to mitigate Causes:
_____

Person responsible for Action: _____

SQC End Time: _____ Total Time taken for SQC: _____

Version: August 15, 2004. Owner: Tom@Gilb.com

Fig. 2.4.14 Simplified SQC form.

relation-to-standards. However, SQC has some limitations in understanding 'how well' specifications will work in practice. Even if specifications exit according to any rules you use to analyze them, there can still be catastrophic defects in them in practice. So, we need to use additional methods to see 'how good' a specification is and, if necessary, adjust the specification.

The SQC ability to measure quality, in relation to standards, is also important when the standards are a major part of continuous process improvement. We can use SQC to measure process improvement efforts! The measurement of defects is a measure of whether people have actually learned, practiced, and understood the continuous improvements intended to increase productivity.

2.4.10.3 Continuous process improvement

The DPP exploits grass-root everyday experience with DDP, as well as making use of your data about defects from 'test' and 'field' situations. DPP is the 'engineering and management' version of what Deming in 1993 and Juran in 1974 taught manufacturing industry, starting in Japan. Experience shows that 40% annual productivity improvements are possible, when this is done properly (which is rare).

Possible Purposes For Using SQC

-Reducing Time-to-Delivery
-Measuring the Intelligibility of a Document
-Measuring the effectiveness of engineering specifications
-Measuring the ability of the Process producing the Document to follow best practice rules.
-Enabling Estimation of the Number of Remaining Major Defects
-Identifying Major Defects
-Removing Major Defects
-Preventing consequential 'Downstream' Defects being generated by removing existing Defects
-Improving the Engineering Specification Process (better standards, like rules)
-Improving the SQC Process (better rates, better entry exit conditions, better procedures)
-On-the-Job Training for the Checkers
-Training the Team Leader
-Certifying the Team Leader
-Peer Motivation (getting people to learn, and follow the rules)
-Motivating the Managers (to deal with problems early)
-Helping the Writer (learn to write clearly and have effective ideas)
-Reinforcing Conformance to Standards
-Capturing and Re-using Expert Knowledge (by use of Rules and Checklists)
-Reducing Costs
-Team Building
-Fun – a Social Occasion

Fig. 2.4.15 Possible purposes for using SQC. Any one or several can apply at anytime.

Section **Three**

Product development

This page is intentionally left blank

Chapter 3.1

3.1

The business objective

Marc Annacchino

3.1.1 Background

3.1.1.1 New product development and the economy

New product development is an integral part of a healthy, growing economy. This chapter will start out with a review of the various means for economic development, including manufacturing and how product development plays a role. The various types of product development and how they draw on and pay out to the economy are also included. In addition, we will look at our world and what it would be like without some famous new product developments. Finally, there is a review of what happens to a company without new product development, from a purely financial perspective.

3.1.1.1.1 Economic development

New product development contributes to the economy by generating revenue and profits to a corporation that otherwise would not have been generated. The revenue then is paid out to vendors (other manufacturers). The vendors themselves pay out to their sub-vendors and personnel, or retain the earnings. Salaries are paid to personnel and they, in turn, spend funds to purchase goods and services from other profit-making enterprises. The retained earnings fund the long-term growth of the enterprise and increase the value of the business. The profits are taxed and that goes into the pool of funds to govern the community and provide for the common good.

This is the role that private investment plays in the economic development arena. The public sector contributes to the economic development by providing incentives to encourage manufacturers to establish their businesses in their locale. They provide the means for funding business expansion and growth. A collateral activity is to network with other manufacturers on your behalf for future business. Figure 3.1.1 illustrates the typical funds flow in a manufacturing enterprise. As will be seen later, the service sector of the economy has a funds flow that is different from that in manufacturing.

Service businesses are not characterized by the leverage associated with manufacturers. Instead, they are characterized by a smaller investment in capital equipment, a smaller investment in each revenue cycle, and a generally faster revenue cycle; as such, they generate incremental profits from incremental investment with low fixed costs. This is shown in Figure 3.1.2.

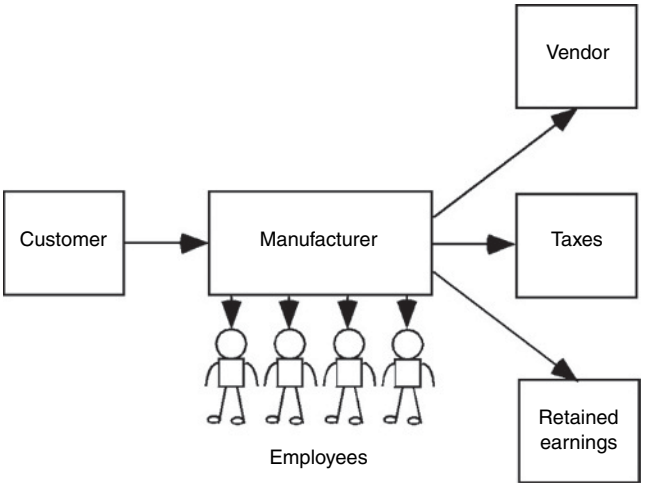


Fig. 3.1.1 Manufacturer's funds flow model.

The Pursuit of New Product Development; ISBN: 9780750679930
Copyright © 2007 Elsevier Inc. All rights of reproduction, in any form, reserved.

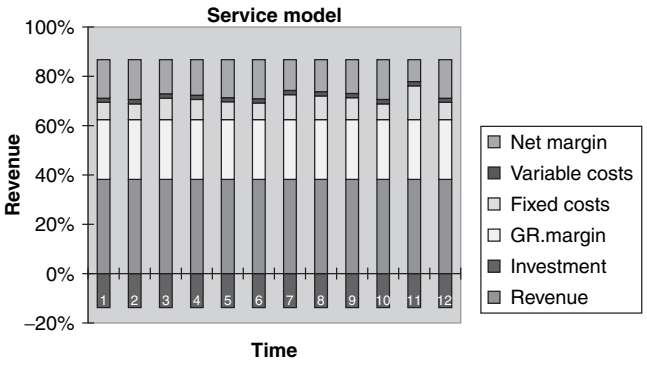


Fig. 3.1.2 Service business funds flow model.

As shown, the service model has relatively constant net profit. This is because there is no major initial investment to absorb and there is a small incremental investment with each order. The other costs are somewhat variable and track with the incoming order rate. Incremental profits can track with incremental revenues. Service businesses generally have a lower barrier to entry than do their manufacturing counterparts.

Manufacturing businesses differ markedly from service businesses in that leverage can occur. This leverage is characterized by larger returns for their investment. A larger investment is required at the outset, but revenue is generated through a stream of returns from year to year through the products' life cycles. This assumes there is sufficient volume to offset the fixed expenses and the absorption of the initial investment. There is significant investment in capital equipment and processes, along with significant payoff if you have hit on the right opportunity with the right product.

The manufacturing financial model (Figure 3.1.3) shows that net margin increases at a disproportionate rate as revenue increases. This is because the fixed costs are already absorbed by some base level of business and the incremental revenue does not require incremental investment with each order. Fixed costs decrease as a percent with increasing revenue.

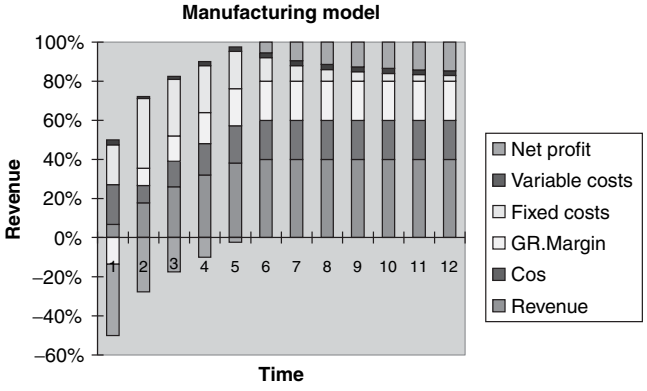


Fig. 3.1.3 Manufacturing financial model.

However, there is a large initial investment to design the product, which can run the length of the diagram in time. Figure 3.1.4 illustrates how a manufacturing enterprise's funds flow as a result of new product development.

An idea is conceived and qualified in terms of opportunity and overall business sense. There are costs associated with these activities, such as market planning costs, surveys, customer visits, and demographic data analysis. There are also costs in taking the opportunity, feeding back the market data, and coalescing them into a product opportunity. The scope of the target market is accounted for in this stage, and the product platform must be laid out to reach the market at a cost-effective price. These costs are generally low compared to the other costs in the product development arena.

The next phase is the investment and development phase of the program. This phase takes the product concept and creates the intellectual property required to take a concept and reduce it to bills of material, manufacturing processes, and define a manufacturable product that the market will purchase. There are technical, labor, development, tooling, and other capital equipment costs in this phase.

The next phase in the model assumes that all the development is complete and manufacturing takes place. Here, there are set-up, material, labor, and overhead costs required to produce the units. The product moves out of manufacturing and into the sales channel for placement at the customer. In this stage there are a host of sales, general, and administrative (SG&A) costs incurred.

Finally, the product is at the customer and funds for the product have changed hands. The manufacturer now can begin to enjoy the profits. These profits come after the funds are distributed appropriately. Dollars must go back to fund the sales expenses, general expenses, and any other expenses previously incurred. Manufacturing must be reimbursed for the materials labor, manufacturing expenses, and inventory carrying costs.

The development of the product must be amortized so as to be able to fund future developments. In addition, funds must also be channeled back to the product planning and market planning function to allow the proper expenditures to verify product viability, market program viability, and future enhancements and products.

3.1.1.1.2 Types of new product development and their contribution

There is a variety of different types of new product developments in existence today. Each is used for a different reason, and each has its own objectives and dynamics for execution. The following is a common list of the different types and their attributes and contributions.

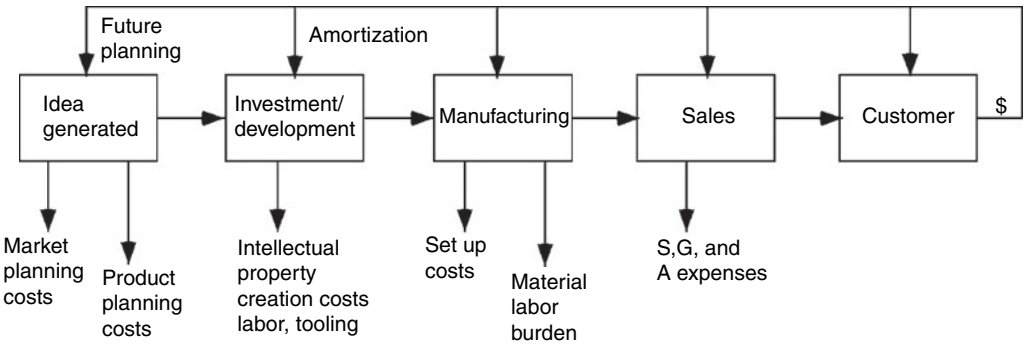


Fig. 3.1.4 Manufacturing funds flow as a result of new product development.

3.1.1.1.2.1 "New to the world" products

These are somewhat revolutionary in the marketplace in that the marketplace never had exposure to the product directly—perhaps only as a concept or prediction from a futurist. They generally create entire new markets that never existed before. An example would be the cellular telephone. Predicted only vaguely by the Dick Tracy cartoon strip and personified as a "communicator" in Star Trek, the cellular telephone has revolutionized person-to-person communications in modern-day society. One such product looming on the 10-year horizon today is mass-market fuel cells.

These product development programs generate entire new markets that were not previously there. They enable true growth in the economy by generating revenue to the enterprise. They also have a multiplication effect in the economy by generating requirements for parts and subassemblies that need to be developed and supplied by the vendors. In many cases, they generate new channels of sales and new routes to market.

3.1.1.1.2.2 New product lines

These new categories of products allow entry into newer markets not previously participated in by the manufacturer. By adding the categories, manufacturers must be careful to protect the positioning of their existing products, which generate the existing business. Failure to do so will place them in danger of converting loyal customers away from one already successful product to a new one, with no net gain in market share. Perhaps a good example of this type of product would be the Hyundai Azera, a higher-end automobile offering from Hyundai. Here a large manufacturer with many product lines generated an entirely new category of car to serve a more discerning customer base. Careful not to jeopardize their existing base, this new car company initially launched its product in the same value-focused dealerships selling the existing products. This is different from the Acura offering distancing itself from Honda.

The new product lines generate incremental revenue to the manufacturer by leveraging the market's familiarity with the manufacturer into new categories of products. In many cases, the market's familiarity with the manufacturer paves the way for new categories of products. Sometimes these products go into new markets, but can also be an alternative to existing ones.

3.1.1.1.2.3 Additions to existing product lines

These efforts support existing product lines by creating line completers to extend the influence of the original products' brand to larger audiences or extending range, power, and scope. All are done in the attempt to secure more of a market. An example of this type of product would be the M&M Candy Company extending their product line to M&M peanut and M&M almond and seasonal M&Ms for Christmas and Easter. Another example would be tomato sauce versions—hearty, traditional, roasted garlic, Alfredo, and vodka tomato. By taking the basic product and modifying it, a wider market share is realized.

The addition to existing lines has a similar effect on the company's revenue as the new product lines. They generate incremental revenue by leveraging the existing product familiarity rather than the company familiarity. These programs generate incremental improvement in the economy, but generally fall short of the contribution made by the totally new products.

3.1.1.1.2.4 Improvements and revisions of existing products

As time marches on, customers have higher expectations of your product and the competition adds features to their offering. It becomes necessary to improve your company's offering to increase market share or to retain it. By redesigning the product or repackaging it, your company can offer a greater value or satisfaction to the customer. It is possible to temporarily affect this by enhancing perceived value; however, an ever-more-informed customer base will respond to actual value increased in the long run. An example of this type of product development is the automotive companies adding features to their base models each year as standard.

Generally, the improvements to existing products do not generate additional revenue to speak of. They are simply a means to retain the market share or to slightly improve it. They are defensive in nature and in many cases are stopgap measures until a new product program can be introduced. These programs do little to generate a vitalized economy in the long run, but can provide time and revenue to pursue the development of a replacement.

3.1.1.1.2.5 Repositioning

Another means of increasing or maintaining market share is through repositioning. A repositioning is an exercise in changing the perception in the mind of the consumer. It generally can happen with products that are lower in value (dollar amount), or the consumer spends little time evaluating the actual data. For high-dollar decisions, the consumer will generally take the time to evaluate the facts and make his or her own decision. Repositioning is truly a marketing activity rather than a development activity. An example of this is a change of advertising by focusing the audience on a possible linkage drawn between certain brands of cereal and a high-fiber, lower-cancer-risk diet.

Repositioning is another stopgap measure for generating revenue from an existing product. It does not generate overall growth in the economy per se; rather, it is similar to an improvement or a revision except that it doesn't even necessarily require a product change. It simply repositions the product in the mind of the consumer.

3.1.1.1.2.6 Cost reductions

These programs are strictly a means for reducing the cost of products to offer similar value. They generally are the result of a competitive initiative, either generated internally or from external forces. In many cases, it is simply a means to generate more volume, which will generate less incremental profit (but perhaps more overall profit). Whatever the motive, a cost reduction is generally meant to increase unit volume through the channel. This becomes easier with capital equipment costs and development costs absorbed, and the manufacturer wants to capitalize on the sales channel.

Cost reductions are helpful to the organization by generating additional margin from the existing product. This margin can absorb development costs and manufacturing set-up costs. In many cases, they enable a period of time to continue with the product, generate the revenue, and allow the organization to position itself with a new product. They do not, however, generate any real growth in the economy.

Figure 3.1.5 summarizes the different types of product developments in terms of (a) time required to develop; (b) the revenue to the economy; (c) the revenue to the company; (d) the company's positioning; and (d) the margin impact.

3.1.1.1.3 Narrative and financial review of a nondeveloper contribution

Figure 3.1.6 illustrates the dynamics of the income statement of a company that has little new product development. As shown in the financial analysis, a company can grind to a slow halt by not participating in the dynamics of the ever-changing marketplace.

Along the top are the years under study. Along the left side are the income statement categories. The following summary discusses the perspectives and how they affect the overall operation.

3.1.1.1.3.1 Cost of sales

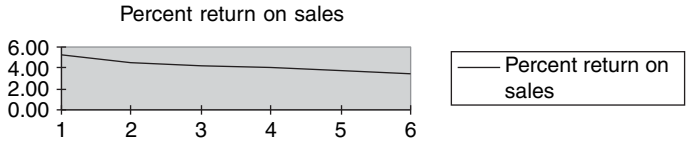
The cost of sales has a natural tendency to increase over time. This is a result of the increases in direct labor costs in manufacturing the unit as well as the effect of vendor increases in pricing. For most manufacturers, the product maintenance function of development has to initiate cost reduction wherever possible, just to stay even. Therefore, to progress down the learning curve, significant initiatives must be made to effect cost reductions because these changes must offset the increases already built-in.

Type of development	Time to introduce	Potential revenue contribution to economy	Revenue contribution to company	Company positioning strategy	Potential margin impact
New to the world	Longest	Highest potential	Highest potential	Market development	Highest
New product lines	Long	High potential	High potential	Market development	High
Add to existing	Medium	Medium potential	Medium potential	Line complete	Medium
Improve or revise	Short	Little potential	Medium potential	Market share	Medium
Repositioning	Shortest	Little potential	Medium potential	Market share	Medium
Cost reductions	Shorter	Little potential	Medium potential	Raise margin	Medium

Fig. 3.1.5 Types of product development programs.

	Base Year ($)	Year 1 ($)	Year 2 ($)	Year 3 ($)	Year 4 ($)	Year 5 ($)
Net revenue	100,000	101,000	102,000	103,000	104,000	105,000
Material	45,000	45,950	46,400	46,850	47,300	47,750
Labor	10,000	10,600	10,700	10,800	10,900	11,000
Burden	25,000	25,250	25,500	25,750	26,000	26,250
Total cost of sales	80,000	81,800	82,600	83,400	84,200	85,000
Gross profit	20,000	19,200	19,400	19,600	19,800	20,000
Engineering expenses	4,000	4,200	4,410	4,630	4,862	5,105
Administrative expenses	3,500	3,675	3,858	4,051	4,254	4,467
Sales expenses	3,000	3,150	3,307	3,472	3,646	3,828
Total expenses	10,500	11,025	11,576	12,155	12,762	13,401
Net profit	9,500	8,175	7,823	7,444	7,037	6,599
Provision for tax	4,275	3,678	3,520	3,350	3,166	2,969
Net profit after tax	5,225	4,496	4,303	4,094	3,870	3.629
Percent return on sales	5.23	4.45	4.22	3.98	3.72	3.46

A



B

Fig. 3.1.6 (A) Income statement, nondeveloper; (B) percent return on sales.

3.1.1.1.3.2 Profits

Assuming the expenses stay at a constant amount or grow slightly, this increase in materials and labor and manufacturing expenses (burden) creates enormous offsets in the net profit of the company. Salaries and benefit expenses increase with time because the manufacturer must remain competitive with the other, more dynamic companies in the industry. Sales expenses will increase as more concessions, incentives, or travel must be done to achieve the same result from year to year.

3.1.1.1.3.3 Strengths

The financial strength of a company degrades with a lack of successful new product development. Unfortunately, the effect is not known immediately. The financial picture initially looks better because the expenditures normally spent on a development are saved. This is, however, a short-term gain, because the market will soon bypass the manufacturer with new products from other firms, leaving the manufacturer with little to sell in the evolved marketplace. In later analysis, the strength of the enterprise diminishes.

3.1.1.1.3.4 Market share

If the manufacturer cannot keep up with the marketplace's appetite, eventually the package of values it offers does not meet the customers' needs and it begins to lose market share. By losing the revenue from lost market share, a dangerous downward spiral begins that is difficult to reverse.

3.1.1.1.3.5 Response to changes in environment

If the manufacturer cannot keep up the pace established by its industry by evolving product lines on a normal evolutionary path, it is at great risk from the negative effects of changes in the playing field, or legislation changes, or salient attacks from nontraditional competitors. This weakness will manifest itself as a financial problem over time, which will eventually place limitations on the organization's ability to respond in order to survive.

3.1.1.1.4 A look at our world without some new product developments

3.1.1.1.4.1 A world without Apollo

Not all new product development comes from private investment. The U.S. government regularly conducts research in a variety of areas in order to generate understanding, research, and analysis of data. One of the more scrutinized and high-profile development programs in the 1960s was the project to land a man on the moon and safely return him to Earth.

Aside from the sheer magnitude of the program, the raw objective was to develop methods, produce machinery and hardware, and execute a mission to land and return. The task represented enormous technological hurdles in numerous disciplines. The achievement of this objective in July 1969, must be historically underscored by the numerous spin-off technologies that have changed our world.

The Apollo program accelerated the development of technology in an unprecedented manner. The need drove the course of action: If it existed, use it; if it didn't exist, invent it. Time was literally of the essence. Money was no object, and resources were granted on the basis of how much time it would save, not how much it cost.

There were several technologies that experienced accelerated development during that era. The term accelerated is the operative one in this case. Most technology eventually gets developed as people become interested in its potential and entrepreneurship drives these elements into society. However, the Apollo program literally rocketed development of certain technologies that would have taken many more years to implement without the driving requirement that necessitated its invention.

Rather than listing and discussing each one, it may be more effective and interesting to look at our technological and societal world without the Apollo program. Let's examine the latter half of the 1990s as we look at "a world without Apollo."

You wake up at 6:00 A.M. The mechanical radio alarm clock clicks on with the local news. The clock is mechanical because a digital alarm clock is too expensive to build out of discrete transistors. The integrated circuit showed a lot of promise in the 1960s as an idea but was never funded, and so its widespread use is just beginning. However, that is not on your mind.

What is on your mind is the temperature outside and the chill you feel inside. There's a 20°F cold snap outside. Back in the 1970s, oil prices skyrocketed, and so Americans living in North America needed to hold down expensive oil and natural gas usage. Your home is insulated with the 1960s technology and cannot hold in the heat very well. The less-than-optimal insulation used is due to lack of funding to develop the new technology in the 1960s. The Apollo program hardware utilized specialized insulation and polymerized film materials, and solved its heat retention and heat rejection problems by the use of advanced insulation systems.

You are a jet aircraft pilot and have to fly to two cities and return home early this evening. You are planning to go to the baseball game tonight to see your favorite team play, so you hope it warms up later in the day, upon your return.

Today, your children are going to perform in the school play and your wife is planning to go food shopping and investigate a new hairstyle at the beauty shop.

You are pressed for time because you forgot to buy movie film to record the kids' school play. You can only film three minutes of the play and have no means for sound recording to coordinate with the movie film. The video camera has not been invented yet. Because there was no need to develop the technology for portable television cameras used on the Apollo missions, there was no spin-off to personal use.

When your wife gets to the food store and is ready to check out, there is a long line at the checkout. Because all groceries are checked out by hand, it takes longer to process the customers. Bar code scanners haven't been invented yet, nor is the raw technology in a cost-effective package for use in the commercial sector because integrated circuitry has not progressed far enough down the learning curve, so she waits.

Now, it's off to the beauty shop to experiment with that new hairstyle. Your wife will be taking a significant risk in experimenting because she will have her hair cut very short for the new hairstyle. In addition, the new style requires a permanent-wave treatment. Both items are a one-way commitment to the new style for two to three months. The image processing technology developed for the Apollo program has not yet been invented, since there was no need. Consequently, no spin-off technology allowed the beauty shops a means for taking your wife's photo and digitizing it to try out different hairstyles before cutting or waving.

Your wife now starts to plan the budget for the next month. Unfortunately, she is calculating and adding bills manually and makes a mistake, allowing more expenses than can be covered by the checking account's current balance. The handheld calculator has not been invented yet. There has been no market need for integrated circuits or cost reductions, so there are no calculators yet. Eventually, you will have one as the world progresses into the acceptance and the implementation of these technologies at the normal, nonaccelerated pace.

Meanwhile, you are at the airport and ready to start your day. You climb into the flight deck and see the traditional stick and rudder used to fly the airplane. There are very few avionics and no inertial navigation system. Everything in terms of flying the plane is done by manual calculation. There are no major systems, such as

navigation or communications, in place. The Carousel system for inertial navigation and fly-by-wire technology have not been invented or implemented in the fleet of planes yet. There was no requirement for inertial navigation for long distances like in the Apollo program, so it never made its way into the commercial marketplace. This system, combining computer technology, electromechanical physics, precision machining, inertial navigation algorithms, and a worldwide infrastructure, will be available through natural, incremental evolution around the time you plan to retire.

You land in the first city of your journey and listen to the television in the airport terminal. There is a news story about an entire family who perished in a home fire. How sad and tragic, you think, visualizing your own family. This family, like many others, died because the fire consumed them while they were asleep. There was no means for early fire detection or warning; the smoke detector was never invented for home use. The Skylab program in the early 1970s used mostly Apollo hardware, and the National Aeronautics and Space Administration (NASA) needed a means for early detection of smoke and fire for America's first space station.

Your attention shifts back to the weather, and you are thinking about that game tonight and want to know whether it will be too cold to go. You do not have any way of planning for the weather because modern forecasting methods are limited to predictions only—satellite photos using image enhancement have not yet been invented. There was no requirement for it on Apollo, so no spin-off.

Your children go to the school library to do research for a term paper. They comb through the card catalog and cannot find any article or reference material for their papers. The library's focus has not yet shifted from collection of books to access of information.

Your kids will have to wait for the reference material to be sent through the mail because there is no Internet with universal access yet, no low-cost computers affordable enough for the municipal budget, and no network of other users.

You finish your flights, arrive back home, and plan to go to the game. Your wife asks you to hang a new picture she bought today. You go down in the basement to get your electric drill to drill the hole required for the anchor. Unfortunately, you loaned the extension cord to a neighbor. You have no means for electrical power in the area where you will place the picture—cordless tools have not been invented yet. They will be an eventual spin-off from the drill used to core lunar samples on the Apollo missions. You will have to wait until tomorrow for your neighbor to return the extension cord.

You are off to the game at last. It never did warm up, nor did you ever get any advanced warning. It would be nice on a night unseasonably cold as this to have a cover over the stadium to retain the heat. Traditional roofs over stadiums are expensive and require huge means for support. Lighter, more affordable fabric used in domed roofs for stadiums and airports hasn't been developed yet. (They use spin-off fabric technology from the Apollo Moon suit developed for the astronauts to traverse the lunar surface.) It's so cold in the stadium that you get some hot coffee rather than a cold beer!

This short vignette represents only a few technologies developed for the space program that are in the everyday products we now take for granted. It has been often said that the space program was wasteful in terms of spending the funds on space travel rather than "curing" the problems on Earth. By conservative estimates, there are more than 30,000 products developed and launched as a result of the space program.

The reality of these results is that the tax on profits alone from the revenues generated from these products and technologies has paid for the government's investment in the entire Apollo program many times over each year.

Here are a few examples where continued product development has benefited the customers with better, safer products:

Automobiles: The Model T progressed to the Lincoln Town Car.

Aircraft: The Ford Tri Motor progressed to the Boeing 777.

Computers: The Electronic Numerical Integrator and Computer (ENIAC) progressed to the Cray supercomputers in technological integration and their pervasive use in numerous mass market products.

Vacuum tubes: The vacuum tube amplifier progressed through discrete transistors to integrated circuit amplifiers on a chip.

Incandescent lights: The original incandescent light spawned fluorescent tubes and bulbs with numerous multiples of extended life. Fluorescent technology applied to bulbs using embedded electronics enable fluorescent light in a standard bulb socket.

The importance of continuing the product life cycle preserves the customer base. It drives competition between manufacturers in the marketplace. Competition is desirable and necessary. It prompts everyone to action to improve the products that they offer, so the customer and the manufacturer both benefit.

3.1.1.2 Differentiating research and development

3.1.1.2.1 Research expands core technology, development implements core technology

There is a unique and interesting relationship between research and development (R&D) in an organization. The two disciplines are diverse and separate, although they

are often thrown together in conversation and lumped as one entity. In actuality, the two must function like a relay race, with research establishing the lead position and handing off the intellectual property and know-how to the development people to apply and create new products.

Research can be thought of as a strategic element of the organization, whereas development can be thought of as more operational in nature. Research is science-oriented. It translates phenomena into deterministic events, removes uncertainty, and is a resource tool for the development group. Often there is a disconnect between the week-to-week efforts of researchers and their respective development groups.

The operational issues with research include loose definitions of residual uncertainty; loose definitions of the delivered development tools; a loose definition of deterministic; and often a disconnect with the end-user of the technology, resulting in so-called projected customer acceptance.

Many problems in new product development occur when researchers have not completed their effort in totality. Time lines are not met, costs increase, and they forget that they are an integral part of the revenue-generating organization that must produce results on a specification and a time line to hit a window of opportunity.

Development, on the other hand, must focus on the creation of a product that can be manufactured and sold at a profit and achieve actual customer acceptance. Development engineers are charged with creating an accurate bill of materials, enforcing change management, creating documentation for manufacturing and product support, and testing. They also must identify and cement the product usage boundaries to ensure safety. Finally, they are charged with the human-factor engineering of a new product. Development has operational issues in establishing and maintaining linkages with research, ever more pressure on time lines, full absorption of uncertainty, funding constraints, compromises, and decision management between safety and cost. In total, they must translate customer needs into certified hardware.

For fast-moving technology companies, even research needs to be considered as operational, meaning it needs to have usable results, to be on time, to be well-defined, and to be easily transferable to development for industrialization and commercialization.

There are three basic types of R&D used in one form or another in industry today: incremental, radical, and fundamental.

Incremental research is best characterized by a small amount of research and some amount of development. This development is manifested by small, incremental improvements in the product, manufacturing processes, and cost reductions. Not much research is required; however, continuous development is. These efforts generally have large, aggregated payoffs over time but can be unsung heroes of the corporation's profitability index.

The radical type is best characterized by a large amount of both research and development. This is the type of development that requires new knowledge to be discovered and understood, as well as development of the means to embody the know-how into a product for a specific purpose. Huge expenses in development follow huge expenses in research. The organization launching into this type of development requires immense amounts of capital to see the programs through to profitability.

The fundamental type of program is best characterized by a large amount of research and little or no development. Many times these programs degenerate into gathering knowledge for the sake of knowledge rather than gearing the research toward a specific goal. They may be strictly strategic in nature and won't have a payoff for many years. The senior management must have a great amount of faith and long-range vision because the rewards of the decision and subsequent sacrifice will accrue to the next generation of management.

3.1.1.2.2 Research expands core technology

Research and development have three major strategic goals in an organization. Their purpose is to contribute to the existing businesses by allowing the organization to defend market and product positions, support sales growth, and expand existing business.

Research fundamentally generates intellectual property for the organization. If the activities of the research element are directed properly, the intellectual property will have tangible value on a balance sheet. This asset will be eventually used to develop new products.

It is therefore incumbent on the senior management to carefully select the programs to fund. They need to select programs that are tied to operational objectives, have a time table and an end, and result in development generating revenue for the organization through product sales. Management needs to think through the vision of the future in today's time frame and assign research work that will generate timely results for the next product developments.

3.1.1.2.3 Development contributes and makes money through product sales

Development, on the other hand, takes the research, intellectual property, and know-how and generates products that can be sold at a profit sufficient to offset the other expenses of the organization. Figure 3.1.7 is a simplified illustration of the role of development.

It can best be characterized as a machine that has marketing input, operational input, and core technology

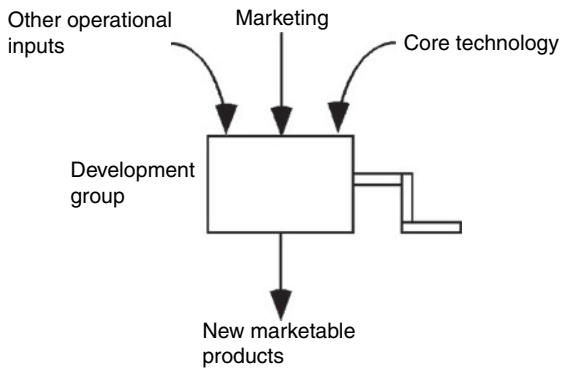


Fig. 3.1.7 The role of development.

available, and generates salable products that can be manufactured at a profit.

3.1.1.2.4 Development is much more measurable

The actual progress of development can be easily measured, since the development team is charged with the responsibility of committing to a schedule to produce results that are quantifiable. With the fundamental technology understood and made deterministic by the research function, development must concern itself with items such as bills of materials, performance, functionality, manufacturability, and cost. Milestones can be established and tracked with action items, responsibilities, and completion dates. Work content can be managed by manpower hours and loading in general, and results of expenses and time invested can be evaluated.

3.1.1.2.5 The need for research to be tied to operational goals

Because research generates the knowledge and is an integral part of development productizing the technology, it is very important for the research expenditure to be carefully made. Costs of progress from research to development increase dramatically, so diligence is required at the onset. Simply stated, ensure that research function is tied to operational goals. This means that the area of study must directly source the needs of development, and it, too, must occur on a time line. If not, development ends up cultivating the core technology and the project becomes irretrievably delayed.

3.1.1.2.6 The model for research and development

The following is a model for the research function and how it fits into the organization. Although organizationally the reporting relationship may vary, Figure 3.1.8 represents, in a general way, the flow of information and knowledge. It shows that information comes into research from several sources, including cooperative agreements with other companies' shared development in a consortium, from university sources, or from other sources as required. The corporate research department then generates the core technology to lay into the development programs so products can be developed to hit a market opportunity.

On a strategic basis, Figure 3.1.9 illustrates how profits are derived from the efforts of research. The goals and objectives identified for the organization's future drive a strategy to achieve them. Research contributes to the development process, and profits are realized through operations. It is a critical link in the chain, and results are required in sequence and on time to be effective.

On a project basis, the research function must contribute knowledge at various points in the planning and development process. Figure 3.1.10 highlights this contribution of core technology. The project starts with a concept. The marketing analysis defines the customers' needs. Research then contributes the core technology—the necessary knowledge to determine project feasibility. It

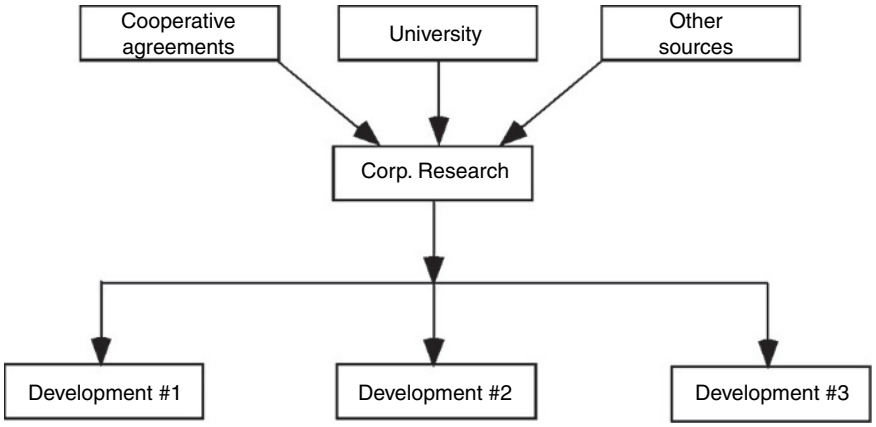


Fig. 3.1.8 Core technology flow.

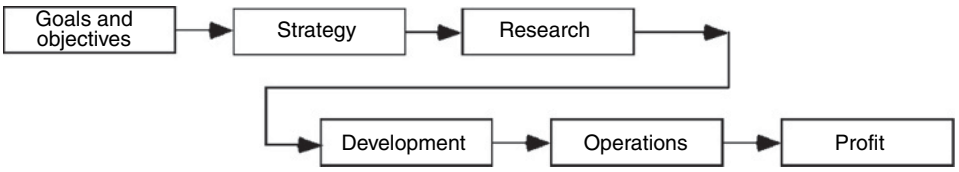


Fig. 3.1.9 Research sequence to profits.

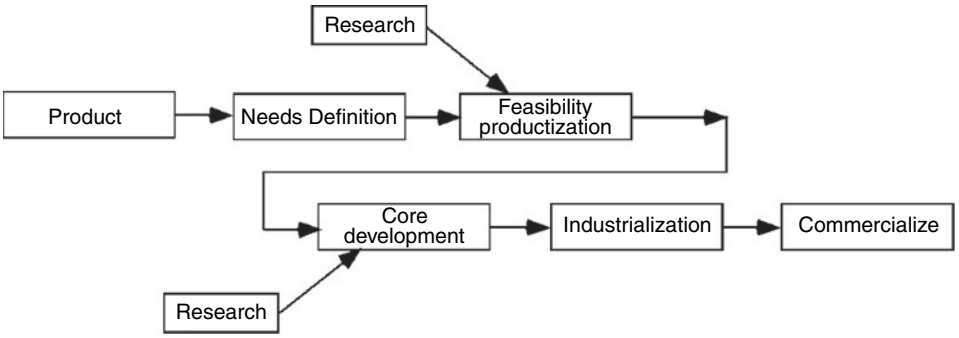


Fig. 3.1.10 Research sequence to product.

must answer the question: can we as a company absorb and cultivate core technology and produce it cost-effectively to create a profit? At the point that the affirmative conclusion is reached, the core development of the technology begins. Research then transfers the technology to development. An integral part of the development is the industrialization of the product, and then commercialization. As shown in Figure 3.1.10, research contributes at two critical junctures in the process: at the feasibility stage to determine if the technology can be brought to product status, and at core development when research transfers the technology to development.

3.1.2 Perspectives internal to the organization

3.1.2.1 The product fit

3.1.2.1.1 Selecting the new product opportunity

Opportunities abound in this world of ours! The challenge is to select and focus on the ones that will yield the desired corporate result. However, what is the desired result for the company? Is it revenue generation, cash flow, profit, or is it tangible, long-term growth for the company, strategic positioning, and satisfying operational requirements? Is it achievable, desirable, and agreed-to? All of these questions must be addressed to determine if the opportunity in question is the right one for your company at this time.

Not all opportunities are truly opportunities for a company. Although some of them represent good prospects from a pure market perspective, they may not fit with the company that is evaluating the opportunity. It is very important to factor into the decision all of the company attributes that will contribute to or detract from success of the new product. This exercise can be referred to as the fitability of the new product into the company.

3.1.2.1.2 Understanding the company's capabilities

A new product opportunity must easily fit into the operational culture and temperament of the firm. A company evolves with several interlocked disciplines that must work together to effectively assimilate, integrate, and execute a new product program. This must happen from a sales and marketing perspective, an engineering perspective, and manufacturing, service, and quality control perspectives.

The new product opportunity must fit with most of the disciplines to be pursued effectively. If there is an area that is not initially compatible or leverageable with the opportunity, then steps can be taken to correct or bolster the weakness. If, however, there are too many areas that are inconsistent with the new product, then the chances of developing the product within the market window of opportunity are reduced, especially while effecting changes in the organization and supporting the existing business. The new product idea must be tested not only in the marketplace, but also within the firm to determine its fitability with the sales organization, the

marketing organization, the current and anticipated capabilities in development engineering, and the integratability of the product into the manufacturing organization.

3.1.2.1.3 How does this opportunity fit into the goals and agendas of the company?

Every company has goals and objectives that generate its agenda. Even if the company doesn't have any stated goals or objectives, there is still an agenda that factors into the new product analysis. It may be as simple as: the objective is to maintain the status quo. In that case, the agenda is to do nothing in the way of product development. All new opportunities are evaluated and summarily rejected for one reason or another. Others are evaluated against agendas, such as target growth rates of 15 – 20% per year. Still others are evaluated against requirements demanded by customers, other parts of the organization, or that are purely financially motivated.

There are also the ownership issues that affect the company's ability to conduct a healthy product development program. A publicly held company with significant resources will evaluate an opportunity differently than will a privately or closely held company. A subsidiary operating from a remote location will act differently than will a division that is part of a larger group within the same venue. A set of poorly prepared objectives for a general manager of a relatively autonomous organization will dictate new product development activity differently than will ones given by an on-site senior manager.

The current financial status of the organization will also determine the risk and size of the development program. In the case of a subsidiary or a division, the status of the parent company must also be examined.

All of these issues must be factored when evaluating the opportunity. The operative point, simply stated, is that a new product opportunity must be evaluated not only from the external factors affecting the success, but also from the equally important internal factors that must contribute to its success.

3.1.2.1.4 Who makes the sacrifice?

As the saying goes, "No pain, no gain"; so, too, with new product development. Who in the organization will make the sacrifices? Is it the sales team that will make the extra effort to introduce the product? Is it the marketing team that will make the extra effort to prepare an adequate specification that will stand the test of development time? Will it be the development team that will shed the internal conservative agenda and break new ground in terms of costs, functionality, and manufacturability? Or, will it be the service team that will drive the product to success even through initial quality and reliability problems? Each and every part of the organization needs to contribute the extra effort to ensure success of the new product, in spite of the pitfalls it encounters.

If the fit is right, the team has a chance at success. The fit alone will not guarantee success; success also requires the commitment of the team members and the company as a whole.

3.1.2.2 The product strategy

3.1.2.2.1 What is the company's strategy in pursuing a market?

To effectively evaluate the fit of a new product, it is necessary to understand the company's overall product strategy. What has been its track record for embracing new technologies and product concepts? What were the expectations of each, and what were the results? Often, many companies do not have a clear objective and thus cannot generate a clear strategy.

The previous section discusses leveraging the organization's segments as part of the fit of the product. The product strategy should attempt to leverage the available product opportunities to create a steady stream of products to create a new, sustainable business unit for the company. This forces examination of the interrelationship of the product as well as individual contributions.

Regardless of how lucrative an isolated product opportunity looks, few products can succeed as a one-time hit without subsequent market and product development. It simply puts the organization in a market position that is not defensible. In addition, a single success in an uncontested market will not remain so for very long, necessitating the need for continued development.

Again, the strategy is dependent on several factors, including the structure of the company. A product or service that is in its infancy and has "a lot of runway" may be a good choice for a company that will eventually be taken public. It has the longevity and growth potential required to attract investors in the initial public offering. A product or developing technology that may be required to complete someone else's product line may be chosen to groom the company as an acquisition target. A privately held company will tend toward market niches that are narrow enough to be defensible and wide enough to be worthwhile. Consequently, a well-thought-out strategy executable toward an objective is invaluable to the organization.

3.1.2.2.2 Why does the business development team think this opportunity will take off?

New product development is a business venture. It must be evaluated like a banker or venture capitalist evaluates any business venture. Some developments may have very

interesting technology and be very fun to work on. However, each must result in tangible value to the organization funding the development activity.

As will be discussed later, the development team consists of several interdisciplinary elements of the organization. Therefore, the development team must make the case for the development program. Why will it be successful? What is unique about the product or technology that will be desired and accepted by the marketplace at this particular juncture in time and space? These are the hard questions that must be answered satisfactorily very early in the process.

The strategy, the timing, the functionality, and the company's plans must be instilled into the development team and must be eminently clear to all involved. The team must initiate, champion, and carry the torch of the new product business plan.

3.1.2.2.3 Where is the connection between the market and company?

At the risk of sounding like P.T. Barnum, I would stress that every product should have a distinct marketing advantage that it brings to the marketplace. Very few products are accepted that are "me, too" in nature, so there must be a clear marketing advantage at introduction. This is what starts the momentum that will carry future sales.

The marketing advantage is generally, but not limited to, the integration of features within the product that differentiates it in the marketplace. Longer, lower, wider, less expensive, more functional, easier to use, and more available are all product design tactics designed to make the product more marketable.

If the so-called hook exists, it is because the team placed it there, and it is there to generate sales. If the product is a "new to the world" product, its novelty alone goes a long way. However, it is then a sales and training job to gain market acceptance.

3.1.2.2.4 Think the sales scenario through with contingency planning

Each new product must have a vision of how it will play out in the marketplace. The scenarios for success and for failure must be thought through, and a recovery procedure for each worked out. This contingency planning is similar to conducting battles in a war. Competitive battles resemble the tactics and strategies of warfare. For example, if you introduce a new product and it has existing competition, you are the invader and the competitor will attempt to defend its territory. If it has uncontested reign of the market, and it is a worthwhile market to go after, the competition will enter the race with an offering of its own. Then the competition is the intruder. Each attack has its own set of measures and countermeasures as well as scenarios that must be developed by each participant to survive. It is best to attempt to think through the interchanges beforehand to establish contingencies.

3.1.2.2.5 The benefits of strategic planning

The term strategic planning has been a very overused term and, in many cases, is underimplemented as long-term activity. It needs to be specific enough so that its results define the product development and market development direction.

Strategic planning has been traditionally used to prescribe desired market position, dollar amounts of business segments, growth targets, means for justification of acquisitions, and more. It seems to be employed to reset a company's focus and direction initially; however, it soon seems to lose focus in the third, fourth, and fifth years.

Unfortunately, strategic planning has the most value as a learning tool for the organization by setting direction at the beginning of a period—executing a strategy or program and evaluating its success within the strategic framework initially laid out. Most important, the success or failure of the first year must be factored in when setting second-year and subsequent goals.

Consider a strategic plan that calls for accelerated growth by internal development of a steady stream of new products. Also consider that all-important review after that first year—showing the initial target and missing that target by some amount, for example. It is an imperative requirement for the accuracy and legitimacy of the plan that this be factored into the subsequent year's planning. If your company exceeded the first-year objective, what have you learned? What worked and what didn't work? Draw the correct conclusions from successes as well as failures.

If the ground rules of strategic planning are well understood and the company has the discipline to effectively execute them, then the planning process offers several benefits to the organization. As a review, here are the benefits of the strategic planning process:

1. It improves profitability by focusing the organization on two levels of profit: gross margin and net profit. It improves profitability regardless of the measurement criteria: return on investment, return on net assets employed, or returns on total capital employed.
2. It yields a higher growth rate for the enterprise with it than without it. Without a plan for growth, it becomes difficult to determine the best opportunities to pursue. This compromises the overall growth rate. The organization may happen to stumble onto a great opportunity; however, consistent increase in business and market share is not likely without integrated planning.

3. It reduces wasted time, effort, materials, and resources of the organization due to lack of focus. Following on the previous tenet, chasing loose opportunities can consume time and effort.
4. It provides focus on the marketplace and examines the business from the customer's perspective. Every business is about customers. It is they who make the business possible; it is they who comprise the marketplace. Strategic planning, although an internal goal-setting exercise for the organization, forces the management to focus on the customers, their problems, and the opportunities to resolve their problems.
5. It articulates and details future plans to the workforce to allow early buy-in to a program. It also serves as an early check for opposition to the future directions of the company. This point is not to be underestimated. Strategic planning is an excellent means to trial balloon issues and initiatives to test the waters of the organization. People make up the organization, and people can facilitate or destroy initiatives.
6. It fosters greater personnel commitment to the agreed-upon goals. As an essential part of the management commitment process, a personal commitment must be made and followed-through. The strategic planning process allows the participants to contribute and commit to the goals established in the process.
7. It provides for a cohesive management team. One of the most divisive elements in an organization is the fractionalization of the management team. Strategic planning allows for galvanization of this team. If it is not evident that the team can be galvanized in this fashion, it allows for change of the players, if necessary.
8. Strategic planning provides the framework to understand the competition in terms of competitive advantage, workforce, value added, position in the marketplace, and ability to react. This is one of the most lucrative elements of the entire process because it requires the organization to evaluate the competition in a dispassionate, objective way and to determine how the organization intends to compete and prevail in the markets it chooses.
9. It provides high visibility on quality from a holistic perspective: quality of the product, the process, and the entire revenue cycle. The strategic planning process forces quality into the organization. By its very nature of long-term goal and objective setting, the process dictates that the moves must be well considered and execution cannot be sloppy. The philosophy permeates all of the programs generated from the process and drives a higher level of quality onto the organization.
10. It yields greater customer satisfaction. Higher quality, an organization that knows where it is going and how to get there, and a realistic customer focus allow a greater level of customer satisfaction.
11. It provides a framework for continuous improvement. Because the process is long-term and annually reviewed, the organization can continuously improve through introspection, modification, and feedback.
12. It directs a day-to-day vision that is real and achievable. Although strategic planning is long-term, it affects the organization on a day-to-day basis by reinforcing goals and schedules. It keeps the organization focused and on track, rather than reactionary to every diverse opportunity that comes along.

3.1.2.2.6 The framework for planning

The planning process has a general framework that must be followed to generate the required results. A well-executed process has three distinct parts.

The first is introspective in nature. It encourages examination of the company itself to determine strengths, weaknesses, and historical perspectives.

The second part is creative in that the company must determine its future and its desired position in the marketplace.

The third part is pure planning. It defines the means to the end described in the second part. The following is a framework for the strategic planning process that can be used as a general format for initial discussion.

3.1.2.3 Introspection: where we are

Initiate a search of records to create a five-year history of the company's performance. This would include financial information and a narrative about the company. (Figure 3.1.11 presents a suggested format.) Analyze and draw initial conclusions from the obtained data. Review trends in sales engineering and manufacturing expenses, in addition to the product-related performance.

Use a narrative to describe the general health of the business. Interview a cross section of the employees to find where "the bones are buried" as far as product line performance, results of management attitudes, predisposition to new initiatives, and acceptance criteria are concerned. Do not focus on individuals; rather, focus on the results of the organization.

Next, create a product line analysis (as detailed as possible) to describe the performance of individual products. Examine the costs, features, benefits, trends, and driving forces for each product. It is also important to include the SG&A expenses for each because costs may not be allocated on an activity-based cost basis.

Category	Year 1	Year 2	Year 3	Year 4	Year 5
Sales					
Gross margin					
Net profit					
Expenses					
Warranty					
Inventory					
Return on investment					
Return on Capital					
Return on net assets employed					
Accounts receivable					
Cash reserves					

Fig. 3.1.11 Analysis of business.

Product maintenance is what absorbs time and slows down new product development, so it is important to examine, from a historical perspective, how the company handled earlier products and supported them. In some cases a product appears profitable from a margin analysis, but in actuality may break even and absorb a disproportionate amount of the organization's resources.

Figure 3.1.12 is a framework for this analysis. Be sure to allow for narrative for some of the intangible elements of the analysis.

Here is a review of the information required for this analysis. The product should be replaced with the description of the product. Consider individual products that are part of families as separate. Next, enter the price that your company is getting in the marketplace for the product. Cite important features and benefits that the customer has given as feedback. Do not use sales and promotional literature for this part; use actual feedback from customers who have used the product.

In the next row, describe the customer base for the individual product segments. Next, discuss the technology employed in your company and describe any differences from the competition. Also include manufacturing costs, fully burdened. Look at the current and near-term

Category	Product 1	Product 2	Product 3
Price level in marketplace			
Feature/benefit			
Customer base			
Technology base			
Cost			
Market driving forces			
Percent of total sales			
Contribution to profit			
Life cycle stage			

Fig. 3.1.12 Product line analysis.

market driving forces that will affect the acceptance of your technology, product feature/benefit set, competition, and demographic preferences.

Use accounting data to generate the row of sales data contribution to profit data for each one. Finally, enter the best assessment at the life cycle stage for each product. This may be numeric or a narrative. It will be used to gauge product development efforts throughout the overall framework.

3.1.2.4 Future: where are we going?

This section of the strategic planning process identifies where the organization wants to go strategically in the future. This section is one of the most important of the three sections because it is the goal that must be set for the organization. It defines the call to action, the point in the future that the organization wants to attain. The following are the basic elements of this section and represent the basic requirements for defining the organization's goals.

3.1.2.4.1 Theme or mission statement development

This is an often-overused term, and is frequently invoked by senior management as the only means of reinforcement of long-term plans. Simply stated, the theme statement is the short, concise description of the business of the company and its objectives. It describes the company's identity and role in the marketplace.

3.1.2.4.2 Narrative of the vision

Create a detailed narrative of the vision of what you want your company to be in the future. It should discuss the time frame, expected results, financial position, market position, product position, and competitive comparison. It should also outline the ownership structure at that time, and whether any changes are expected.

3.1.2.4.3 Narrative of how the plan and reality will fit together

Create another narrative that weaves together all of the requirements for the future dream and the present and future operational issues. This is done for the sole purpose of outlining the plausibility of the plan and goals, and to ensure that they are consistent and compatible.

3.1.2.4.4 Five-year product line catalog

This is an interesting and enjoyable exercise because it allows management to sculpt the future offerings of the organization by outlining the five-year product and services catalog. You most likely will be required to go through a detailed product and services evolution flow chart to generate the complete catalog. That process will be outlined later. However, at this time it is sufficient to list the products as they would appear in the catalog.

3.1.2.4.5 Product scope definition

The product scope definition is a means to describe the product portfolio in a detailed manner. It should show how the products fit together and leverage off each other, and where the boundaries are. Also, take time to establish low, medium, and high unit volume scenarios for each of the products. This will become helpful for the identification of the core requirements of technology that the company must have.

3.1.2.4.6 Specific market segments

Next, detail the specific market segments for each product. Do some market research to determine the trends in the marketplace for these areas and plan scenarios as to how your effort will fare competitively. Management should, as part of this exercise, question why this strategy will work and why this product segment will succeed in the dynamic marketplace.

3.1.2.4.7 Industry trends

Finally, the team needs to look at overall industry trends to determine whether they will be active through the planning cycle. If basic natural resource industries, such as the energy industry, are not active from a capital spending point of view for the near term, be aware that you may position the organization for the end of the planning cycle; however, it will be dry run through the planning cycle. Strategically, it's a good plan; however, from a cash generation point of view, it may not be fiscally allowable.

3.1.2.5 How will we get there?

This section is the pure planning section. It specifically outlines how the organization will move from point A to point B in the planning time frame. Whereas the previous narratives describe what the organization will look like, this section outlines what specific steps will be taken. Each area of the organization will be examined and affected in the transition. There will be marketing issues, development issues, manufacturing issues, service issues, and financial issues.

The following outline lists the elements of this section of the plan that must be addressed.

3.1.2.5.1 Segment strategy

This portion of the plan views the market strategy of each segment. It should start with a current market assessment of the segment. The purpose is to find out what is going on in the marketplace and to view the trends, both

IMPLEMENTATION MATRIX		Year 0	Year 1	Year 2	Year 3	Year 4	Year 5
Product 1							
	Internal development	X					
	Joint venture						
	Brand label						
Product 2							
	Internal development						
	Joint venture		X				X
	Brand label						
Product 3							
	Internal development			X	X		
	Joint venture						
	Brand label						
Product 4							
	Internal development					X	X
	Joint venture						
	Brand label		X			X	

Fig. 3.1.13 Implementation matrix.

near- and long-term. It is also imperative to cite critical success factors. These two elements then drive the strategy—how the company will position itself and what role the company will play in the marketplace. The plan also needs to define how the company will play this role. By what means will it prevail in the competitive environment: innovation, price availability, or training? These are the questions that must be answered to complete the strategy.

3.1.2.5.2 Implementation

Since not all growth is internally generated (especially in a five-year scope), some of the businesses and product lines may be acquired or brand-labeled, or be a result of a joint venture development. In this part of the plan, each segment is identified and a strategy is developed. In addition, an implementation plan accompanies each one to determine whether it is an internal development or some external mechanism. The chart in Figure 3.1.13 illustrates the plan.

3.1.2.5.3 Investment/return

In order to add financial validity to the plan, it becomes necessary to identify the scope of funds required to carry out the plan. Each investment or market segment that has been identified, each strategy that has been established, and each company organizational requirement that has been outlined must now have a financial picture attached to it. An investment/return money line should be established for each. This is illustrated in Figure 3.1.14.

Be sure to include all investments and both the development expense and the capital equipment expense. Although this is not completely pure from an income statement/balance sheet point of view, you are looking for the scope of funds to be expended. This is most often executed with cash, and therein lies the requirement.

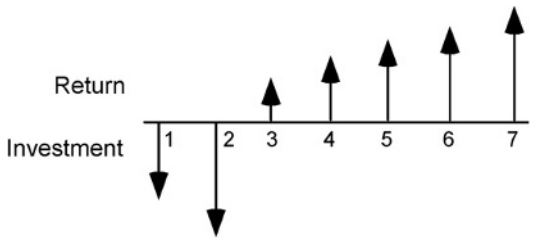


Fig. 3.1.14 Investment/return.

The return should be calculated as follows: Take each individual segment and determine the number of projected unit sales. Then generate a projected gross profit based on revenue and costs, and post the product of the gross profit and unit sales to the return stream of payments. It is best not to post revenue only as a percent of market share obtained or as total dollars (independent of cost and units) because this often lacks substantiation. Also, the intent here is to evaluate the opportunity of the segment. As such, you want to compare real dollars invested against real dollars realized.

Next, calculate a growth rate for each segment and establish a trend measurement for each. In this way the opportunity can be evaluated against a desired financial profile and also against each segment.

Finally, aggregate all of the investments and all of the returns and post these results to a money line.

In this way, the financial pro forma accompanies the narrative of the plan, and senior management will get a financial feel for the plan and the scope of the expenses.

3.1.2.5.4 Requirements

There will be two sets of requirements for the organization and the plan. One will be internally focused and the other will be externally focused. Both sets of requirements must be addressed and cannot be glossed over.

3.1.2.5.5 Internal

There may be infrastructure changes as a result of the plan. These should be outlined and specifically listed (with timing) if they are critical to the success of the plan.

3.1.2.5.6 External

There may also be external requirements (e.g., in the sales organization, marketing channel, or vendor links). These need the same treatment as the internal changes to the organization.

3.1.2.5.7 Recommendation

The recommendation is one of the most important parts of the strategic plan. All of the data observations and opportunities must be coalesced into a firm recommendation for the company to follow. It also determines the acceptability of the plan from a management and commitment point of view. There are several elements in the recommendation section that should be included for the plan to be complete and to increase chances for acceptance.These are discussed in the following.

Trial balloon the plan with internal people from different elements within the organization. This can be done in stages, in parts, or both. By doing this, you can get initial feedback from the organization and a measure of the acceptability to certain groups.

Next, trial balloon the plan with the sales channel to get their perspective. This will afford you the opportunity to see who in the channel will support the plan and who is not on board with it. This information will be helpful at a later time when evaluating constituent players as the plan is executed.

Finally, visit a cross section of the customer base and present it to them. This feedback is the most helpful because it is the customers who will eventually support the plan by continuing to purchase existing and new products from the company.

It now becomes necessary to evaluate the feedback from the internal organization, the channel, and the customer base to factor their input to any modifications that may be required to ensure the plan's success. Any modifications made should be communicated back to the people who suggested them to enhance acceptance and buy-in.

Learn the dynamics and the biases of the organization and retreat from confrontation in the recommendation section of the plan. The objective is to create agreement, not to teach or preach to the organization.

Next, leave the plan alone for several weeks, if possible, and work on something else. For example, visit some customers, manage other projects, or simply distance yourself from the plan, if possible.

Then, reread the plan and examine to see if it's reasonable. Does it still read like a plausible plan? Does it still seem achievable? If so, proceed.

Finally, make the recommendation to management and get their agreement.

3.1.2.5.8 Action plans

After you have agreement of the principals, it is necessary to generate an action plan. Without it, the plan becomes a conversation piece and none of it gets executed. Worse yet, senior management may assume that action is taking place when it may not be. This will impact credibility and eventual funding. The action plan should be very specific with times, assignments, and completion dates. Even if it is not totally complete, it is important to establish momentum early and set the pace for the organization. If funds are necessary to be appropriated by senior management, put them on the action plan with an assignment and a completion date. You will find out early if this is a real plan or just wishful thinking. The final element in constructing the action plan is to refrain from taking half-steps. If the plan is reasonable and the recommendation is well-considered and achievable, then do not compromise at this stage by taking half-measures. Execute the plan as designed.

3.1.2.6 Dead-ended single-product programs

3.1.2.6.1 A new product should fit into: a family supported by a strategy

There has been significant discussion regarding strategy, vision, and operational planning and all of the steps associated with it. The reminder here is to refrain from one-time-hit programs. The single-hit programs are characteristically seductive, lucrative, and have huge upside potential. These programs also have generally longer lead times, significant uncharted development waters to navigate, and often require huge capital outlay. Long-term success of an enterprise relies on the need to stay on strategy, and nurture and execute a plan quickly. Focus on solid strategies; then execute tactical and operational planning to carry them off effectively.

3.1.2.6.2 Product development initiatives should not be easily copied by competition

Take care in selection not to choose programs that are all alone, like an island. An enterprise must have significant financial resources to embark on a development program with no previous core technology or background. These firms are at risk when embroiled in these types of programs because they can consume tremendous amounts of capital and personnel. Progress may be slow and obstacles might require disproportionate effort to overcome. They generate characteristic revenue curves, making them easy targets to knock off. They are financially stretched; they have no depth to the core technology; they have not productized it before; and their slow progress allows a competitor or consortium of competitors to leapfrog them technologically with their product offering. The marketplace is very dynamic and fast-paced. Loyalty is as strong as your last product advantage. Alliances between companies can be fleeting, and your ally today may be your opponent tomorrow.

3.1.2.6.3 The family of products retains momentum and therefore strength

As discussed previously, a family of related products under development, in which the technology employed is one of the core technologies in-house, serves the best long-term prospects of the organization. It retains momentum, offers flexibility, and fosters financial and competitive strength for the company.

3.1.2.6.4 Practicing the concept of leverage

This is the basis for practicing the concept of leverage. Combine efforts to serve a strategic goal or use a single effort to serve related, multiple goals. This philosophy leverages the product development resources in an organization. The managers of several programs should spend their time looking for and creating this leverage. Start with modest goals that are achievable and build on the successes. Along the way, gather ideas and execute selected ones that contribute to an overall portfolio.

3.1.2.7 Consistency

3.1.2.7.1 Is this a reasonable opportunity?

As part of determining if the prospect in question is a real opportunity, one needs to conceptualize and visualize the product offering as part of the company's future standard offering. You need to visualize a sales call with your sales team working the channel. Evaluate the plausibility of the product. Can you see your team assimilating and effectively promoting it?

Do not assume that the organization will step up to the plate to promote the new product. Take care to strike a delicate balance between a new product stretch for the organization and something that is unachievable given the resources and mix of players.

Given the development cost, resource allocation, and lost opportunity cost incurred in the creation of the product, a certain amount of introspection and due diligence are required to ensure that the organization can be effective.

3.1.2.7.2 Setting the criteria for evaluation

To provide some consistency in the evaluation process, it is helpful to have a criterion by which to evaluate these new product prospects. A framework could be as simple as a list of criteria that must be satisfied in order for the product to fit in the organization. This has little to do with the market fit or opportunity for market success; rather, it deals with the internal limitations of the organization.

The benefit of this type of evaluation is that it removes the emotionalism in the decision-making process. It also limits the effect of personal preferences and agendas, making the selection more objective and driven by a corporate decision-making process. To make this even more effective, factor in the individual company's time to develop into the criteria, to determine the ability of the company to capitalize on the market window of opportunity.

Perhaps the best way to accumulate the criteria is to establish a rating system that assigns weight to each of the criteria and allows for an arithmetic means for selection. In addition, a single opportunity may not have to satisfy all of the criteria; in some instances it may only need to satisfy 70 – 80% to qualify. However, a few of the criteria may be mandatory. The following are examples of criteria for a company that has changed ownership and is

highly leveraged. In this case, opportunities taken must be deliberate and must have large payoffs to offset the risk in valuable capital:

1. Is this a finite opportunity or a new market?
2. Are the projected sales at least $2.5 million per year for four years?
3. Can our manufacturer's reps sell this in most of the territories?
4. Can we design and build it?
5. Do we have to install it?
6. How much after-sale support is required?
7. How many people will we have to add to headcount to support this?
8. Can we sell it to other customers?
9. Can the design be leveraged from another design or to another design?
10. Can we develop it in less than a year?
11. Is the gross margin more than 50%?

As an example, for the highly leveraged organization that must make every dollar spent payoff, projected sales, corporate capability, channel leverage, and development leverage may be absolutely mandatory, whereas there may be some latitude in the remaining criteria.

3.1.2.7.3 How much of a stretch is this for sales, engineering, and manufacturing infrastructure?

It is important to assess how much of a stretch the new product will be for the organization on all of its operational fronts. As discussed earlier, one cannot blaze new trails on every front and expect any leverage to take place. A degree of stretch is desirable in each of the operational elements to keep the organization fresh and competitive. A visual way to look at this is to generate the graph in Figure 3.1.15.

Radially place lines from a center point as shown. Have each line equally placed at the same angle. The length of each line indicates the measure of difficulty to implement the new product. A separate line is used for each of the critical operational areas that contribute to the product's success. There is one line each for sales, engineering, manufacturing, quality, purchasing, service, and repairs. Assess where the organization is with respect to the competition and plot on each of the operational areas. Also plot where the new product will fall with respect to the present internal capabilities of the organization. In this visual way it is easy to assess the degree of difficulty in implementing the new product.

The black line represents the company's capabilities in the various disciplines. The red line represents the present capabilities of the marketplace, and the blue line represents the present requirements of performance for the new product or business. By looking at the relative comparison between these various assessments, one can evaluate the company's capabilities to execute the new product development program.

As can be seen, the company is a little short in the sales and marketing discipline and significantly short in service. The diagram also shows that the product requires service capabilities even ahead of the market's

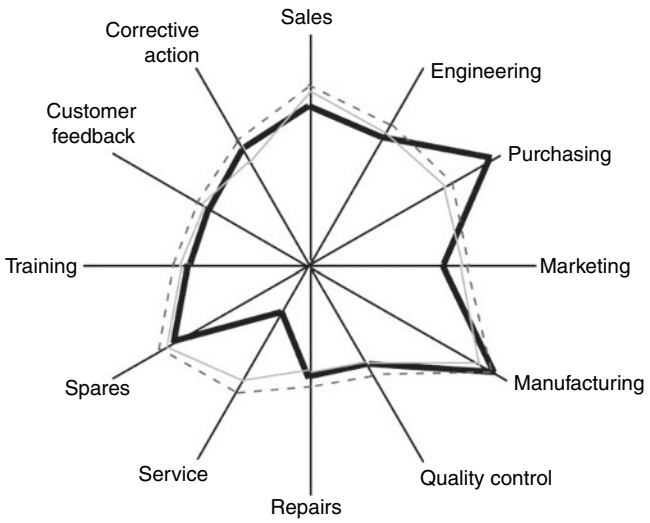


Fig. 3.1.15 Fitability graph.

expectation for service; therefore, the company will have to make some changes.

3.1.2.7.4 Ensuring that all of the business stakeholders are committed

As stated earlier, it is crucial to the new product development's success that all essential stakeholders of the organization are in agreement and are committed to the execution of the program. It is therefore incumbent on the person entrusted with the leadership role to assess this commitment as early as possible. Interview the key players in the organization to assess their levels of commitment to the program. Correct or alter any attitudes that will stand in the way of success. If you cannot, change the players. The energy needs to be spent in executing the program to position the product favorably with respect to the marketplace. Do not expend the energy trying to convert nonbelievers within in the organization.

3.1.2.7.5 The danger of "promised commitment" or getting commitment early

There is a danger in getting what seems to be commitment early from several people in the organization in word, but not necessarily in deed. Consequently, it is good practice to involve all the players in the organization at the outset. Require their participation early in the process and observe their attitudes. Don't shy away from confrontation; in fact, invite some of it. In this way the truth comes out and it can be resolved early on. Also, there is a benefit to some confrontation in that some good ideas and observations may be brought to light (which otherwise would have not been discussed) and integrated into the program. In addition to inviting the confrontation, it is the manager's responsibility to drive it to a conclusion for the team. In this way the group can see how the objections, discussions, and disagreements can be a positive force in bettering the program.

3.1.2.7.6 Taking the "basal temperature" reading

As the program progresses, it becomes necessary to periodically take the basal temperature of the organization with regard to the program. This serves as a reaffirmation of the commitment of the key managers and the key players to impart a proper sense of urgency to the project. Do not let apathy or negativism begin to permeate the group, because this begins to feed on itself and eventually will slow progress. In the leadership role, your responsibility is to smooth over internal speed bumps and focus on the competitive- and market-driven issues.

3.1.2.8 Continuity

3.1.2.8.1 The continuity of development is as important as the development itself

The continuity of development can be considered as important as the development itself. What this statement means is that a company must become proficient at recognizing an opportunity, initiating action to capitalize on it, and driving a program through to success with it. If an organization can achieve this with the determination to weather the obstacles along the way, it will eventually be able to get to anywhere it desires to go.

3.1.2.8.2 On again/off again

Conversely, the developer's nightmare is a management structure that issues on again/off again orders and constantly changes priority on projects. In these cases, the bulk of the effort that needs to go into the development of the product actually goes to winding down a program, only to later reinitiate it.

3.1.2.8.3 Building momentum

Figure 3.1.16 illustrates this point. If there are interruptions in the development processes, the returns in revenue cannot be additive, according to the model for leverage discussed earlier. According to the leveraged revenue model, each successive program builds the revenue and profit curve toward a more exponential shape. Interruptions will therefore alter the timescale and shift the revenue (missed market opportunity) and profit-generating potential scale away from the desired characteristic shape.

3.1.2.8.4 Geometric progression of rewards

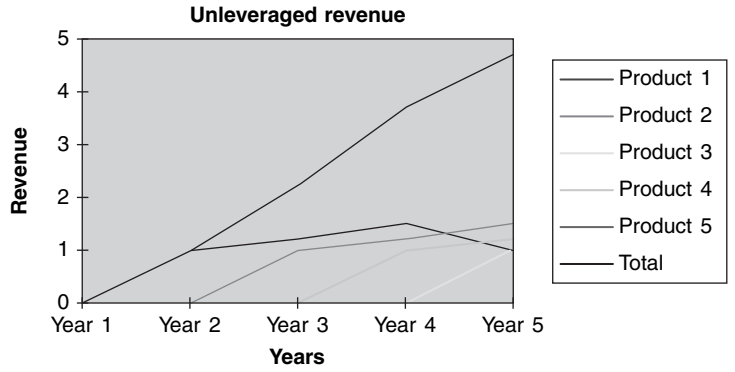
The objective, therefore, is to reap a geometric progression of rewards for the invested development. The operative word here is investment, in that a company desires to harvest as much revenue and market share as possible from an investment made in product development. Future revenue dollars are discounted by inflationary pressures, and development dollars are invested presently at full value and generally after tax (outside of R&D tax credits used from time to time). Therefore, almost by definition, a certain amount of leverage is mandatory.

3.1.2.8.5 The pursuit is as important as the attack

The important point of this section is to understand that continuity is critical in new product development. The company must initiate a program, execute it in a timely manner, and harvest the reward of its investment. To that point, the pursuit of the business is as important as the

Unleveraged revenue					
	Year 1	Year 2	Year 3	Year 4	Year 5
Product 1	Delayed	1	1.2	1.5	1
Product 2		Delayed	1	1.2	1.5
Product 3			Delayed	1	1.2
Product 4				Delayed	1
Product 5					
Total	0	1	2.2	3.7	4.7

A



B

Fig. 3.1.16 (A) Product introduction timing; (B) unleveraged revenue.

original attack. Each program and action must be deliberate and have a payoff. There can be no voids in the continuum because this wastes time and dollars; hence, the pursuit is as important as the initial attack on a market opportunity.

3.1.2.9 Leverage

3.1.2.9.1 Each new product development effort should leverage the organization

The key to effective new product development is to create leverage in the development arena similar to creating leverage in the manufacturing and sales arenas. It simply is unaffordable for most companies to continuously start from ground zero on every new product development; consequently, the desire is to leverage past development efforts when starting new programs. Figure 3.1.17 illustrates the issue.

As shown in the figure, successive development programs can be thought of as developing concentric rings around a base of core technology. As the programs become more complex and involved, the core increases. Each activity then pushes the outer envelope in terms of technology, purchasing, sales and marketing, and manufacturing systems. In this way the company develops expertise in all areas with each program and experiences real growth in sales, margin, and the infrastructure to support it. A less abstract and more

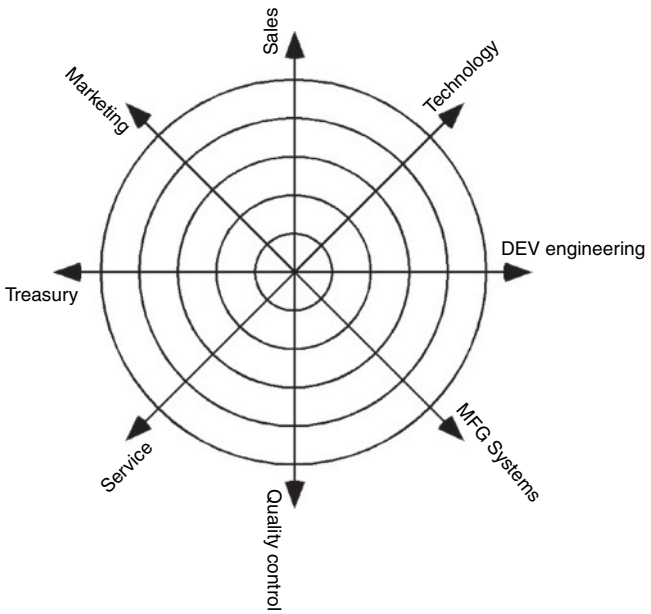


Fig. 3.1.17 Extending operational capabilities.

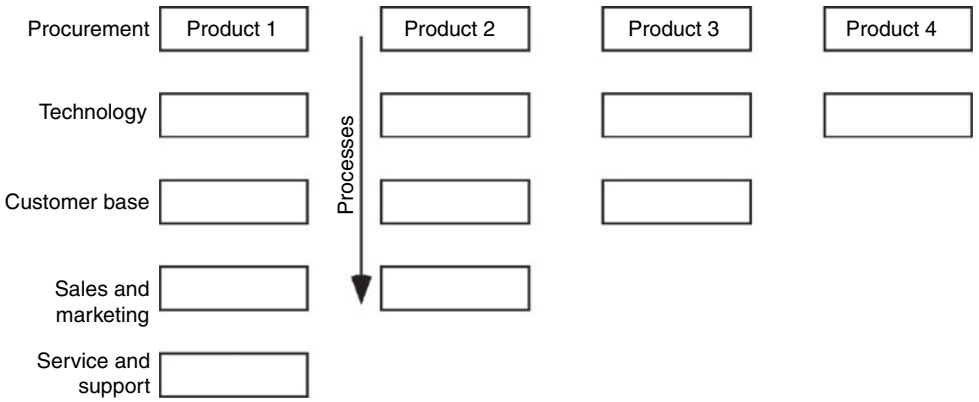


Fig. 3.1.18 Company capability development.

operational way of looking at this is shown in Figure 3.1.18.

Figure 3.1.18 shows that a single product demands company resources in the procurement of materials, the engineering of technology employed, the manufacturing processes, and the sales channel to get it to the marketplace. As new opportunities are evaluated, a portion of the evaluation needs to address the leveragability of the product within the organization for the successive programs. Building a business using similar manufacturing processes, engineering and development, and routes to market contributes to the leverage discussed.

3.1.2.9.2 The result should be 1, 2, 4, 8 versus 1, 2, 3, 4

By leveraging the SG&A elements of the business, a successful series of programs will generate financial leverage in the form of growing revenue and profit. Because the knowledge is being cultivated in-house and applied to the programs, the leverage shows up as faster program execution times and lower costs. Faster time puts the company in a better market position, and lower costs make the program easier to amortize. Revenue- and profit-generating programs then sum up to a result that generates a profit curve that begins to look exponential rather than linear. The difference is the leverage that is created. This is illustrated in Figure 3.1.19.

Each product generates growth for the company, starts out small, and builds momentum in the marketplace. As the company then structures the programs in succession (and does this for several product lines), the revenue and profits grow disproportionately. This is the leverage that is desired; however, there is also a collateral benefit, as we shall see in the next section.

3.1.2.9.3 More flexibility in the event of failure

This philosophy will allow the organization to recover from a single failure more easily than with other strategic philosophies because there is overlap in all areas, and each obstacle does not result in navigating uncharted waters. This philosophy also takes more time and effort to initiate, but results in a more solid momentum after you get it going.

3.1.2.9.4 Improving the chance of success

Consequently, this will better the company's chance of success in a fast-paced, changing marketplace. Most articles on new product development will discuss the percentage of failures in the marketplace and the historical reasons for failure. This philosophy represents tangible, positive steps that can be taken to ensure a higher percentage of successes. It also becomes important when discussing platform development and recovering a program that is in trouble.

3.1.2.10 Flow

3.1.2.10.1 Take and execute manageable chunks of development

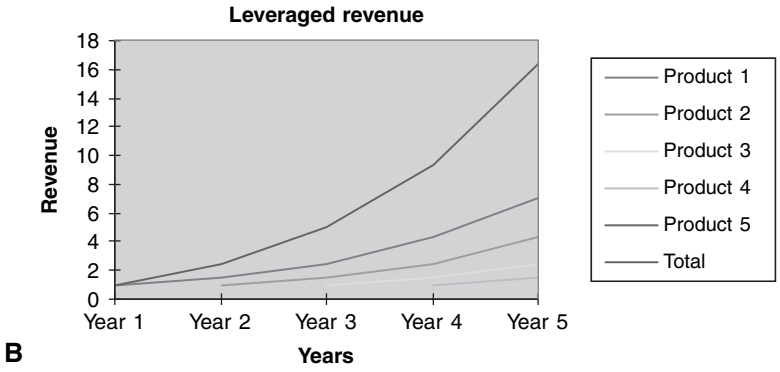
To keep the flow of a new product development program smooth and timely, each step of the process needs to be identified and broken down into component parts for analysis and execution. It is important to refrain from selecting huge chunks of activity to complete without detailing the specific parts. By taking small elements, the team can use trial-and-error in managing uncertainty and essentially practice its problem-solving techniques. As will be discussed later, the smaller the component part, the easier it is to populate the team and create a timely flow of progress.

3.1.2.10.2 Get great at executing small, incremental improvements

The company needs to get highly proficient at executing these small, manageable chunks of development

Normalized sales results					
	Year 1	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Product 1	1	1.5	2.5	4.375	7
Product 2		1	1.5	2.5	4.375
Product 3			1	1.5	2.5
Product 4				1	1.5
Product 5					1
Total	1	2.5	5	9.375	16.375

A



B

Fig. 3.1.19 (A) Normalized sales results; (B) leveraged revenue.

activity so as to remove the uncertainty associated with the huge project at the outset. By doing these small chunks effectively, repeatedly, and predictably, one can then manage a large program within a budget in terms of time and money. The only way to confront the uncertainty hidden in a large program is to break it down to its component parts for execution. This also has a collateral benefit of ensuring completeness by including all of the component parts required for the program.

3.1.2.10.3 Smoothing the progress and making it measurable and predictable

The benefit here in terms of project management is to make these component parts measurable in terms of effort and talent required, and to be in a position to plan and complete activities in a predictable and repeatable fashion. This takes practice and several programs to become proficient. To achieve this means that almost any program can be broken down into its component parts for analysis and a measurable amount of effort can be applied, and the result can be predicted within a time frame.

3.1.2.11 Pursuit

3.1.2.11.1 Becoming relentless at improvement and execution wins the business

The company should evolve a culture of consistent improvement, incremental product development, and timely execution. The competition does not maintain the status quo, either in isolation or collectively; this mandates your firm to initiate consistent, predictable actions that result in increased market share and better competitive posture.

3.1.2.11.2 Key vocabulary: small, incremental, definable, measurable, predictable, and fast

Once the strategic vision is locked-in, the company must perform. The watchwords of this performance are small, incremental changes that are definable, measurable, and have leveraged payoffs. Any actions that the company takes need to be fairly predictable at the outset, executable in a short period, and relentless in the succession of these steps. The firm needs to maintain the strategic vision and strive to relentlessly execute steps to achieve it.

3.1.3 Perspectives external to the organization

3.1.3.1 Defining the marketplace

3.1.3.1.1 Who is the customer?

Always understand and know who the customers are. They are the ones generating the revenue through acceptance of your solution to their needs. Continue to solicit their opinions, their wants, their changing customer needs, and their pressure points. This will ensure that as changes in the landscape take place, your company will be positioned to respond appropriately.

3.1.3.1.2 Demographics

Do not be lulled into a false sense of security by demographic summaries. Because so much of our lives is governed by lumping people together and treating their issues as common, keep in mind when serving an original equipment manufacturer (OEM) market, the customers are trying to differentiate themselves to create a competitive advantage. Consequently, they need your company to customize your solution for them. In trying to serve a consumer market, keep in mind that there are different classes of customers with different needs that have to be met by the product or service. Products cannot necessarily be lumped together and mass-marketed with success, as a demographic summary may indicate.

3.1.3.1.3 Users, OEMs, resellers, and agents

Remember that a market is a collection of customers, not a uniform, homogeneous purchasing machine that can be pushed, prodded, and calculated with decimal-place accuracy every time. There are intermediaries in the process who have their own problems, issues, and agendas. Leveraging the initial product development and the launch requires real-time monitoring and action to build and preserve momentum.

3.1.3.1.4 What are the customer's alternatives to your solution?

Always remember that the customer has alternatives, including competitive offerings at a lower or higher cost, and functional displacements. In some cases, there may be personal apathy on the part of the customer. The operative point here is that the customer your company is approaching to sell to may not be predisposed to the solution you are offering. To understand this better, place yourself in the customer's frame of mind and evaluate his problem and look at all of the industry solutions. Then measure your solution against his alternatives.

3.1.3.1.5 How is your competitive advantage embodied in the product?

Often this is evident in examining competitors' solutions to customer needs. Many times a unique or lower-cost method may be used to accomplish the same function. The difference may be that the development engineers have biases and mental paradigms, as we all do in addressing and solving problems. The innovative engineers, designs, and companies break these paradigms to offer alternative approaches embodied in their products for satisfying customer needs. That is the business objective in new product development.

3.1.3.2 Understanding the marketplace

3.1.3.2.1 It's an ever-changing, complex world

The world is becoming more homogeneous as time progresses. Actions taken overseas affect transactions and business here. Changes in currency affect global trade and individual buying decisions. The marketplace is simply becoming more complex and initiatives and actions need to be multidimensionally evaluated before implementation, or they may be ineffective or disastrous.

As an example, let's assume your company supplies electronics to an OEM in North America for sale to end-users in North America. A sister company to the OEM (half ownership) resides in the Tuscany region of Italy and sells its equipment to the European continent and the rest of the world. The Italian company is the primary mechanical manufacturer for the domestic OEM and has electronics supplied by a privately held Italian electronics manufacturer, supplying the same package near your cost. Do you have a domestic threat from the Italian electronics company here? What are the critical factors to secure continued domestic market share? Can you maintain domestic loyalty here even if the Italian mechanical manufacturer prefers its own supplier? Can you be functionally displaced by the Italian electronics manufacturer who is working in close conjunction with the Italian mechanical manufacturer on a new platform that will be brought to North America?

All of these factors must be evaluated. You will find that it begins to mimic a chess game with an international flavor. The operative point is that the world is a complex place.

3.1.3.2.2 What are the customer problems to focus on?

The effective way to navigate the marketplace is to focus on the customers' needs. Their problems are the ones that need to be solved in the embodiment of your products and services. By keeping close to the customers' wants and needs by resolving their problems, you

position yourself as a problem solver and a consultant to be trusted and relied on (i.e., a business partner), a position that every supplier aspires to.

3.1.3.2.3 Understanding marketplace dynamics

Because the marketplace is a fast-paced, dynamic medium, it is most important to understand the dynamics of a market. Your company is attempting to do one of two things. Either you are trying to develop a new market that currently doesn't exist, or you are trying to increase market share by taking it from others. Both actions drive reactions in the market. The first will draw attention to the lucrative opportunity. The second is basically a catfight to secure the business. Knowing the marketplace dynamics will assist in developing the strategy for each. Failing to understand the dynamics can be fatal to the effort.

3.1.3.2.4 Understanding customer perceived value purchasing triggers

What does the customer purchase—a product, the satisfaction of a need, or peace of mind? This question and its answer are the essence of the marketing and sales of a new product. The products and services are fundamentally the embodiments of the customer's satisfied need. Many marketers assume that the product they make is what the customer buys (i.e., the customer buys these features). The customer actually buys benefits and solutions to problems. If you fail to understand that or provide for it in the product embodiment, sales will be fleeting.

If you fail to satisfy the need, there is no compensation that will offset what is lacking. Meet the need well and there are many things that will be overlooked, all else being equal. Remember that the product you sell and get profit from is simply the embodiment of a satisfied need and is crucial to formulating any program. Failure to understand and accept this fact may cause wasted time and effort for the company. A basic example of this is an insurance policy: is the customer purchasing a document with legalese and clauses to satisfy a need, or are they purchasing peace of mind from the insurance company?

3.1.3.2.5 Positioning the product

This is why it is so important to position the product in the mind of the consumer. This positioning must tell the story in a convincing manner so that the buyer's perception of the product satisfies the need. It is important to understand that, at the outset, it is the customer's perception that is making the decision. Subsequent buy decisions after the first sale will be based on product performance in meeting the need. In other words, no matter how good any product is, fundamentally, the customer must have the perception that the product will satisfy their need at first sale or you will never get the opportunity to let performance satisfy the need. Positioning the product creates that initial perception.

Positioning the product means different levels of difficulty for different industries and products. For example, repositioning a product by repackaging the same basic elements is easier than redesigning a product platform for an added or different feature set.

3.1.3.2.6 What triggers the purchase order?

A study of business transactions would indicate that, during the exchange between the company representative and the customer, there is a point at which the customer has all or most of his perceived needs met and decides that this is the best deal he can reasonably get. This is the precise point where a closing needs to take place. This is what will trigger the purchase order. Asking for one before this point is reached is pointless, and waiting past this point is dangerous. It becomes a critical success factor in a sales transaction to seize the opportunity to create this point and act on it.

3.1.3.2.7 Practice, learn, adapt

They say that practice makes perfect. That is generally true because inherent in practice is scientific methodology; try, evaluate, learn, modify, try, evaluate, and so forth. This is the value of the lost sales report—it is a key element in the analysis of the transaction and is integral to any improvement change to become ultimately successful.

3.1.3.2.8 Testing the effectiveness of different sales styles

There are different sales strategies and styles for different products, and it is important to match the sales style to the product type. If, for example, you have a technological leadership position, the sales style would be different than if you were supplying a "me, too" type of product. The difference is a value-based style versus a price-based style. This is an area that must be identified and coordinated with the product launch and that is integral to the marketing plan. If it is not, you may be wasting the sales personnel's time and risking their disillusionment. Many companies have failed in product launches because they assumed that a brand new, highly technologically advanced product could be marketed by an existing sales force with no background in the technology.

3.1.3.2.9 Modifying the sales approach for effectiveness

Marketing and selling a product is similar to telling a story to the customer, at least from the promotional aspect. Ideally, the interaction needs to symbiotic but, for purposes of this discussion, the marketing "story" needs to

be consistent with the launch objective. If it is observed that, for whatever reason, the presentation needs to be modified, it should be reevaluated in conjunction with the entire product plan and product launch.

3.1.3.2.10 Tiered market structures: high, medium, low

Also, different tiers in the market demand use of different sales strategies and tactics. If the product line consists of vehicles that traditionally had status and utility associated with them, there are several strategies to use to be successful with each tier of luxury, value, and low-cost transportation. Parts of the promotion need to appeal to the different wants and needs of the buyer. The operative point is that the multiple-tiered market must match up with the multiple product line to be effective.

3.1.3.2.11 How can it be sold?

Product marketing personnel need to constantly evaluate the effectiveness of the promotional programs and continuously ask how the product can be sold. Because the landscape is constantly changing, the positioning must be modified to keep pace. This can be quite difficult to do during an initial launch because it requires introspection and openness in assessing sales levels and degrees of success.

3.1.3.2.12 Where is the value created and exchanged?

To effectively market and sell products, the company must determine where the value is created and where it will be exchanged. This is the essence of the transaction. The customer is looking for the solution to a need at a good price. The manufacturer must add value at a profit to continue operations and to fund future developments. The value must be created and exchanged for revenue so that both parties come out to the good. Anything less than this is temporary and fleeting.

3.1.3.3 Global product and business development

3.1.3.3.1 Small world, isn't it?

Disney Corporation dedicated an entire Magic Kingdom attraction to it, the Internet has made it an information reality, and Boeing's 777 made it easier to circumnavigate. It is indeed a small world and getting smaller everyday. What does this mean for new product development? It simply means that isolationism in one market, country, or continent is no longer possible. What your company does affects, to a degree, the global marketplace for that particular market. What the rest of the world does may indeed affect your company. Product development and market development no longer occur in a vacuum. The downside is more and more intense competition. The upside is instant access to information and data.

3.1.3.3.2 Typical foreign content of today's products

If it is not obvious that the world is getting smaller from a communications aspect, consider the foreign content of many of today's products. The foreign content has become so commonplace that several products publish it in terms of a percentage. Consider the Hyundai automobiles manufactured in the United States: they are manufactured by Americans for a Korean company. It would be an exercise just to document in detail the funds flow in terms of raw, manufactured cost and intellectual property amortization. A similar split in content occurs in cameras, digital video disks (DVDs), cellular telephones, and computers.

An interesting exercise would be to document the mileage on shipped parts in a subassembly through manufacturing to its final destination. Consider the following route a power transistor takes in getting to its final destination: the power transistor will be used in the final output stage of an electronic drive for controlling the speed of an electric motor, which is part of an elevator drive system in Algeria. The power transistor is manufactured in Japan and is shipped to a distributor in North America. That distributor ships it to a drive manufacturer in the Midwest for integration into a drive. The drive is shipped to the drive manufacturer's subsidiary in the United Kingdom as part of a stocking order. The drive is pulled from stock and shipped to Bologna, Italy for integration into the elevator control panel. The elevator control panel containing the drive is tested and shipped to the end-user, a building in Algeria!

3.1.3.3.3 Analysis of exchange rate sensitivity on foreign content

The presence of foreign content in products also adds a dimension to product costing. The actual value paid for foreign content components is subject to the changes in currency from one country to another. This can place pressure points on a product design from a cost sensitivity point of view. In addition, the supply of these products can be affected by global competition, export compliance, and political turmoil.

3.1.3.3.4 Venues for procurement

This means that the venue for procurement of the component must be carefully selected, in addition to the specification and the cost. Failure to consider these ramifications may cause loss of supply midstream or intense

cost pressure. Later in the text, we will examine a vendor profile that goes beyond parts specifications and costs.

3.1.3.3.5 Venues for manufacturing

The same can be said for manufacturing; however, an additional dimension occurs here. What happens if the local lawmakers pass legislation that affects the plant? Can your capital equipment be idled by changes in law, political pressure, or local labor initiatives? In this case, your company is at greater risk than in the procurement case because your manufacturing capacity is affected.

3.1.3.3.6 Local laws

Local laws can place greater demands on new product development because some require adherence to international standards. These may be stricter than U.S. laws and standards; consequently, the development may have to be more complex and costly.

3.1.3.3.7 Export compliance

The United States requires manufacturers to adhere to export compliance legislation. This is mostly a marketing effort to be selective in targeting certain locations to exclude from sales and shipment. Failure to comply may result in revocation of an international license, or worse. The export compliance guidelines exist to limit shipment of U.S. technology to foreign countries that may be a security threat. It can significantly affect your company's marketing effort.

3.1.3.3.8 Technology transfer

Technology transfer is generally a company decision, although the U.S. government can play a role regarding certain technologies. Some foreign governments mandate a transfer of technology to their country as part of a sale. This can adversely affect your company's market position and strength elsewhere, and the decision to transfer technology must be carefully evaluated.

3.1.3.3.9 Joint ventures

Joint ventures are complex partnerships, with differing cultures, profit motives, and taxes—know your partner's agenda. Such partnerships can be very successful or very costly. Watch where the money changes hands to determine allegiances, agendas, and motives. This will be discussed in detail in Chapter 3.2.

3.1.3.3.10 Cost roll-up spreadsheet with sensitivity

Figure 3.1.20 summarizes the effect on product factory cost impact from foreign content. As can be seen, the effect can be favorable or unfavorable depending on current market-conditions. The object point is that your company may enter into an agreement with the conditions being favorable, and lock into terms that will force cost pressures and eventual market pressure if the conditions go unfavorable. This puts your effort in a precarious position. In addition, the manufacturing venue can cause greater problems depending on the percent content of labor, local laws, and relative numerical difference in wages. This also assumes that the hours required to produce the product are uniform and homogeneous across the globe; however, they are not, because different venues have different workforce skill levels. As can be seen in Figure 3.1.13, many factors contribute to the product factory cost and, as such, it is the responsibility of the manager to account for these issues. The architecture of the product transcends the bill of material; it is affected by all of these contributing issues. The product needs to be planned out from all aspects to ensure continuity in sales and manufacturing.

3.1.3.4 Market investigation

Market investigation is so important to the long-term growth of a company that it is included in the following presentation. It is presented here from a "fit" perspective to provoke the strategic thought and planning of new product development early in the process, as it should be. In Chapter 3.2, the mechanics of marketing will be covered in detail. The objective here is to present how the concepts relate to the overall evaluation and selection process, and what one should look for in a new business development program.

3.1.3.4.1 Primary market research

There are two basic types of market research—primary and secondary. The primary type of research is one of the most direct and accurate forms available because it deals with information obtained directly from a customer. There are a variety of ways to secure primary market feedback, but most fall into the following categories: surveys, interviews, and demonstrations.

Surveys can be administered in a variety of mediums and are structured means for obtaining information. To be conclusive and effective, they need to be administered in the same way each time, with no changes in format. Interviews are a face-to-face means of gathering information directly from the respondent. These are very valuable for securing qualitative information, exploratory information, and discovery of latent needs among customers. In fact, this type of sampling can be used as research for determining how to structure a survey. Demonstrations are another type of primary market research. This type of research offers information in the form of a demonstration and solicits feedback from the respondent. These are critical to confirming features,

BOM	with labor adjustment							
***								
Description	**P/N**	**Qty**	**Cost each**	**Extended cost**	**Domestic**	**Foreign**	**Currency**	**Cost impact**
							Favorable/unfavorable @Standard cost	
Part 1	654321	2	43.25	86.5	Yes	No	1	86.50
Part 1	654320	2	0.23	0.46	Yes	No	1	0.46
Part 1	654319	2	0.48	0.96	Yes	No	1	0.96
Part 1	654318	6	4.30	25.8	Yes	No	1	25.80
Part 1	654317	1	6.20	6.2	Yes	No	1	6.20
Part 1	654316	4	7.90	31.6	Yes	No	1	31.60
Part 1	654315	8	12.30	98.4	Yes	No	1	98.40
Part 1	654314	10	8.75	87.5	Yes	No	1	87.50
Part 1	654313	3	3.40	10.2	No	Yes	0.99	10.098
Part 1	654312	7	1.25	8.75	No	Yes	0.98	8.575
Part 1	654311	5	1.65	8.25	No	Yes	1.12	9.24
Part 1	654310	2	2.40	4.8	No	Yes	1.12	5.376
Part 1	654309	1	2.75	2.75	No	Yes	1.12	3.08
Part 1	654308	3	3.45	10.35	No	Yes	1.12	11.592
Part 1	654307	4	6.45	25.8	No	Yes	1.12	28.896
Part 1	654306	5	4.12	20.6	No	Yes	1.12	23.072
Part 1	654305	7	2.34	16.38	No	Yes	1.12	18.3456
Part 1	654304	8	45.70	365.6	No	Yes	1.12	409.472
Labor		12	35.5	426	Yes		1.05	447.30
Burden		1.75		745.5	Yes		1	782.78
Miscellaneous expenses		100		100	Yes		1	100
			Total cost	2,082.40			**Adjusted cost**	2,195.24
							Impact	112.84
							% Impact	5.42

Fig. 3.1.20 Foreign content sensitivity analysis.

benefits, and acceptability of a product to a customer. Each type of primary market research has its advantages, and each needs to be used in different aspects of the information-gathering process.

3.1.3.4.2 Secondary market research

Secondary market research is characterized by an indirect means for gathering information from readily available sources. Typically, research of this type consists of, but is not limited to, the following means for obtaining information.

3.1.3.4.2.1 Trade shows

The trade show is an excellent means for obtaining product information as well as determining the scope of companies in any one field of endeavor. In addition to

enabling personal interviewing, observation, and literature search, trade shows typically publish a program that lists companies by product type or process. These are typically segmented several ways to allow a quick view of the market.

3.1.3.4.2.2 Trade literature search

The literature search is an ongoing activity to keep in tune with the marketplace to see who is promoting what to whom. It is an excellent means for gathering information for a future competitive comparison.

3.1.3.4.2.3 Articles

Articles are excellent for documenting trends in the marketplace. They are, by virtue of their authors, expert sources for further information and perspectives on the marketplace. They are very useful to the industry novices who need to familiarize themselves with a product area.

3.1.3.4.2.4 News pieces

News pieces are timely information about the actions of the leaders in the marketplace in a given area. They tell you who is active in an area and to what extent, and what boundaries are being pushed.

3.1.3.4.2.5 Observation of trends

This is an internally generated activity that looks at what is happening in the marketplace from period to period. By looking at the historical activity, one can extrapolate trends to the future. This will be an invaluable exercise when producing a product evolution flow chart.

3.1.3.4.2.6 Financial reports

Financial reports offer a wealth of information about competitors and, as such, collectively about the market. These can be monitored for information and financial information to get a picture of the finances, the principals, and the capitalization of a competitor. Monitoring of the financial markets and their news can provide insight into acquisitions and mergers affecting business development.

3.1.3.4.2.7 Textbooks

Textbooks offer the novice background information on products or services and technology. These sources can be useful to understand elements of the technology.

3.1.3.4.2.8 Patent search

Patents are available for search to determine the state of the art in a field of endeavor, as well as to determine your company's position in the state of the art. It can also be a good source of competitive information on technology.

3.1.3.4.2.9 Lost sales reports

These internally generated documents can be helpful in determining a competitor's ability to use their product to secure business. By evaluating their technology and your company's performance against them, these reports can be an invaluable source of information.

It almost goes without saying that the Internet is of the most prolific means of information-gathering about companies, products, competitors, pricing, trends, and countless other data. It can be an invaluable resource for immediate data-gathering if the data is available and accurate.

3.1.3.4.3 Interviewing the customer

Although previously mentioned as one of the primary means for market investigation, interviewing is a valuable technique because, if approached properly and under the right conditions, the customer is the most accurate means for assessing your company and its standing in the marketplace. If you listen and attempt to internalize what the customers are telling you in interviews, they will communicate most (if not all) of the qualitative data you will need.

When conducting an interview, it's best to first review the customer's company operations. This will give you a feel for the operation and how this customer makes money. It will also give you a perspective on how your company and its products will fit into the customer's overall plans, as well as determine the legitimacy of the respondent's comments by materiality. The interview will allow you to discover latent needs that the customer may not readily know they have. This generates opportunities for new product ideas.

It's best to have samples, models, examples, and props to enhance the exchange and draw out responses. Do not be averse to broaching issues that may be controversial (pertaining to the subject matter at hand), because this will give perspective on the attention level of the respondent and his or her biases, and will yield perspective on other comments.

3.1.3.4.4 Engaging the customer

As part of a customer visit, you will want to engage the customer in a frank discussion about their business, its driving forces, its vulnerability, and how they navigate the competitive threats to find success. This discussion should result in symbiosis: to have a real appreciation for the customer and their needs, it becomes necessary to understand their business.

The process of customer engagement consists of three basic themes:

1. First: listen to the customer. You have created the venue for an exchange, so take the time and energy to really listen.
2. The customer has an issue or problem to resolve. Recognize it!

3. This is why you are here—to pick up on subtleties and act accordingly.
   - The customer is not predisposed to your solution, so do not try to convince them at this time.
   - The best evaluation of competitive standing is to remove yourself from the sales arena and place yourself in the customer's perspective.
4. Second: move the interchange to a process of discovery.
   - Conduct a dialogue, not a sales presentation!
   - The interchange should be a mutual process of discovery.
   - Probe for conditions, customer alternatives, acceptance scenarios, and prerequisites.
   - Understand and frame the customer's business, alternatives, decision breakpoints, and their product pathways.
   - Determine and reduce the information to a flow chart and breakpoints.
5. Finally: look for the hidden opportunity for your company.
   - How can your company create a product or service that enhances the customer's ability to pursue their market more effectively?
   - Discovery should produce the customer's hidden opportunities, competitive situation, and threats to securing the business.

To get started, consider the following as a framework for an initial discussion. These are questions you may want to use to engage the customer. These discussion interrogatories are open-ended and require thought to create mutual understanding.

1. What are the areas in your organization that currently fall short of the skill set and performance index required to achieve your business vision? Who, what, where, how, and why? (This allows for an open-ended forum between you and the respondent. It allows you to draw out complaints initially, and then to get specific ideas onto the table for discussion. Share your company's experiences, also.)
2. From a historical perspective, what are you doing about them? (This determines the commitment level of the management and the organization to change, and to the strategic vision.)
3. What is your business vision? (This is a good test to determine whether the organization has established a vision and whether it has filtered down through the organization.)
4. What operational or product changes are you making to achieve this result? (Again, this determines the effectiveness of the management structure to effect change toward a specific goal.)
5. What do you need to make your operations easier and more efficient, to increase output, and to exercise better control? (This open-ended discussion allows the respondent to identify any latent needs in the organization.)
6. If you were asked to double your output with the same resources, what would you expect to be required to achieve this? (This removes the specter of the status quo and forces the respondent to push performance of the organization. It helps the respondent to identify future needs.)
7. What is currently considered impossible or impractical, but if you had access to it, you could increase your effectiveness? (This question allows the respondent to dream of a future workplace and identify latent needs to achieve it.)
8. What do you look for in a supplier? (This identifies the customer's expectations.)
9. How can we become a preferred supplier to your company? (This determines whether you have a chance at the business and the account.)
10. What are your critical competitive pressures? (This allows you to determine a way to partner with them.)

The key to identifying the potential new product is to synergistically combine all of these ideas and requirements into the tangible improvement of the next-generation product.

3.1.3.4.5 Surveys and customer feedback

Surveys and customer feedback should not be conducted only at the launch of a program. These need to be an integral part of the process of refinement, product evolution, and subsequent product developments. Never get too far away from the customer because this will lull an organization into complacency or, worse, remove the feedback loop in the process.

3.1.3.4.6 Extracting the opportunity within a customer complaint

The customer grants a significant favor to you by complaining. Your company sold them something that fell short of their expectations, and they thought enough of it to complain. We as market and product development professionals may not like or want to listen to the complaint; however, there is a lot to be mined out of one. In effect, the customer is telling you what you did wrong. If you evaluate the math of the situation, a single complaint is worth quite a bit to the organization if you are going to do something about it. If you view it as a nuisance, then it has no value; however, if you plan on changing to meet the customer's expectations, then it has great value. In other words, how many other customers

who didn't complain would have to stop buying from you for you to lose enough market share to notice? How many customers would you have to survey to find out what this one complaining customer is telling you directly? Clearly, it is a valuable means of feedback.

3.1.3.4.7 Exchange of knowledge and experiences

During your discussion, it is critical to establish an open and free exchange of ideas, problems, and opportunities between the two companies. Establish a spirit of cooperation at the management level and drive it through the organization. Create the atmosphere of going after a market opportunity together, with each entity contributing a portion of the talent and effort to get there. Constantly look for how the two companies can profit from the market segment, since very few companies are absolute leaders in a market with a position strong enough to dominate and affect control of the segment by themselves.

3.1.3.4.8 Establishing rapport, expertise, assistance, improvement, and authority

This exchange should establish a lifelong rapport between the personnel of the two businesses that will transcend any minor market fluctuation, upset, or competitive thrust. Your actions need to position you as an indispensable partner who is integral to your customer's plans. Make sure to plug the customer (now your partner) into your organization. Balance the sacrifice you will make in availing the customer of your resources with the positioning of indispensability within their organization.

3.1.3.4.9 Correlating the data

Examine the responses carefully and group the data in some meaningful way to draw product conclusions from it. Look for product configuration patterns in the summary of data so that product planning can serve the market with the proper product.

Do the data received make sense with other information gathered? Have the data been corroborated in some way to substantiate the results and accurately draw conclusions? Did you notice any conflicting data that would require more research? Summarize the data in some kind of a spreadsheet to allow viewing and sorting. Organize it for eventual presentation to management. The results should tell management the story of the opportunity.

3.1.3.4.10 Testing the data

Finally, test the results and the data supporting the conclusions by visualizing the product available for sale today. Evaluate how it will sell in the marketplace today. Would it sell? Why or why not? Visualize the product selling in the future, project the future market needs, and (after development time has elapsed) determine whether it will still sell or whether you would have lost the window of opportunity. Do the results of this reinforce that you are targeting the right generation of product?

3.1.4 Summary

This chapter had several objectives for the reader to achieve. The first theme was a discussion on how product and business development relate to the economy. The thematic thrust of the chapter was to draw the relationship between product development as the business objective and to examine it from perspectives external as well as internal to the organization. It was also designed to give a qualitative feel for the different steps in the process of new product development. As the chapter unfolded, the reader learned the various aspects involved with the process. There was a discussion of the new product opportunity's fit in the organization and how it ties into the strategy of the organization.

There was also a discussion of the directed purpose of research and the function of development, how to leverage efforts, and create continuity. Business development programs should each contribute to a corporate goal, and the market investigation is used to refine the goals and the opportunities.

The chapter should also give the reader an appreciation of the interrelationship of the various disciplines within the enterprise that must function in concert to successfully implement a program. Some of the material presented here in broad overview will be discussed in a more detailed manner in later chapters. However, it was presented here for continuity in outlining the process of taking an idea from inception to completion. Some of the processes are discussed in detain in Chapter 3.2.

This page is intentionally left blank

Chapter 3.2

3.2

The market opportunity

Marc Annacchino

3.2.1 The marketplace within our world

3.2.1.1 The faster-changing world

The world is an ever-changing place, and the changes that we are currently undergoing are occurring at an ever-increasing pace with wider swings and larger differences. The process of new product development must factor in this uncertainty. Product iterations are coming faster and demand for changes is more prevalent than before. Not all of these changes are necessarily good, nor do they necessarily affect a program; however, they must be recognized, evaluated, and factored into a program (and possible modifications of that program) for it to be successful in the marketplace.

In terms of evaluating humankind's knowledge base, each discipline needs to be examined for its accumulated knowledge and real-time progress. Many disciplines increase their body of knowledge in very short periods of time. Since this accelerates with the experience and the passing of time, the amount of change in certain disciplines is enormous. Consider, for a moment, some of these fast-changing technologies.

3.2.1.1.1 Computer technology

3.2.1.1.1.1 Wireless technology

The following looks at data that is somewhat representative of the growth of knowledge, so as to flavor the discussion of how new product development must operate in these fast-changing times.

The U.S. patent system can be considered a measure of knowledge by certain standards. All else being equal in terms of novelty, technological innovation, and application, the number of patent applications and issues is a measure of a level of knowledge development. Figure 3.2.1

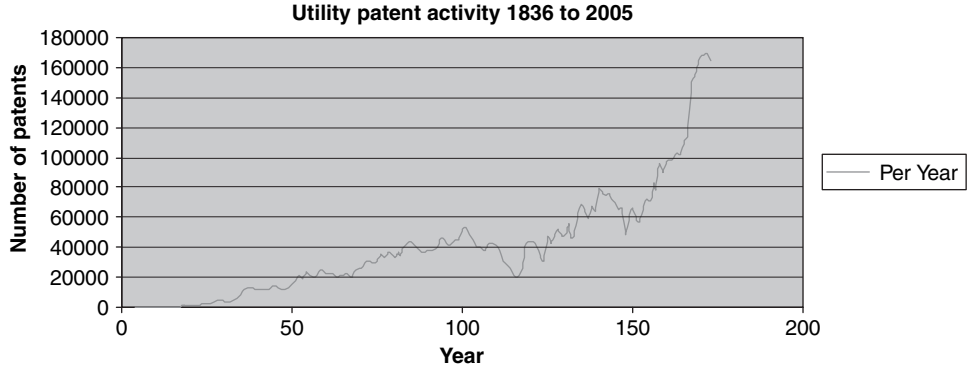


Fig. 3.2.1 Patent activity trends.

The Pursuit of New Product Development; ISBN: 9780750679930
Copyright © 2007 Elsevier Inc. All rights of reproduction, in any form, reserved.

is a graph showing the number of U.S. patent activities from the period from 1836 to 2005. This 169-year period in our history shows a steady increase of the number of applications and patents issued.

3.2.1.1.1.2 Population and time

Another measure of humankind's knowledge level is attributable to the sheer growth in the population. More people, more studies, more ideas, more creativity, and more applied knowledge development time ultimately generate a larger knowledge base from which to draw. In addition to these, the advent of computers and Internet communication schemes accelerate information transfer to lightening speed.

3.2.1.1.2 Effects of faster doubling of human knowledge

3.2.1.1.2.1 Faster pace of life

The effects of this increase in humankind's knowledge are creating a faster pace of life. Information flows faster, with more sources of input. However, there is also greater inaccuracy in the information, causing false starts and changes in direction. Many aspects of life in modern times have become competitive from a quantity rather than a quality point of view. People are racing toward an ever-more-mobile target. The result is a faster pace with questionable value added in many of the transactions.

3.2.1.1.2.2 Increasing pressure

This faster pace is causing increased pressure on our society. We can see the pressure manifested in loss of patience, wider swings in violence and crime, and loss of reflection and relaxation in people. In many cases this pressure, unnatural by its extreme, is causing much anger. The pace is also accelerating as time progresses. There is a higher level of learning among people, with requirements and demands that are constantly increasing.

3.2.1.1.2.3 Disconnection index

With all this excessive movement forward, the danger of becoming disconnected is increasing. This so-called disconnection index is a measure of a person's lack of keeping pace with the changes around himself or herself. Fifty years ago a person could have conceivably left the Earth for 10 years, returned, and essentially picked up where they left off. For example, between 1840 and 1850 the amount of change in life was rather small and not easily discernible, compared to the changes from 1990 to 2000. A person leaving the Earth for 10 years in 1990 and returning in 2000 would find sweeping changes in foreign policy, security of the European continent, vast changes in communications, computer power, and availability of and access to information.

3.2.1.1.2.4 Aging population

Unfortunately, these changes are being thrust onto the populous as if people were homogeneous and equal in terms of their ability to assimilate, absorb, digest, and accept these "improvements."

For example, how many people of the baby-boom generation can easily program a digital video disk (DVD) recorder without instructions? How many elderly people feel comfortable with the seven-segment readout for numbers? How many people enjoy reprogramming all of the clocks in the house after a momentary power outage? As more products get more sophisticated and the population ages, this disconnection index will grow naturally. From a product development perspective, it will have to be addressed in order to appeal to the mass market with a single product platform.

3.2.1.1.2.5 Technical focus versus value focus

Another change that has been taking place is that young people, in their quest for keeping up with these changes, have become technology-focused rather than value-focused. This fact is probably little observed as we lead our busy lives, but it is crucial to recognize this in understanding the demographic changes, the needs, the preferences, and the general attitudes of the different segments of our population. Although it might seem that this is only occurring in the United States (because it is more noticeable here), these changes are occurring all over the world, just at different rates.

3.2.1.2 The seasons of change

It would seem that each era spawns a new generation of knowledge that emerges, grows, flourishes, declines, and dies off to give way to the next generation. This closely parallels the business development cycle in that the product emerges, gains acceptance, is copied by competitors (indicating, oddly enough, acceptance in the plurality), flourishes, and then eventually dies off, making way for the next generation of product.

It's important to understand how knowledge levels affect our society and its development. Just as the steam engine was a technological revolution that drove far-reaching changes in industrial society, the technological revolution that facilitated knowledge transmission will revolutionize and accelerate drastic changes in post-industrial society. The caveat for business and product development is that one must understand that change and growth are an inevitable part of the cycle of life as well as the business and product development process.

Development patterns exist in our civilization as time progresses. They are often hard to wrestle with on a day-to-day basis but, when considered within the historical perspective, they are readily apparent. These patterns

drastically affect the marketplace and the market's acceptance of new products.

One such development pattern has a time base of approximately 80 years, or a single human being's lifespan. The pattern seems to be broken up into four distinct segments, as follows:

1. High phase
2. Awakening phase
3. Unraveling phase
4. Crisis phase

These four segments can be likened to the seasons of change, namely spring, summer, fall, and winter, respectively. In the high phase (spring), a new beginning emerges. This usually follows a crisis of some sort and, as society emerges, there are unbridled optimism and growth. Society expands, prosperity increases, civic responsibility increases, and advancement occurs in most aspects. This phase is followed by an awakening phase in which society begins to question itself and its direction. Spiritual agendas begin to dominate the social landscape. Scientific agendas begin to lose influence, public order declines, and society's general level of enthusiasm begins to wane. War and conflict, driven by newfound idealisms, break out. The next phase is the unraveling phase. This phase emulates the tone of its title. Society begins to acquiesce, public trust declines, guilt begins to dominate the moral landscape, and general attitudes of pessimism abound. People "feel" but no longer "do." Wars and other conflicts are fought with moral fervor and debates on idealism and morality abound. There are "liberating" forces at work, fragmenting society as they have had their fill of "rebirth and soul-searching." It has low resistance to external forces, placing it in extremis.

As can be expected, this phase leads into the crisis phase. This phase separates the conversation and soul-searching of the previous phases into actual conflict. The threats become real, there is a sense of public urgency, and the family unit strengthens. War and conflict are prosecuted furiously and with purpose. The government governs and extraneous laws are shunted aside to make way to meet the crisis head on. Crime diminishes and optimism increases as the society shifts from individualism to national or civic purpose. This phase then wanes and ushers in the next high phase. Each phase seems to average 20 years in length.

For a historical perspective, World Wars I and II represented international crises where national and international purpose drove many changes in society. The government governed and created a Promethean war machine. Society shifted to a more unified entity with purpose.

After World War II there was a tranquil period of growth or "high" in the 1950s, followed by the beginning of the awakening in the late 1960s and early 1970s. The awakening evolved into the unraveling of the 1980s and that brings us to the eve of the next crisis. The 1950s and early 1960s were a tranquil time when growth and prosperity abounded. This high spawned the moon venture, consumerism, freedom, and enthusiasm.

The late 1960s and early 1970s ushered in the awakening, a period of searching, questioning, and shift to individualism. The unraveling continues until the next crisis. In recent times, elements of all four phases are evident in different locations. This would suggest that many time-shifted phases are occurring simultaneously in different parts of the world and within different parts of a same location. Accordingly, products serving these societies must be targeted to serve the phase at each place and time.

The history of cycles and trends needs to be understood and considered within the framework of new product development. Although some products may be achievable, they are not right for the times. For example, would instant access, connectivity, and the freedom of the Internet have been successful in the 1940s and 1950s with the Red scares and security breaches? Would products promoting individualism and personal freedom be successful in an era when national and civic purposes dominate the society? Or, more basically, would fuel efficiency and environment issues embodied in automobiles be accepted in an era where performance and styling dominate the taste of the consumer? These are just a few examples of why product planners need to understand the periods we are in and factor this into the product and company strategy. In a crisis period, products need to rapidly get practical.

The market is looking for products that are functional, affordable, able to be manufactured with available materials, and reliable. Technology generally needs to be proven because the market will have little tolerance for ushering in new technology. After all, they are navigating crisis. In a period following the crisis or high, the market is much more accepting of new things, implementation of new technology, and new products and markets. Regardless of whether one accepts the model or the time line, one must accept that the products marketed within a phase must fit the acceptance patterns of that phase.

3.2.2 Marketplace dynamics

3.2.2.1 It's a fast-changing world

The world economy shows a nonhomogeneity of needs and satisfactions all over the world. There are different peoples at different stages with different politics and different goals. The new world order is a work in process. Some societies are way up on the hierarchy while others are struggling at the base for basic needs. This disparity

drives product development directions differently. For example, in South Africa a domicile may require 150 watts of electrical power. In contrast, the average U.S. requirement is 2.0 kW, spiking to 18.0 kW.

Disparity in the product evolution and infrastructure implementation will change the assumptions. For example, telecom is evolving to a wireless system in the United States after a landline infrastructure had been installed for many years, whereas in other parts of the world telecom will start off with wireless infrastructure immediately and landlines will probably never be implemented. This will change the mix of telecom products, the product evolution, and the obsolescence, and legacy pass-downs of used equipment will evolve.

The same thing may happen with distributed and centralized energy! Once the bastions of a society, centralized electrical power generation may give way to decentralized, distributed electrical power where products enabling an existing infrastructure and support system of fuel delivery allow immediate implementation.

3.2.2.1.1 Reduced ability to cope with changes

For technology to have a long-lasting benefit to humankind, it must be accompanied by the wisdom to use it. Failure of the wisdom element in the equation results in abuses, misapplication, and eventual rejection. With many of the recent technological advances, we have seen the accompanying moral dilemmas in the medical and health-related fields. Failure to garner the wisdom to utilize the technology can result in an inability to cope with the changes that it brings.

3.2.2.1.2 Diverse value systems

The diversity of the world results in the diversity of values all over the world. We would like to consider them to be homogeneous, but they are not. Accordingly, the development of business and products must factor in these value systems and how they impact acceptance.

3.2.2.2 Market characterization

3.2.2.2.1 Market maturity

One of the most valuable exercises is to determine the level of market maturity that exists in the specific market under consideration. Depending on the degree of maturity, the market can be low-growth or high-growth. A mature market will be flat or no-growth. Conversely, an emerging market will be more growth-oriented. Figure 3.2.2 illustrates the difference and will be used as a basis for illustrating the market share and profit potential of each.

As shown in Figure 3.2.2, the differences in emerging and mature markets manifest themselves differently.

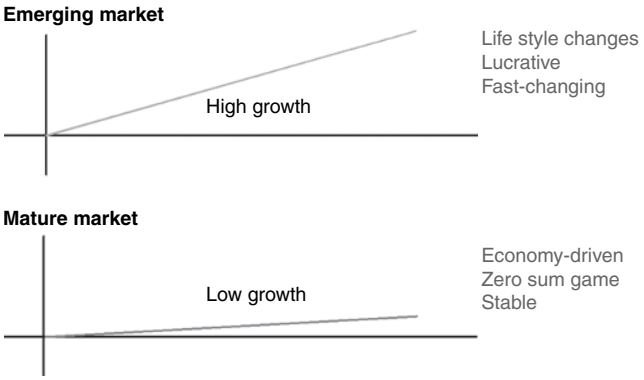


Fig. 3.2.2 Emerging versus mature markets.

This affects the business development plan. The high-growth markets have more opportunity associated with them. They generally effect lifestyle changes in the customers. An example of this is the automobile industry in the early 1900s. It was a growth market that caused sweeping changes in the lifestyles of the customers; it was very lucrative and fast-changing, which challenged the manufacturers to keep up with the changing market demand.

A low-growth market has less opportunity, is more price sensitive, and can be thought of as a zero sum game. What market share you increase is obtained only from competitors, not from a growing number of customers using the new product or technology. Company growth is determined by swings in the economy, more than other factors, and there is stability in the marketplace.

Figure 3.2.3 shows the impact on the companies participating in these markets. The benefit shown in the emerging market is that the company can more easily gain absolute shipment dollar revenue through increasing its market share in an increasing market than in the low-growth markets.

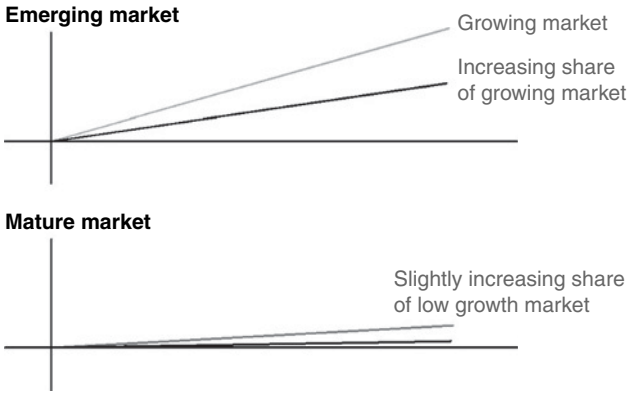


Fig. 3.2.3 ■

In either case, as part of the business plan there should be a clear understanding of what the market dynamics are and how you intend to participate in them.

3.2.2.2.2 Life cycle and market cycle implications of profit

Figure 3.2.4 illustrates the specific impact on a company operating in these diverse markets. In the emerging market, the economic benefit to the firm in dollars of margin far outweigh what can be generated in a mature market.

These examples are for illustrative purposes only, and the absolute values or percentages are somewhat exaggerated for effect in illustrating the difference. Also, the market size and revenues have been scaled for the example. However, the analysis shows the vast difference in actual gross margin between the two markets. In the emerging market, there is a growing percentage of market share to work with. In the mature market, the behavior is relatively fixed. The graphs show the difference.

Another critical point to remember is that markets are not exclusively either mature or emerging. They start out as emerging and move into mature markets. Therefore, the business plan developed for new products and business has to factor this into the development and account for the transformation.

Figure 3.2.5 indicates how this may look in terms of market growth over time. The market growth starts out small and increases dramatically, and then remains somewhat constant up to a peak growth rate. The growth rate then diminishes to essentially zero, where the market size remains constant. This assumes, of course, a normal distribution and behavior without overt manipulation by any one dominant competitor who might be large enough to unilaterally effect a change in the marketplace.

3.2.2.2.3 Low-cost producer

There are certain strategies that a company can use to navigate the changes in markets and use them to their advantage. In the graphs in Figure 3.2.6, the various behaviors are presented in the framework of the various markets to show why it is desirable—from a survival point of view—to be a low-cost producer by the time the market has matured.

The graphs in Figure 3.2.6 show a contrast between a mature market and an emerging market implementing older, less cost-effective technology and newer, cost-effective, state-of-the-art technology. The behavior of prices and costs tells the story of company profits.

3.2.2.2.4 Leading-edge technology

In the first scenario (a mature market being addressed with older technology), the cost remains relatively flat while the prices are driven down. In the second example (a mature market with state-of-the-art technology), the decreasing prices are offset by decreasing costs. The third example shows that in a growth market the prices remain stable while the demand for functionality can quickly outstrip the older technology's ability to supply the market demand. In this example, sales cease after a time because the older technology is limited and cannot satisfy the market needs. The fourth and final example (a growth market with state-of-the-art technology), shows prices slightly declining with costs declining at a greater rate, thus preserving the margin.

As shown in Figure 3.2.6, the best overall strategy a company can take is to be at the leading edge of technology in order to keep current with the market requirements for functionality and price, as well as cost containment, to maximize profits. It simply puts your company ahead of the competition initially so that you can establish and maintain a market position.

The danger is in allowing the company to be placed in a squeeze situation wherein the market is demanding a low-cost, highly functional design and you have a high-cost, barely functional design to offer. There will be lackluster sales and minimal profits, resulting in thin profit margins and eventual inability to reinvest properly.

3.2.3 Product concept synthesis

3.2.3.1 The business concept embodied on the product idea

3.2.3.1.1 The better mousetrap

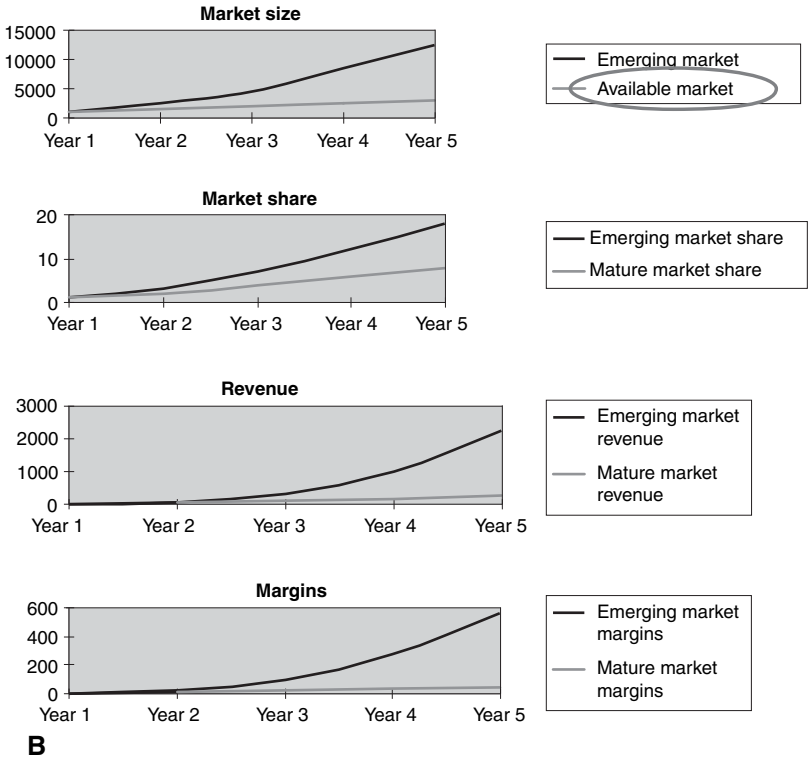
"Design a better mousetrap and the world will beat a path to your door." This has been the battle cry for new product developers for many decades. However, there are increasing requirements for success today, such as: Who else makes mousetraps? What must our mousetrap do to induce people to buy it over others? Are we a low-cost mousetrap producer or are we a niche mousetrap player? Have we complied with the new mousetrap standards? What steps have we documented to ensure that humans will not get injured using our mousetraps? Is there a significant influx of foreign mousetrap manufacturers encroaching on our market? Are there new ways to functionally displace mousetraps, such as high-frequency deterrents? All of these issues make the age-old adage now multidimensional.

3.2.3.1.2 Most systems generate the idea and fail to encompass a plan

Most systems for generating new product ideas focus on the idea, the market opportunity, and the timing. The concept for a new product must encompass the plan to

Growth patterns					
	Year 1	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Emerging market					
Available market (000's) Dollars	1,000	2,500	4,500	8,500	12,500
Market share (%)	1	3	7	12	18
Revenue (000's) dollars	10	75	315	1,020	2,250
Cost of sales	0.6	0.65	0.7	0.725	0.75
Margin (000's) dollars	4	26.25	94.5	280.5	562.5
Mature market					
Available market (000's) Dollars	1,000	1,500	2,000	2,500	3,000
Market share (%)	1	2	4	6	8
Revenue (000's) dollars	10	30	80	150	240
Cost of sales	0.7	0.725	0.75	0.775	0.8
Margin (000's) dollars	3	8.25	20	33.75	48

A



B

Fig. 3.2.4 (A) Growth pattern comparison. (B) Growth patterns.

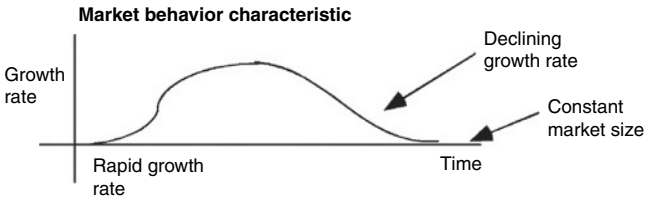


Fig. 3.2.5 ■

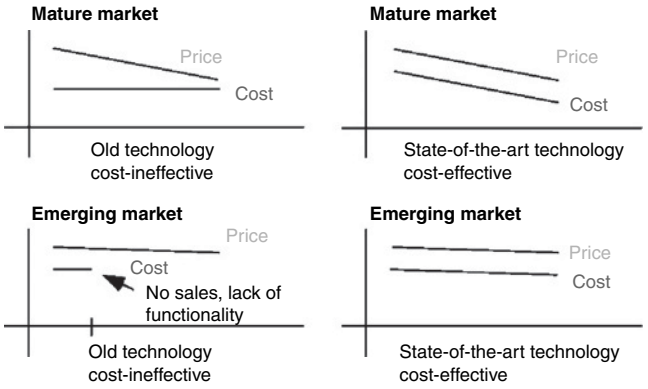


Fig. 3.2.6 Market aging characteristics.

carry it off effectively. The business concept must be entwined with the product concept. Each will draw on the other in the development phase. Each contributes to the other's progress.

3.2.3.1.3 Tie the business operations to the idea

The manager needs to visualize the business operations as the product is being outlined. This visualization will assist in defining product features and configuration that will aid in future business operations. For example, the demographic data may indicate a number of versions of a product; however, the business operations—manufacturing, procurement, and inventory control—require a finite, practical number of versions that are more manageable.

3.2.3.1.4 Critically look at the business of the product

It is essential to critically look at the business of the product. Where will the money change hands? Why will the customer be willing to part with money for the product? How will the transaction take place? Provisions and steps must be taken to address these issues to have an effective launch and program. As an example, the electric car is an age-old idea which, with the advent of computerization, modern batteries, electronic drive technology, and improved internal combustion engine technology, and the hybrid vehicle (a cousin to the electric car), is becoming a market reality. If a hybrid electric car is designed, produced, and introduced into the mass market, there are latent infrastructure issues to deal with. Issues such as safe disposal, battery remanufacturing, financing, and supply/demand changes could affect the price of conventional fuel and thus could affect operating costs. Operating cost changes could affect market acceptance. It is not just a great product that makes great businesses. It is also the infrastructure that supports the product that must be integrated.

3.2.3.1.5 Tie it together

The saying goes, "To the victor go the spoils." So, too, with the business of business and product development. The program that covers all or most of the bases to ensure success will most likely be successful. This is because the manager asked those hard questions of the program, rather than ignoring the issues altogether only to find the marketplace asking the same questions and displaying little patience for lack of planning.

3.2.3.2 Solving the customer's problem in the product

3.2.3.2.1 Customers purchase perceived value of the solution to a problem

The buying decision is based, in part, on the perceived value of the solution to the customer's problem. The customer buys a solution; the manufacturer constructs a bill of materials and processes to assemble it into a product. It is therefore a requirement for successful marketing, design, and sales to embody the solution into the product. The value placed on it is generally governed by the customers. They will determine the amount paid based on this initial perception, and subsequent follow-up sales will be influenced by experience factors that may weigh the value differently.

3.2.3.2.2 How does this new product development solve a customer's need or problem?

This is the basis for a marketing plan. Effectively answering this question will directly tie the customers' needs into the product concept. The pathway is the sales presentation. It relates to the customers and how the solution to their problem or need is embodied in the product your company is selling them. As we know, different issues affect classes of customers differently. What is a critical buying decision for one customer is not necessarily all that important for another. The market must be broken down to effectively create the product's architecture so that the lion's share of the available

market can be addressed with a basic platform design. This platform design allows a wide variety of features and a feature gradient to be incorporated in a basic unit, with options to scale up performance or packaging of values for the customer.

3.2.3.2.3 Under what conditions does the customer commit?

What is the universal model for customer acceptance? The answer to this question is as complex as the number of different customers there are in the marketplace. The arithmetic that takes place in the mind of the purchaser varies with situation, need, time pressure, alternatives, and many other factors.

For example, under normal circumstances a cool glass of water is little more than an act of kindness, worth a generous tip in a restaurant. Under conditions that are much starker and manifested by shortage, a cool glass of water can be worth considerably more if the need is great enough. Certainly, in a desperate situation one would be willing to pay considerably more money for a less-than-cool, less-than-clear glass of water than under normal circumstances. What would be the quality or price gradient in this example?

The operative lesson in the example is to understand and know the customers' conditions and frame of mind when they make that buying decision. Failure to understand and plan for it properly may position your company with product inventory and wondering why customers aren't buying.

An additional means for gathering data on the customer buying-decision dynamics is to monitor competition. How are they selling products? What is their pitch and is it effective? Gather and analyze brochures, data sheets, marketing literature, lost sales reports, and field sales reports to get a feel for their strategy and tactics.

3.2.3.2.4 How can this new product development fit the demographic conditions?

The marketing assessment outlines the market segments (i.e., that group of customers who have identified features, needs, and preferences that are similar). The manager needs to break down the market opportunities into their component parts so they may be reassembled into product versions to be used to address the market. What version of the basic platform will be used where, and why? Will it offer a competitive advantage? In a more macro sense, what will the product mix consist of? What will it cost and how will it be priced? These decisions need to be made correctly early in the process. A sensitivity analysis needs to be conducted on the cost and the price. If it's expected that the price will lower after introduction and competitive response, this new price needs to be factored in to see if there are any latent pressure points in profitability due to a skewed mix.

3.2.3.3 Idea evaluation within the framework of the business

3.2.3.3.1 Creating a system for objective evaluation

To begin to objectively evaluate a wide variety of diverse new product ideas and business opportunities, it becomes desirable to have criteria. These criteria remove the emotion from the evaluation and allow different personnel in the organization to do the evaluation. Each idea then must be tested against the criteria in the same way, with the same questions.

3.2.3.3.2 Setting up the criteria

One of the challenges is in setting up the criteria. What questions are important? What are the company's pressure points? What are things to avoid? For example, is the opportunity dependent on the development of technology that requires a few selected people? Are they available within the time frame you desire? If they leave the project, what means for recovery do you have? Given that this is an example of a pressure point as it relates to personnel, there may be a host of other pressure points that affect the selection. As a rigorous part of the business development plan, it is important to develop these criteria customized for your specific business.

3.2.3.3.3 Testing each idea

Each idea should be tested against the criteria, as previously stated. Figure 3.2.7 is a format that can be used to evaluate an idea. It is structured with a weighted sum arrangement of criteria. Each criterion is listed as a separate item. Next to each criterion is a preference assessment of that specific criterion. The preference assessment relates directly back to the strategic initiative. Does the business desire that attribute in a new product development? The range is from -1.0 to $+1.0$. Next to this preference value is the assessment of how the product relates to that specific criteria. The columns are multiplied and summed to get the weighted value.

By observing the individual values posted for each criterion or attribute, one can learn a fair amount about how the organization will absorb and implement the program. As can be seen, the organization has a high desire to select an opportunity that has sales of over $2.0 million over four years. They are highly technical and are not concerned about absorbing the technology. They do, however, have an aversion to field service and support, perhaps due to a weak field organization. This criterion shows up as a negative value. Consequently, if

Evaluation criteria for new product development					
Evaluation criteria	**Company desired weight (–1.0–1.0)**	**Actual score (0.1–1.0)**	**Product 1 weighted score**	**Actual score (0.1–1.0)**	**Product 2 weighted score**
Sales > $500k per year for four years	1	0.9	0.9	0.9	0.9
Existing sales channel secure orders	0.75	0.85	0.6375	0.6	0.45
Absorb the technology and design the unit	0.85	0.45	0.3825	0.75	0.6375
Training, service, and support in the field	–0.25	0.2	–0.05	0.75	–0.1875
Sold to a variety of customers	0.1	0.6	0.06	0.8	0.08
Already our customer base	0.9	0.7	0.63	0.5	0.45
Leverage to another product line	0.5	0.35	0.175	0.2	0.1
Projected margin over 50%	0.65	0.4	0.26	0.6	0.39
Installation is required	–0.5	0.95	–0.475	0.85	–0.425
Total			2.520		2.395

Fig. 3.2.7 Evaluation criteria.

the product being evaluated has a high dependence on this for success, multiplying the preference and product assessment will be negative. When summed up with the other criteria, these will detract from the overall weighted value. This analysis simply serves as a numerical evaluation (which is objective) rather than talking yourself into the assumption that the product or business won't need that much field organization support.

Another aspect is to admit that the criterion of field support is not one that your company is good at, and to reexamine your commitment in that area. In addition, perhaps the product and the sales channel could be designed such that a field support organization won't be required.

As with any weighted summation, the absolute value of the sum does not mean much. When two programs are set side by side, the sums can be compared on a numerical basis to select the best alternative. This can be a valuable tool, given two conditions:

1. Be very careful when setting up the company criteria. Make sure they are an accurate reflection of the desire and capacity of the company.
2. Be accurate and honest when evaluating an opportunity against these criteria.

If you are diligent about these two items, this will serve as a useful tool. Figure 3.2.7 presented here has fixed values and fixed criteria.

3.2.3.3.4 Consistency

Consistency is one of the key benefits in using this objective type of evaluation. As times and personnel change, it becomes a challenge to establish and maintain consistency in the organization. This exists in all aspects of the organization, not just in the identification of new product opportunities. It is most important to be consistent in the new product area, however, since this area drives all future movements. Wild swings in strategy and tactics will misdirect the organization, thus removing whatever shred of leverage it may have had and rendering it ineffective.

3.2.3.3.5 Measure of ongoing effectiveness

As the system is put in place and used for some time, it is important to measure its effectiveness. What ideas were brought in, how were they evaluated, and what were the results in selection? What degree of success was obtained in pursuing the market? Are the evaluation criteria still the right ones? Are they consistent with the business

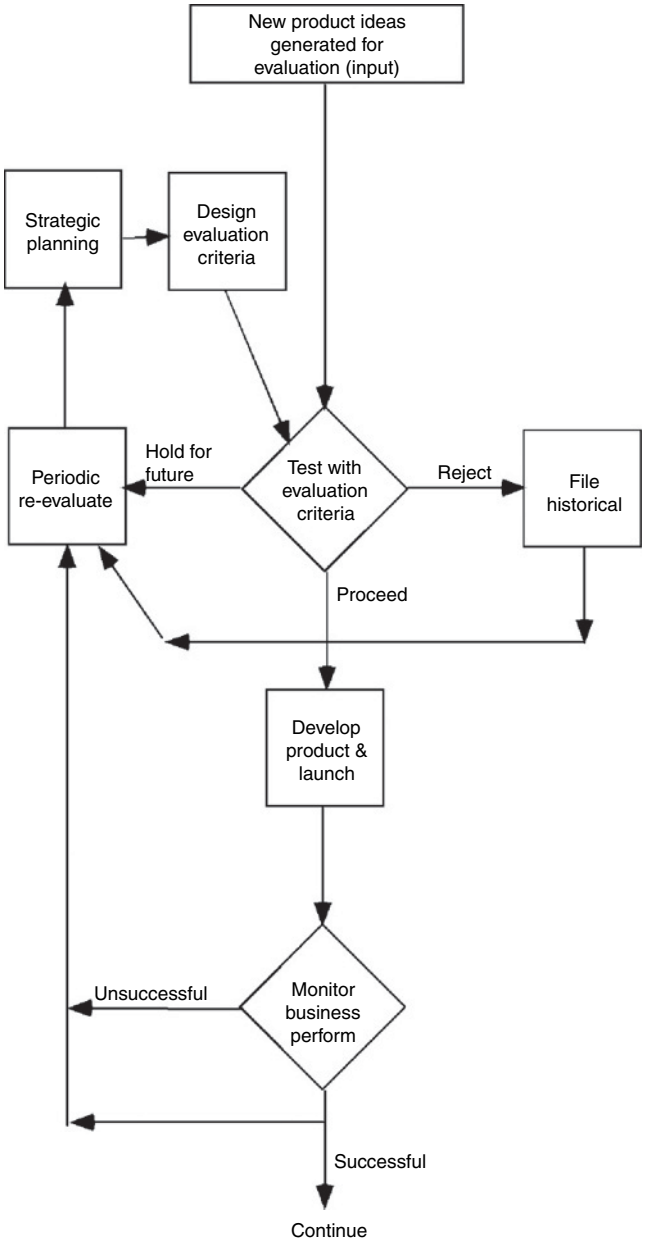


Fig. 3.2.8 Evaluation criteria maintenance.

development plan? These questions need to be periodically reviewed. Figure 3.2.8 is an illustration of how the process should work.

3.2.3.3.6 Objective criteria versus agendas and areas of comfort

The final and summary thought in this section is that new product opportunities must be evaluated with two methodologies in mind. The first is that they need to be evaluated in a consistent manner. The second is that the consistent manner be tied back to the strategic plan and the goals of the company. With these two prerequisites for an evaluation system, new product ideas can be brought in and evaluated in a systematic manner, free from personal agendas and immune to organizational areas of comfort.

3.2.4 Product integration into the business

3.2.4.1 The product as the vehicle to profitability

3.2.4.1.1 How does it use and generate cash, profit, and value?

All products use and generate cash. This becomes an important factor as product development programs become larger and have more impact on the overall finances of a company. Depending on the type of business and the labor and material content, every dollar of added sales requires additional dollars of operating cash to fund it. This cash winds up as inventory either in raw goods or completed assemblies, as well as accounts receivable. The development requires cash for the core intellectual property development and also the capital equipment needed to eventually manufacture the products. Accordingly, the organization experiences cash pressure on several fronts: expenses, capital, inventory, and accounts receivable. This is shown in Figure 3.2.9, where a negative cash flow exists for the duration of the development program. Depending on the cost of capital employed, the inflation rate, and the time to develop, this negative cash flow can impact the organization significantly if it is not recovered rapidly through profitable revenue.

Some products require enormous amounts of real-time cash that could have been retained earnings. The impact is manifested in the value of the stock and shareholder equity. Consequently, timing is of the essence in recovering the initial investment.

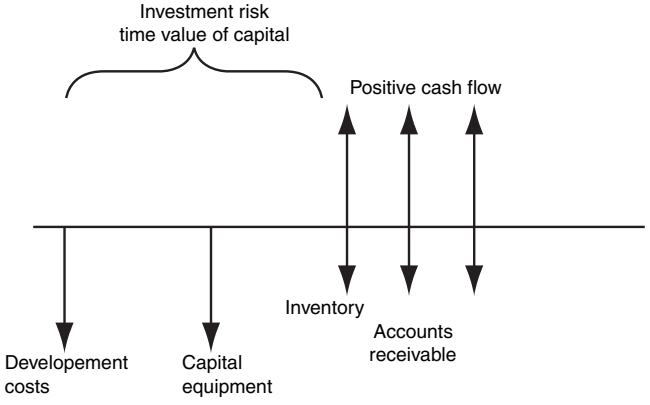


Fig. 3.2.9 Cash flow.

3.2.4.1.2 The goal is dollar and value contribution to margin dollars

To use a British colloquialism, at the end of the day the objective is to have the development program financially payoff in the form of margin contribution and value contribution to the organization. It may be a long road to get there, but if the opportunity is real and the plan is achievable the organization must have the faith to see it through. The following example illustrates this point.

In 1948, the Halloid Company paid $25,000 for the right to develop xerography. Their original target for introducing a Xerox machine was in 1950, some two years later. The model, a copier, had 1260 separate components. Fraught with problems, the first real, marketable unit wouldn't be introduced until 1960. The financial drain on the company to develop and market the product was tremendous. The total development cost was $75 million, which amounted to more than twice the operating earnings of the company during that time frame! However, the first unit sold for $29,500 at introduction in 1960, and the financial impact of the revenue was staggering. By 1970, Xerox's stock had grown in value 66-fold, a fine example of big risk, big payoff.

We must mitigate the financial exposure in the process by holding firm on development costs and pursuing the reduction in development time. The preservation of the market window is one of the most important aspects of the enterprise.

3.2.4.1.3 Integrate the business development team into the business

One of the most positive things a manager can do to improve the efficiency of business development team is to integrate them into the business. Most members of a development team know something about business. They are generally familiar with the cost structures, the material content, and the overhead and burden rates, and have a general product-costing orientation. Very few, however, have an appreciation for the more strategic and tactical aspects of the business. In addition, very few have an appreciation for the sense of urgency required in the product development playing field.

Consequently, training the team in these aspects will have payoffs. The manager needs to outline the impact of slipped schedules and creeping factory costs, and what that means to the revenue line, the recovery of investment, and the loss of competitive advantage. Selected elements of accounting and finance should be reviewed with the team as a group and reinforced on an individual basis where appropriate. Because the team needs to be interdisciplinary, each contributor needs to have broad-based training.

3.2.4.2 Playing the business development game

3.2.4.2.1 Identify opportunities in which small, incremental changes mean big advantages

If you loosely define the competition of new product development as a game, then consider it to be in the organization's best interest to learn to play the game well. The need to become proficient at incremental changes has been previously discussed. An adjunctive activity is to become proficient at identifying those items that are small and can be accomplished with a minimum of effort, and execute them for incremental gains. A series of these will mean big payoffs in the near and long term. They serve as practice for larger programs and also keep competition at bay by constantly raising the barrier to entry in the specific marketplace.

3.2.4.2.2 Determine latent opportunities for competitive advantage, and incorporate them

Here is where the engagement of the customer becomes so important. At this interface, if you listen carefully, the latent needs of the customer will become evident. In that latent need the identification of the product features, configuration, or both will be defined and the value posted by the customer. By incorporating these features, the company can gain a competitive advantage with small, incremental effort and protect margin against erosion as a result of competitive pricing actions.

3.2.4.3 Winning the business development game

The marketplace today is characterized by an unprecedented number of versions of similar products. Everyone wants designs custom-tailored to individual tastes, and the products have unprecedented requirements for documentation and support. Just consider the entire infrastructure that has been enabled to support computers, software, and operating systems. Many products generate an enormous potential for errors in an environment where there is unprecedented potential for punitive damages accompanied by an unprecedented lack of consumer responsibility in using the products. This makes creating a product even more difficult, above and beyond the normal commercial pressures and competitive issues.

3.2.4.3.1 The concept of platform management

Many companies are adopting a product development system that employs platform development as a basis. The product platform serves as a basic building platform on which several versions of a product could be built. In

this way, several versions of a product can be quickly implemented on a platform to serve the market rapidly. At a later date, a cost-reduced version could be implemented after volume begins to climb.

An example of a product platform could be a computer tower. The basic system consists of an enclosure, a power supply, a motherboard, and slots for peripheral components such as hard drives, compact disk (CD) and DVD drives, graphics cards, and user interfaces such as monitors and keyboards. Some of these components are used across many product configurations, allowing the manufacturer to configure a wide range of products to meet specific market needs.

Thus, as market needs change, the manufacturer can reconfigure and reprice an offering very quickly. Conversely, imagine the effort that would be needed to create a unique version for each and every customer from scratch, with each component optimized for the use configured!

The problem with platform development, however, is that a disproportionate amount of development must go into the platform as opposed to a specific version. This is because the designers must factor in compatibility with components, obsolescence, re-sourcing of component parts, and so forth.

3.2.4.3.2 No company or market exists in isolation

To give some perspective on the impact of the faster-changing world, consider that information is readily available almost anywhere and to almost anyone. This makes more knowledge available to more people everyday.

It would be quite straightforward to institute plans for your company if they could be made in isolation. They could be well-considered and every contingency planned out. However, no company really operates in a vacuum, either on the purchasing or the selling side. A network of suppliers, vendors, customers, and employees interconnects each side. Since there is this interconnection, information travels fast and fairly pervasively, albeit not always accurately. Consequently, any advantage is short-lived and information must be acted on rapidly and decisively to capitalize on it.

3.2.4.3.3 The race for the prize

Firms are interconnected in a race to secure incremental business. What one supplier loses in business, another secures. With a fixed market size, it is a zero sum game. In our world today, many of the markets are fixed in size or are growing at a very slow rate. To create real, large growth in a market, one must enable and cultivate a new set of customers. This is done by ushering them into the world of commerce as participants. By developing new customers, markets can grow, thereby relieving the constant competitive pressure.

Therefore, an individual firm must move rapidly to arrive at the opportunity before its peers do. If currently there is no competition in an uncontested area, there will soon be. Undiscovered or newly discovered niches and segments will attract competitive activity. Lucrative niches developed by one firm will be plundered by another. It is the manager's challenge to carve out the company's place in the market and hold onto it amidst all the activity.

3.2.4.3.4 Six degrees of separation, and networking

In 1967 a social scientist, Stanley Milgram, at Harvard University conducted experiments that resulted in the "six degrees of separation" theory. The hypothesis was first proposed in 1929 by the Hungarian writer Frigyes Karinthy in a short story called Chains. This theory addresses the connectivity that exists in our world, suggesting that we are all interconnected by a vast network of people that essentially can link us with any other person, within six or fewer links. Figure 3.2.10 illustrates this point.

This is important in that it describes a sphere of influence that surrounds us all. This being the case, we can gain access to people and information, through each other and daisy-chain to reach otherwise unreachable people. The same can be said of organizations. They each have a sphere of influence that extends to other companies, which can allow access to otherwise unreachable information, contacts, customers, corporations, and suppliers. This is illustrated in Figure 3.2.11.

Your company initiatives are communicated through your employee's spheres of influence to suppliers, customers, and competitors. Thus, care must be taken in initiating plans and product developments to prevent

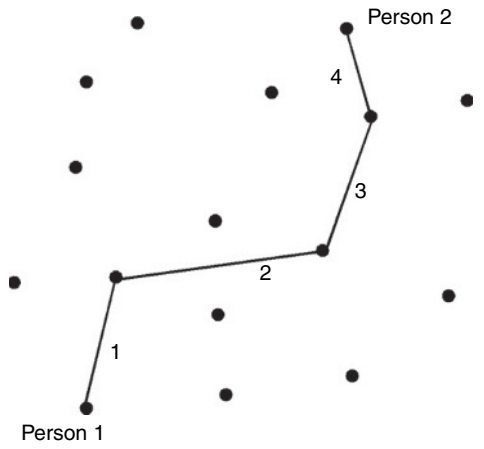


Fig. 3.2.10 The six degrees of separation.

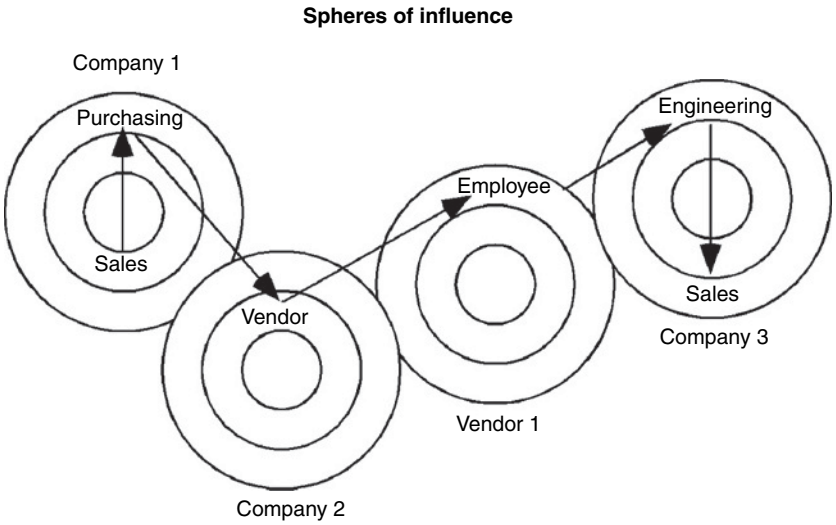


Fig. 3.2.11 Corporate spheres of influence.

inadvertent communication of sensitive material. In other words, if you can gain access to competitive data via your network, outsiders can gain access to your data. It is indeed a faster, changing world.

3.2.4.3.5 More speed, accuracy, and agility before your competitor

How does your company win the competitive game? Everyone seems to be pursuing the same business that you are. How can you differentiate your company from all the others? The perspectives in answering these questions lie in your company's ability and desire to demonstrate speed and agility to the changing market conditions.

It was previously stated that no company or market is an island. It also follows that conditions of advantage and marketability do not remain fixed. Responding to these changing conditions in a timely manner will ensure success.

It might seem from the tone in this and previous sections that winning in the new product arena is a formidable, almost impossible task. It could be, for firms that do not continually strive to improve their products, their marketability, and their position in the marketplace. If, however, your company does execute these continual steps in new product and market development, you will be well-positioned to compete. If you have a strong foothold in the customer base, it will be that much easier to defend it and grow.

3.2.4.3.6 Consultative selling

"Our customers come to us because they view us as the leaders and trust our expertise to resolve their needs." This statement exemplifies the result of continued consultative selling. The company has positioned itself as the resource that the customer will use in the normal course of their business. It also creates a dependence on the company that results in a competitive advantage. In addition to the normal advantage in a bidding or competitive situation, it also positions the company to be on the customer's cutting edge of new developments, placing them there first. This type of selling allows the customer to become a partner in their future success. Such a relationship requires mutual trust and involvement.

3.2.4.3.7 Facilitating the customer's dependence

How can you create this dependence between your company and the customer? Your solution needs to provide features and benefit services that the customer would ordinarily need to provide. By providing your solution to them, you can offload this service burden from the customer. If they redeploy or reduce headcount, then you have created the dependence. You need to be selective in this area because if you are not, you run the risk of attempting to be all things to all people. Create the dependence where it does not cost your company too much in time and effort or, better yet, do this in an area where you are exceptionally proficient.

This type of selling, although very successful in the long run, does require a significant amount of energy and funds to develop and cultivate. Your company needs to make a commitment to select a customer and see it through to completion. By effective partnering with the customer's future plans, links between the two companies can be cemented, which will help transcend competitive thrusts.

3.2.4.4 The new product is the competitive weapon

3.2.4.4.1 New product development is the survival means in a market

Most companies cannot compete on price and delivery alone. In fact, that is quite a dangerous pathway. New product development is often used as a competitive weapon in the marketplace. With most companies, customers and competitors know their history, patterns, and areas of comfort. New product development is a means for a company to gain advantage, secure a position, or win a new customer. It is the ultimate weapon in changing the playing field and should be used as such.

3.2.4.4.2 Leapfrogging the account and the industry

Examine the industry and the marketplace. Determine the products and the technology to be employed to solve customer problems. Cite advantages and disadvantages of each of these implementations, then forecast the future of each of these technologies. This is where the competition will be at your next product introduction. Will you exceed customer requirements, or be at parity with your competitor? Your program needs to leap ahead of where the others will be at that future point. Aim farther as a starting point so that at the time of introduction you will be in a superior position.

3.2.4.4.3 Attacking your entrenched position

So things are going fairly well. No major problems, volume is reasonable, quality is acceptable, and costs permit a good margin. Time to relax and reap the benefit of all the work, right? Wrong. Now is the time to begin to "attack yourself." Do the proper amount of introspection to determine how long you will have the advantage, what the vulnerable points of the business are, what steps need to be taken to secure future business levels and more. The competition will be doing this, so it is in the company's best interest to do the same, and to do it before anyone else. This activity is healthy because it reduces complacency and rekindles a sense of urgency in continually improving the processes, the products, and the relationships with customers. Now is a good time to revisit that loyal customer and solicit their feedback.

3.2.4.4.4 The latent opportunity

Are there latent opportunities within your customer base? What are the specific requirements they need to compete better? Find that specific requirement and seize the opportunity to capitalize on it. These are the actions that bind you to a loyal customer. The difficulty is in finding out what the latency is, because it is not always readily obvious. Often, the customer needs to be led through the discussion to obtain this information. This skill is not readily available through most sales channels and requires people who can communicate effectively and solicit information in an unbiased fashion. However, since it is somewhat difficult to obtain for your company, it will also be difficult for your competitor.

3.2.4.4.5 Raising the barrier to entry

In any situation, it benefits the company to keep the barriers to market entry as high as possible. Surround and imbue the company with expertise and equipment that gives specific market advantage. Investing in capital equipment financed from profits raises the barrier, thus discouraging ill-equipped start-ups from encroaching on your territory. Investment in the business will not only raise the barrier but will also lower costs and improve repeatability and quality. These have a disproportionate effect in terms of competitive advantage and position. Someone chasing your business accounts will find them all that much more difficult to capture.

3.2.4.4.6 Raise customer expectations beyond the competition's ability to deliver

Your company's actions should continually raise customer expectations of provided products and services. By setting the standard for performance in your marketplace, you make the competition dance to the standards you set, not vice versa. This continual resetting of the customer's expectation level will also raise the barrier to entry from a features and business services perspective—another difficult target for others to reach. This can also catch others off-guard because they won't understand the schedules that the partnership establishes for implementation. Your schedules will chart the pathway and will be known only to you and the customer/partner, making it a guessing game for anyone else to intercept.

3.2.4.5 The business and product evolution flow chart

3.2.4.5.1 Identifying the opportunities that contribute to the vision

Business growth through product evolution is one of the most valuable processes a forward-thinking organization can undertake. It is the essence of the road map to success. It defines where the company plans to go, in what time frame, and how it will get there. The content of a well-considered flow chart can be worth more than the strategic plan itself, since this defines the product path to be taken to the future. The strategic plan then defines the financials and the standing of the firm against its

competitors. The first step, therefore, is to identify the opportunities that contribute to the vision of the future and focus on only them. If the vision of the future is agreed to and locked onto, then reject those opportunities that cannot contribute to it, regardless of how lucrative they may first appear.

3.2.4.5.2 Mix of product and service offerings

The flow chart should not be limited to hard products. It should also include the service aspect of the business. In addition, the mix of products and services should also be projected because this will define infrastructure requirements, finances, and personnel.

3.2.4.5.3 Contribution to revenue and profit

Where will the revenue and profit come from in the next several years? What percentage will come from new products and new services, and what will the margin structure look like? These questions, although not having to be answered, need to be reflected on to prepare a well-considered flow chart. Figure 3.2.12 illustrates the concept.

The chart in Figure 3.2.12 represents an attempt to factor profit contribution centers into the product evolution flow chart. The chart is split into two areas, namely, products and services. The products can be existing and new products to be added in the future. Services, likewise, can be existing and new. The material, labor, and burden components, together with the price and quantity, drive the contribution to margin analysis. By extending this chart, trends can be drawn by loading in historical data, and future mix analysis can occur by forecasting and planning new products and services. This will assist in assembling the flow chart as well as adding financial credibility to it.

3.2.4.5.4 Product mix analysis, new versus existing, and technology patterns

The real value in this process is getting a feel for the trends and the degree of sensitivity each item has, and how it affects the company as a whole. In a historical analysis, look for patterns or performance behaviors in products and businesses. In a future-planning exercise, decide which products and businesses will be the big contributors and which will create large variations in margin and profit numbers. This is a more objective way of evaluating risk factors associated with new technologies and the products that are generated from them.

3.2.4.5.5 Mechanics of the product evolution flow chart

It is time to begin the product evolution flow chart development process. As discussed earlier, this chart will serve as a road map for the products that will be developed, acquired, or brand-labeled. The product evolution flow chart, if prepared properly, can be a helpful tool in charting the future course of the company. By its very specific nature, it forces identification of the products and services that will be required to support the narrative of the strategic plan. It also forces product planning to the point of determining contribution to the overall financials of the strategic plan. Figure 3.2.13 is a typical illustration of a flow chart. It shows several things: the pathway that products will take in evolving; how they combine with others in establishing product platforms; and differentiation of products serving different market segments. It also establishes a time frame for when all of these things will happen.

Product and services mix										
Products	Material (dollars)	Laboar (dollars)	Burden (dollars)	Cost (dollars)	Price (dollars)	Margin	Qty	Contribution to profit dollars	% Contribution	% of total
1	1	23	4	28	55	0.49	10	270	8.9	5.4
2	2	24	4	30	67	0.55	10	370	12.2	7.4
3	2	23	5	30	68	0.56	10	380	12.5	7.6
4	2	25	6	33	89	0.63	10	560	18.5	11.3
5	2	56	7	65	125	0.48	10	600	19.8	12.1
6	2	45	8	55	100	0.45	10	450	14.9	9.1
7	2	6	8	16	34	0.53	10	180	5.9	3.6
8	2	7	9	18	40	0.55	10	220	7.3	4.4
Services						Subtotal		3030	100	61.0
1	0.5	39		39.5	75	0.47	10	355	18.3	7.1
2	0.5	32		32.5	75	0.57	10	425	21.9	8.6
3	0.5	24		24.5	75	0.67	10	505	26.0	10.2
4	0.5	76		76.5	75	–0.02	10	–15	–0.8	–0.3
5	0.5	87		87.5	75	–0.17	10	–125	–6.4	–2.5
6	0.5	98		98.5	75	–0.31	10	–235	–12.1	–4.7
7	0.5	12		12.5	75	0.83	10	625	32.2	12.6
8	0.5	34		34.5	75	0.54	10	405	20.9	8.1
						Subtotal		1940	100.0	39.0
						Total		4970		

Fig. 3.2.12 Product services mix analysis.

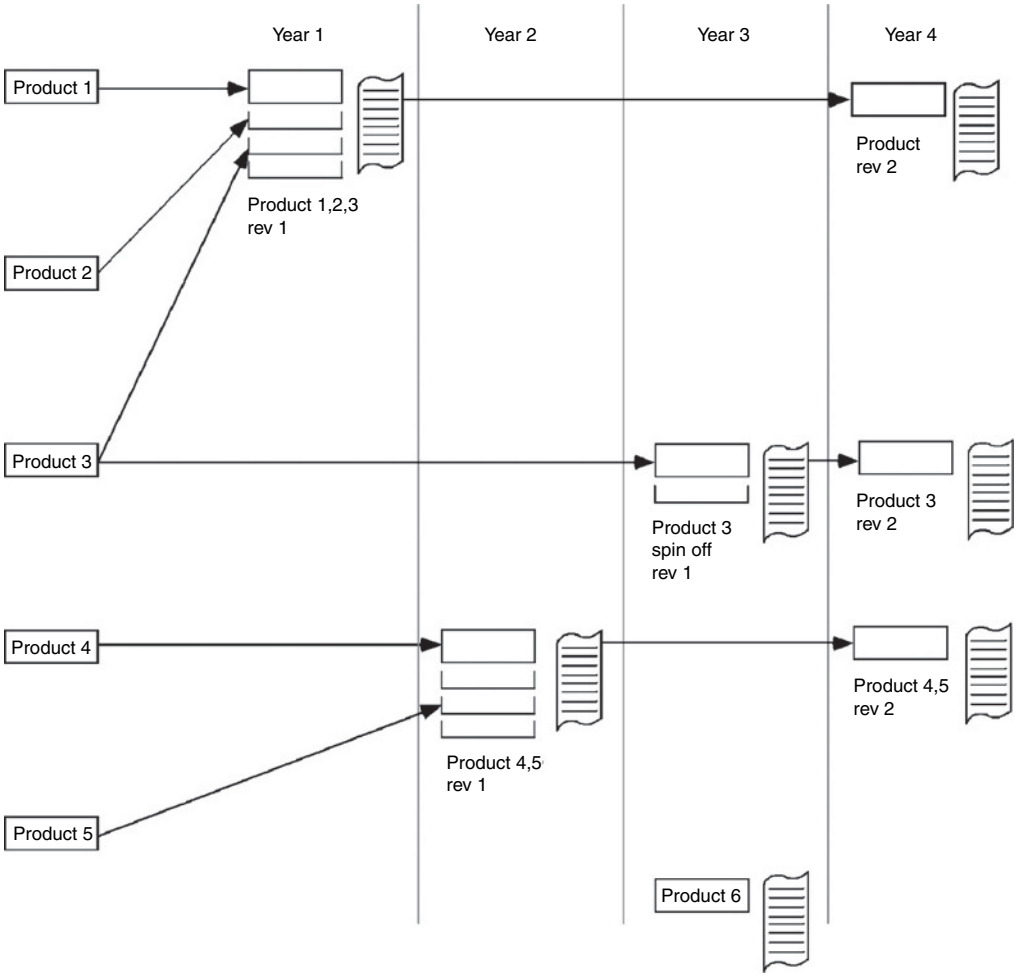


Fig. 3.2.13 Product evolution flow chart.

As shown in Figure 3.2.13, Products 1 through 5 exist in basic form in year 1. The product evolution flow chart shows how these products can be combined from a features and functionality perspective. For example, Products 1, 2, and 3 combine as a result of a development program started in year 1 and introduced in year 2. All of the features in all three products are combined into one product configuration or platform. The basic platform can be configured in one of three versions. This product implementation then runs as a product for years 2 and 3. In year 4 the next-generation product, which combines all the functionality and features of the different configurations, is launched. Newer technology and cost reductions make this condensing of products cost-effective.

In a similar manner, product 3 will spin off a series of product configurations in year 3 from a basic platform. In year 4 a refined product will also be introduced.

To make the flow chart more meaningful, a narrative should accompany each product implementation, which explains the actions taken and for what reason. Also, the objective of the implementation should be included. The narrative should contain the following components: a description of the product, a definition of the markets served, the degree to which the competitive stance is improved, the product configuration, and the expected level of technology employed. In this example of a flow chart, product 6 could be added to the portfolio as a result of an acquisition or a brand label arrangement.

As an additional perspective to this product-planning exercise, it may be desirable to prepare several flow charts showing alternative options and pathways. In this way, the best approach can be selected for implementation.

3.2.4.5.6 Internal requirements, external requirements

Until this point, we have examined the product planning process mostly from an external market perspective. There may be instances in which a requirement for a new product would be a cost reduction or a quality improvement. In such cases, these internally generated

reasons for new product development must require a narrative in the flow chart. From a general management perspective, each product must generate cash and profit. Existing programs that are in trouble, either from a prime cost issue or a cost of quality issue, need to be addressed.

3.2.4.5.7 Future catalog of products and services, features, and technology

What will the future product offering look like? What types of technology will be employed? What level of complexity will it involve? Will training be a marketable commodity or will it be assumed as part of a sales transaction? This process should actually result in a future company catalog with the complete offering included. This 5-year catalog will also drive a thematic narrative of company operations and services in the future. By assembling this catalog, one can see if the strategic story holds together when the products are attached to it.

3.2.4.5.8 Projection of competition, company, framework, products, technology, and positioning

With all this great planning, your company is positioned to capitalize on the future. The competition is not standing still, however, or at least not all of them. Consequently, it is a good exercise to project what they will do based on what you currently know about them. This should be done on the product evolution flow chart and the 5-year future catalog. If for no other reason, it will force consideration of their plans and allow you to evaluate the risk of your plans resulting in future product offering that are only at parity with the competition or, worse yet, nearly obsolete by comparison.

3.2.4.5.9 Who, what, where, why, and how?

A good plan requires specifics of the mobilization to be clearly articulated. What are the specifics of the product evolution flow chart? This product was developed and introduced by such and such date. This technology was absorbed into the organization by another certain date. Additional channels of sale were established, in place, and functional by still another date. These are the specific statements that actions must follow to accomplish results. The flow chart, the narrative, and the 5-year future catalog must answer the five questions of specificity: who, what, where, why, and how.

3.2.4.5.10 Organize stages

Finally, organize the action items and the stages of the plan so that it all ties together into a cohesive set of specific objectives and their supporting action items. Look for commonality of purpose and resources to optimize the work needed.

3.2.4.5.11 Optimize and maximize benefit for dollars spent

Resources for development are always in demand. In addition, the resources for development represent only a small part of the overall investment in a new business opportunity. It is therefore in the company's best financial interest to maximize the results for each dollar spent. Financially sound, well-run small companies consistently deliver the maximum benefit for every dollar spent. They must in order to survive. Larger companies run the risk of failing to do so because of politics, lack of consistent information, and lack of communication flow. In either case, the cost of lost opportunity and the hard cost of research and development demand that the development team maximize the results of dollars spent.

3.2.5 The competitive arena

3.2.5.1 The competitive game

3.2.5.1.1 Upsetting the playing field to your advantage

There are several marketing strategies in use today in terms of leadership roles: guerrilla tactics for niche players, defensive positions, and offensive positions. However, all of these strategies become much more valuable when the new product initiative upsets the current playing field. By finding a way to upset the playing field, you improve your position twofold. The first way is to compete with the existing players with a new product. The second way catches them off-guard as to the approach used. Depending on the uniqueness of the approach, it may be all that more difficult for the competition to respond.

The following example illustrates the concept from two perspectives, long-term and short-term. In the 1970s, U.S. automobile manufacturers lacked certain quality features that the American public was looking for in the product. Several foreign manufacturers embarked on a program to improve the quality and service of their products to the American consumer. By the 1980s, the perception had shifted regarding the foreign manufacturer, from that of shoddy goods to a high-quality, consistent product with exceptional service. Their battle cry: American consumers are entitled to exceptional quality! This changed the playing field for Detroit and, ultimately, severely impacted its market share. This was a long-term change in the playing field that led to a distinct market advantage for foreign manufacturers. Tremendous amounts of time and money were required for Detroit to place the competition at bay.

In more recent history, the computer industry exhibited short-term changes in the playing field to

create market advantage. Different architectures in operating systems, software, and license issues affect access and market position everyday.

3.2.5.1.2 Faster, more-informed solving of the customer's problem

A key competitive advantage is to mobilize the operation to excel at determining business and product opportunities by engaging the customer and becoming proficient at designing and introducing products with customer solutions embodied in them. The organization needs to bring formidable factors (internal and external to the organization) to bear on these opportunities. The better one is at this, the better one's chances for growth.

3.2.5.1.3 Traditional market differentiation has changed

The marketplace is different from that of 10 or 20 years ago. We have experienced well-made products at ever-lowering costs, and better service and more availability of product. Out-of-the-box failure rates are becoming more scarce in different industries. Competitive edge won't come from these areas; they are assumed. The competitive advantage will come from mass customization, applying basic platforms to a diverse set of customers, improved availability, application of specific products, and from low-cost, feature-rich innovative producers.

3.2.5.2 The strategic difference—large versus small companies

3.2.5.2.1 Large companies: leverage buying, dominate markets, intimidate, steamroll, and cover the market

In a homogeneous market condition, a large company can have an advantage. By virtue of its size, its capabilities, and its staying power, a large company can dominate a market. If it is a dominant player in the specific market segment, it can influence the future of the segment and its general progress. An entrenched player has the capability to eliminate competition through pricing actions, availability, supply logistics, and reduced costs.

3.2.5.2.2 Small companies: niche players and guerilla warfare

Small competitors have advantages of their own, but in different ways. A small company can react to changing market conditions quickly, and capitalize on them. The smaller companies will enter a market to exploit an opportunity and be prepared to leave it at a moment's notice. They do not have the staying power or temperament for a long, drawn-out market battle. By virtue of their culture, they will abdicate certain markets under contest and look for new opportunities. "Fast to enter and fast to leave" is often their operational perspective.

3.2.5.2.3 While large companies evaluate and strategize, small companies execute tactics

There are distinct differences in how large and small firms, plan, strategize, evaluate, and then act. Figure 3.2.14 illustrates a review of their respective product life cycles and how large and small companies differ in their approach, execution, and longevity in the specific segment. In addition, the financial profile generated in each of the examples and how they differ are examined. Figure 3.2.14 shows the traditional product life cycle. There are three distinct phases of the product life cycle (i.e., emergence, growth, and decline). The emergence phase is the start of the new product's business. The volume is low and the product is embryonic in its market share. The second phase is the growth phase, where planned volumes are achieved. The third phase is the decline wherein, if left untended, the company will lose market share and volume will drop off. Generally, at this point the company is no longer competitive in that specific market segment.

Ideally, at the time when volume is stabilized and the company is enjoying good profit margin and good volume, the project to replace the product with an updated or next-generation product is initiated. This is timed so that when this development is complete it will coincide with the reduction in volume of the first-generation product. It is through these follow-on programs, product improvements, and cost reductions that the company will retain and grow its market share. At this time the company may want to reexamine the market and segment the product offering to capture different tiers in the market, thus enhancing market share as time passes.

The difference in the size of the company also determines the differences in implementation. As can be seen from the third diagram in Figure 3.2.14, larger companies may be slower at the initial implementation but achieve significant volume once the massive selling and manufacturing machine is mobilized. The small company may be lightning-quick at the first-article implementation; however, it will quickly lose momentum compared to the larger company as a result of lack of critical mass. The consequences are clear when the following financial analysis is evaluated, concurrent with the chart in Figure 3.2.14.

The small company characterization in Figure 3.2.14 reinforces the assumption that as a product opportunity is discovered, a rapid implementation is generated and put into the marketplace for consumption. The firm makes a small investment to test the marketability. Investment in capital equipment is minimal, and the

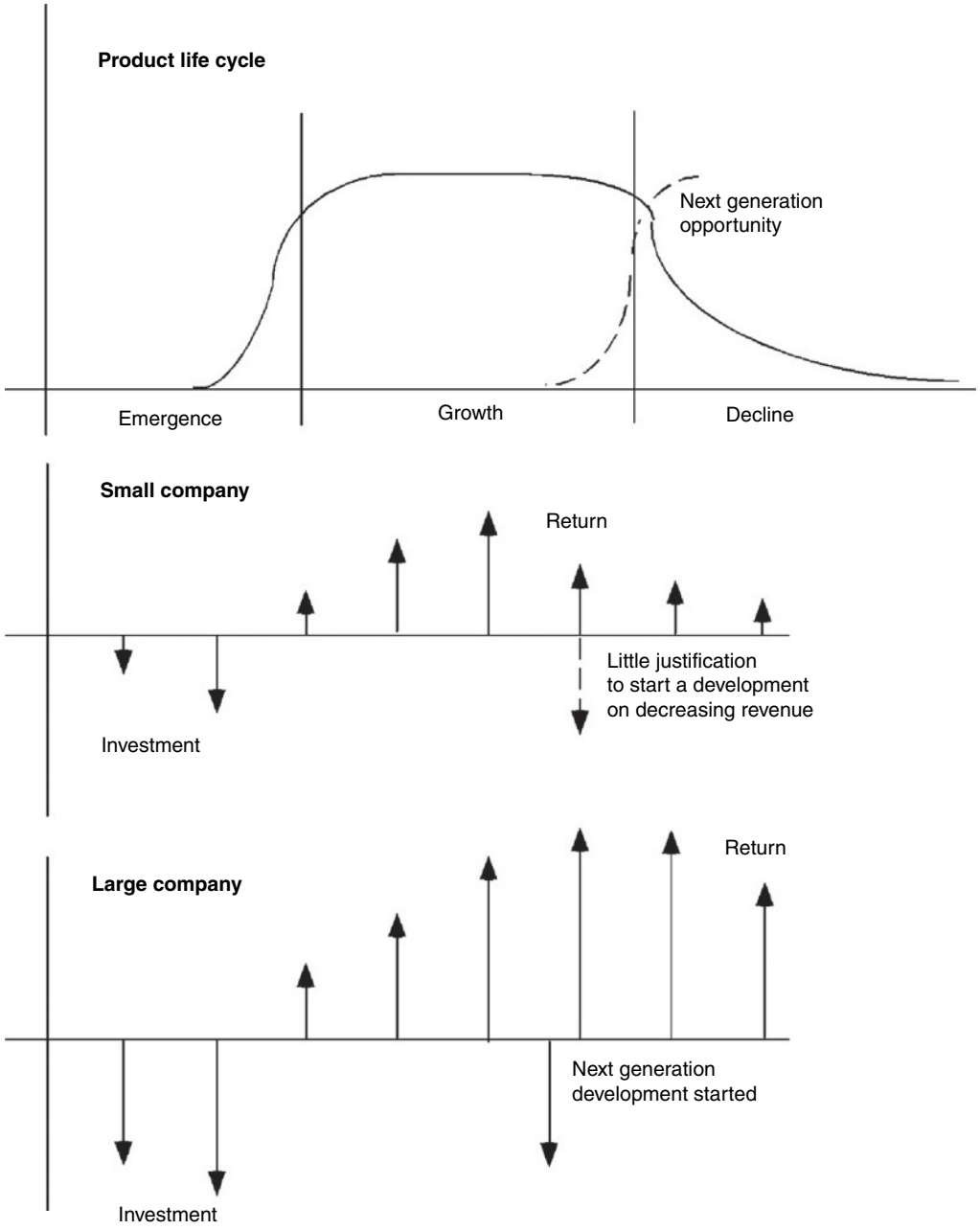


Fig. 3.2.14 Product life cycle: Large versus small company.

product is introduced. Initial volume looks promising as it begins to build up in each successive year.

The problem that now results is twofold. The first is lack of sufficient market share initially established to gain hold of the market. The second is cost pressure due to lack of low-cost product platforms. This results from lack of sufficient capital expenditures to keep the costs low. In addition, there is no adequate barrier to entry in this market from a capital perspective. The development commitment should not be to test the market. It should have been to secure the market.

At the time when the next-generation product should be developed, the outlook for the first-generation product is in decline. Consequently, the company decides that it is not worth pursuing the opportunity, in light of all the other competitors, and seeks another uncontested niche.

In contrast, let us examine a large company's implementation. In this example, the large company researches the opportunity and commits to exploiting it. This involves an interdisciplinary effort within the corporation that focuses on developing the product and the

Differences in product execution: large versus small companies			
	Market condition		
	Emerging	Growth	Decline
Large company strategy	Initiate	Growth and market share	Last man out
Characterization	1. Slow to mobilizet	1. Investment in product	1. Retain with other means
	2. Deliberate and methodical actions	2. Interdisciplinary and massive investment	2. Fight to retain and grow market share
Small company strategy	Develop and secure	Enjoy differential advantage	Abdicate and find new
Characterization	1. Quick to mobilize	1. Immediate limited growth	1. Growing list of competitors
	2. Lack of long-term planning	2. Missed customer signals	2. Failure to redesign to cost

Fig. 3.2.15 Comparison of large company versus small company product execution.

business for the long term. A significant investment is made in development in years 1 and 2. This investment is in developing the product and the manufacturing infrastructure to gain market share. The result is a rapid growth to a favorable market share figure.

As time progresses, the market share and volume increase to a peak. As the volume begins to taper off in subsequent years, the second-generation product has been scheduled for development and is introduced to keep market share stable and growing. The result is a much stronger position from the outset through the product life cycle, thereby generating revenue for profit and subsequent development. This is how a solid commitment in a well-researched program can contribute to long-term growth that can be retained and built upon.

Figure 3.2.15 is a summary of these differences.

3.2.6 The pathways to the product

3.2.6.1 Corporate skills assessment

3.2.6.1.1 Business capabilities—you have the heart; do you have the tools?

The honest assessment of corporate skill level in each area of business development is a very difficult thing for most companies to do accurately. The process calls for deep introspection within the organization. In addition, it is very difficult for biased personnel (everyone has biases) to prepare an unbiased assessment. Since there is no absolutely accurate way of performing this assessment, the best we can do is to invite the firm to address the following perspectives in an open, honest, and complete manner. In addition, it may be a benefit to have the customer assist with this assessment.

1. What do we do best? (Distance yourself from the sales and marketing literature and focus on the customers' and market's perception of who you are.)
2. What do we do worst? (Take an honest look at corporate failures and what the contributory causes were.)
3. Who are we kidding? (Take a look again and now come up with the real contributors.)
4. How do these align with market opportunities? (Take the time to actually reduce this to paper and look at the comparisons and areas where you fall short.)
5. How do we create a sustainable competitive advantage? (Look at each market opportunity to determine how you can create a scenario that allows the company to establish a success in a market and retain the business in that market on a long-term basis.)
6. Can we partner with someone and accomplish our goals? (This is often overlooked in an analysis. Many companies cannot partner with another company and accomplish something. Sometimes the team is not aligned; other times, communications systems are not pervasive or do not have appropriate protocols established.)
7. Do we have the talent to compete? (Are your firm's sales channel and personnel capable of effectively battling for an order or a piece of business and able to prevail over a competitor? This requires a special skill set; if it is missing, one needs to instill the capability in the responsible group.)
8. Do we have the talent to assimilate? (The corporate ego would indicate yes, or that you could soon obtain it; however, in reality, your firm may not be

able to assimilate the technology or integrate the technology into a product within the proper time frame.)

9. Do we have the temperament for business in today's market? (The acquisition of business in today's climate is fast-paced, highly competitive, and difficult to maneuver in. Accordingly, firms must be up to the task. This will not happen with a fixed mental paradigm or a workforce that lacks some diversity.)
10. Do we want the market as badly as our competitor? (Historically, how much has your firm been willing to fight for a piece of business? More important, how much has it been willing to change, adapt, and improve to capture a piece of business? Often, the intent is there in words, but implementation is not there in reality.)
11. Are our people motivated to compete? (Addressing the culture of the company, are the employees motivated to compete and prevail? Do they have a winning mentality and act accordingly?)
12. How do we view our ability to employ technology? (These next perspectives to consider are a little more difficult to understand since they require both internal examination and external examination. The assimilation of technology and developing the capability to be conversant with it to the point of creating a product is a long way from the commitment to do so.)
13. How do we view our ability in application development? (This is an extremely important corporate skill that is often lacking. It is absolutely necessary to develop the right product and to apply the developed product and create a satisfied customer.)
14. How do we view our ability in setting pricing? Are we a cost-plus company trying to operate in a market-price marketplace? Do we know how to design to cost? Do we accept the first cost or do we work the design and procurement until we achieve the desired result?
15. How do we view our ability to absorb uncertainty? (Do we have a firm whose employees are averse to digging into details, questioning information, comparing facts and drawing conclusions, and performing tests to clarify data? If so, the firm's ability to absorb uncertainty is a limiting factor in delving into new areas.)
16. How do we view our ability to derive the optimized solution? (In the same mode as the cost issue, how does the firm operate in achieving the optimized implementation for a product design? Does it go the extra mile in wringing out all the excess cost? Does it achieve the best performance at the lowest cost and still satisfy customer requirements? These and other items define the company's ability to engage in development in a world-competitive manner.)
17. How do we view our ability to profit from an opportunity? (How has the firm historically profited from an opportunity? Is it consistent with the original estimate? Is progress toward profitability measured along the way and steps taken to correct or enhance it along the way?)
18. What areas of the organization need remedial assistance or additional infrastructure to be a contributing factor to success? (There are specific tools required to execute the program, and an introspective look at the strengths and weaknesses is in order. The corporate capabilities need to be integrated and coordinated at near-full value. If one of or a set of these required capabilities is weak or missing, it must be addressed before the program begins. Failure to do this will result in a resource-starved program, dooming it to failure of timing, performance, or momentum. Examine each area of the business to ensure it can support and contribute the necessary resources to effect success of the program. If the organization is lacking, fix it now!)
19. Can you have access to the skill sets? (Is it possible to obtain the tools for the organization from other sources? In the next section, a variety of ways will be discussed to allow companies to compete in nontraditional markets by partnering for capabilities. The objective is akin to obtaining a cut sheet of lumber, not to necessarily own a sawmill. If this is the case, don't go looking to buy a sawmill, unless you use so much cut lumber that it is in your firm's best interest to invest in capital to become a low-cost producer and have control of the operation. This would only take place when usage drives the make-or-buy decision. Obtain the tools that make most sense to the operation and the new product development.)

3.2.6.1.2 Partnering for capabilities

3.2.6.1.2.1 Why collaborate?

Technology progression is moving at an incredible rate. This places the absorption of uncertainty of business development programs and business operations at an ever-increasing expectation to achieve the required assimilation of technology—the cost and time to develop technology can cause companies to expend vast amounts of resources. Remember the forecasts for business levels due to new products many years ago? The maintenance of these products is also costly to your firm in time and personnel, making it increasingly difficult to devote talent to new business development. These factors have

historically driven companies to begin the process of collaborative development. However, there are both commercial and technological issues associated with the collaborative business development process.

3.2.6.1.2.2 Commercial collaboration issues

On the commercial side, business practices and methodologies are being developed rapidly. Also, they are being developed all over the world with different value systems attached to the methods. It is very difficult for a single company engaged in the process of conducting normal operations to be expert or even conversant with these methods.

Having a collaborative partner to assist in tracking the evolution of best business practices and routes to market is a plus. In addition, these collaborative partner's experiences can assist in channel management and how to best navigate market contention issues. It is also quite difficult, given the complexity of our world, for an individual company to select the optimum long-term manufacturing venue and to create manufacturing supremacy.

Will your partner be as willing as you are to make certain sacrifices, given their status? This investigation also extends to ownership structures and motivations that will influence this commitment. Also, what is their degree of sophistication regarding relevant groups and what is their experience with collaboration?

3.2.6.1.2.3 Technological collaboration issues

How will the intellectual property issues be handled in the cooperation? Who will pay and who derives the benefits? Where does the technology reside? Who has the core technology? Who has the right to spin off products? Many times these issues are glossed over in boilerplate contract language and problems can arise later.

Once the products are launched and the business is developed, who governs and provides ongoing product support? If the product requires changes, which partner will provide these changes? If both companies are marketing the product, how do customer-driven changes get incorporated? The ability to adapt to changing market conditions is critical to success. When there are two cooks to the broth, trouble can occur, especially when there has been no previous agreement.

Finally, in the absence of formal intellectual/property protection, trade secret protection can be an excellent way to protect products and processes. The partner company must therefore enforce disciplines to keep trade secrets, secret. Accordingly, this must be a factor in the collaboration negotiation.

Finally, there is the issue of match between the two collaborative partners: are they equally experienced in the process or does one have far more experience developing and launching products in this arena? Take time to define the complementary technical prowess and experience with collaboration.

3.2.6.1.3 Don't be proud; be practical and careful

Corporations are not inanimate entities; they are composed of people. People have needs, emotions, and biases. People feel threatened and react to those feelings. Many a logical, forward-thinking, and beneficial idea has been sidestepped rather than embraced because of these factors. Consequently, the notion of securing assistance or know-how from outside of the organization often does not receive wholehearted support. It is the manager's responsibility to ensure that the organization rises above these factors to embrace the tools required for success, wherever they may come from.

3.2.6.1.4 Getting the tools and remembering the objective

There is an unwritten corporate covenant in business development that exists between senior management and the new product development champion. It is a covenant whereby management will accept the developer's leadership in exchange for a certain amount of due diligence in implementing their program. Consequently, when obtaining the tools needed for a program it is necessary to exercise care in their procurement. Get only the tools needed. Get the tools as a means to the new product end. It is a vocational hazard to quickly get in the habit of procurement on a grand strategic scale and bypassing the critical mass items required for just the program. The operative point is that these tools must be absorbed and amortized by the program, so restraint is best practiced at this time. In the course of procurement it is wise to procure in a forward-thinking, forward-compatible manner. Where it is a small, incremental cost to prevent obsolescence of equipment, expenditures should be made freely. When getting the tools, remember the objective. In the same way you engage the customer to assess where they are, it is a requirement to engage the partner in a more detailed way. Find out their plans, aspirations, and desires. Only in this way can you determine the best deal to strike for both parties.

3.2.6.1.5 Contingency plans, evaluating with a partner, and future in-house capability

"The best laid plans of mice and men often go awry." This saying implicitly defines the scope of new product development, but especially in the framework of partnering. Knowing this is an accepted fact, make sure to plan contingencies within the partnering arrangement.

An unfortunate fact of life is that unforeseen driving forces can alter a partnering relationship for reasons

unrelated to the development, its progress, or its potential. A financial hardship may present itself to the partner of your company and force abandonment of the program and their contribution.

There are also instances in which management focus may change within the partner's management structure. A different agenda may be embraced, or a separate alliance that may potentially overshadow or conflict with the partnership may have been reached. Therefore, it is best to identify alternative and contingency plans at each stage of the process of partnering. As the deal is negotiated, constantly ask yourself, "What if this doesn't work out? What will be my options to achieve the new product development objective?"

One possible alternative may be to plan on absorbing the partner's technology into your organization after the term of the partnership. In many cases, a relationship where this is identified at the outset may eventually be the best for both companies. The initial cooperation is therefore very telling in how the relationships will progress.

Carefully watch the partner's behavior in several situations. Assess their loyalty to the program and to the team of personnel involved. Determine if they act in a long-term, decisive manner or if they operate in a day-to-day, tactical mode with loose focus. It is imperative to assess this quickly in certain programs replete with uncertainty because you will need to evaluate the difference between time lost in overcoming problems or chasing other opportunities before completing the tasks at hand.

Take care to set up a communications infrastructure that preserves secrecy where necessary but encourages a free exchange of knowledge where it enhances the effort, relationship, or the program. Also set up communications at different levels between the two firms, as shown in Figure 3.2.16.

As the figure illustrates, your firm and the partnering firm should set up a communication protocol that is multilevel. Patterned after a hierarchy, bilateral communications can take place at the operative level. For most issues, if nonconcurrence occurs at the lower level it can be arbitrated at the next level. Senior management preserves and orchestrates the relationship between the two firms. Avoid single-point communications because it is only as strong as the individuals at that point. Multiple communications can also allow information to be brought to light in a nonbiased way, whereas the single-point entry can be prone to one individual's biases.

3.2.6.1.6 Collaborative communications

The following is a list of the possible levels of communications in a collaborative partnership arrangement that may be typical:

1. Executive level communications (product planning and product evolution, product management)
2. Development management
3. Manufacturing
4. Aftermarket

There are also protocols for different stages of a program:

1. Beginning of program
2. Development
3. Problem resolution
4. Liability exposure
5. Loss mitigation
6. Reconstitution after a setback
7. Hotly contested competitive areas

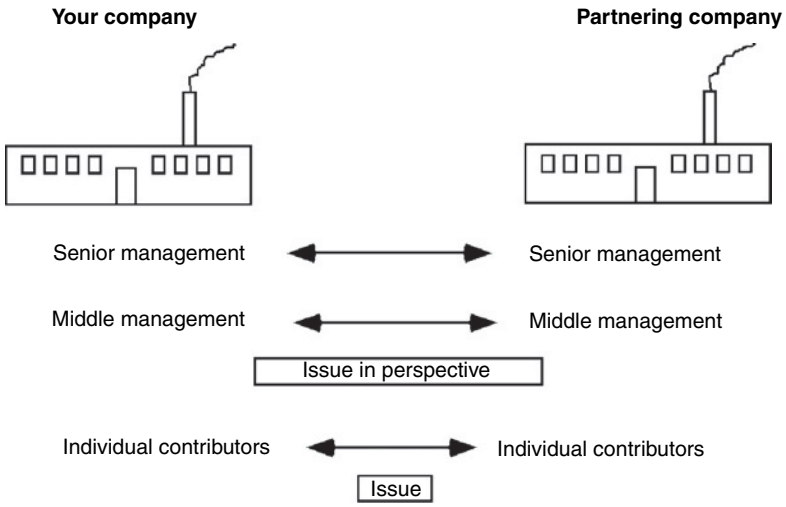


Fig. 3.2.16 Communication model.

3.2.6.1.7 Focus on the mutual opportunity

The basic concept behind partnering is that, together, two entities can achieve more than either one could alone. The partnership fills in the gaps in the individual players. To reiterate, remember this objective: 100% of nothing is not nearly as beneficial to your company as 50% of something.

The very structure and operational dynamics of a relationship between two companies contributes to diffusion of focus. It will soon be apparent that many interesting things surface from the collaboration, not necessarily related to the program. It is the manager's responsibility to refocus the companies on task. In addition, the partnering relationship may not work out for every company. There has to be a certain corporate appreciation for the differences between the players, their history, and (very important) the relative importance of the project to each firm. As you structure an arrangement, ensure that the players selected have the patience for the partnership.

3.2.6.1.8 Concepts of corporate strategy

There are differences in various companies' corporate strategies. Although the following model could have been presented in the planning section in Chapter 3.1, it is more relevant here where it can be recognized in a potential partner during negotiations. Presenting a "hard and fast format" for corporate strategy in the planning section would tend to preempt creative thinking by encouraging comparison and locking onto a specific type prematurely. When contemplating or planning a joint venture, brand label arrangement, or an acquisition, it is critical to determine the seller's corporate strategy. Although most companies espouse some sort of strategy, some do not practice it effectively. In some cases, this depends on the size of the company as well as the ownership structure. For example, a small, privately owned company will execute a strategy differently than a large firm. This is because of three factors:

1. Influence of top management in the process
2. Sensitivity to changing market conditions
3. Inertia of the organization.

A large company sets its pathway and starts execution. It directs its inertia to start in the chosen direction. Changing market conditions generally do not affect this. This is the appeal of the smaller firm, which can react to changing market conditions. This resiliency, however, is a danger to executing a plan because management may change direction more often, thereby negating any momentum. It is incumbent on the acquirer to understand these dynamics.

It would be tempting to create a chart of the various scenarios of acquisitions, joint ventures, and brand label arrangements to summarize the corporate strategies of each, and how to react to them; however, each case is its own set of circumstances that directly affect the dynamics of a deal. Consequently, each must be evaluated and a pathway must be charted through the circumstances.

A large company with a lot of long-term debt may sell off a division at some price not necessarily reflective of its value, and recognize the sale on its financial statement. A smaller firm would execute the transaction totally differently. It is important to investigate the driving forces behind the deal as you are creating the deal. This can have the most payoff when negotiating.

Some of the factors that come into play are as follows:

1. Large versus small
2. Public versus private
3. Technology developer versus pure manufacturer
4. Product offering, which is commodity versus service
5. Profitable versus not profitable
6. Management under fire versus not under fire
7. Availability of cash
8. Competitive nature of the deal
9. Perception in the marketplace
10. Perception of weakness
11. Limiting factors on one firm in the deal
12. Recoverability index for both companies should a deal go bad

These frameworks also factor into the success of a launch of a new product.

This list outlines the various concepts and practices of corporate strategy. It is desirable to study the philosophy of each of the types and to become familiar with them as you negotiate the partnership arrangement. In this way it will be easier to evaluate the partner's motives, pressure points, and actions.

Figure 3.2.17 is a summary of corporate strategies and their characteristics.

3.2.6.2 Internal product development

Once a firm assesses its capability (overall operational ability) to pursue the business and product opportunity, it must now determine the best way to secure the opportunity and make it a sustainable enterprise. This is often a difficult decision for both senior management as well as the business development practitioners. If the firm has the internal development capability and also is conversant with the next-generation platform of the technology, it can start immediately on the program. This is a rare occurrence for a brand-new product. More often, there are certain elements missing from the organization that must be integrated. Although the next sections present a hard, defining line between the internal development and partnering development, the more

Corporate strategies and characteristics			
	Strategic prerequisites	**Organizational prerequisites**	**Common pitfalls**
Portfolio management	Superior insight into identifying and acquiring undervalued companies	Autonomous business units	Pursuing portfolio management in countries with efficient capital marketing and a development pool of professional talenty
	Willingness to sell off losers or sell off winners opportunistically	Small, low-cost corporate staff	Ignoring the fact that the industry structure is not attractive
	Private company or underdeveloped capital markets	Incentives based on business unit's success	
Opportunistic fix-it strategy	Superior insight into restructuring opportunities	Autonomous business units	Mistaking rapid growth or a "hot" industry as sufficient evidence of a restructuring opportunity
	Willingness and capability to intervene and transform acquired business units	A corporate organization with the talent and resources to oversee turnarounds and strategic repositioning of the business units	Lacking the resolve or resources to take on troubled situations to intervene in management
	Broad similarities among the units in a portfolio	Incentives based largely on acquired units' results	Ignoring the fact that the industry structure is not attractive
	Willing to sell off when restructuring is complete and market is favorable		Paying lip-service to restructuring but actually practicing portfolio management
Transference strategy	Proprietary skills in activities important to competitive advantage in target industries	Largely autonomous but collaborative business units	Mistaking similarity or comfort with new business opportunities as sufficient justification for diversification
	Ability to accomplish the transfer of skills among the business units on an ongoing basis	High-level corporate staff whose roles are those of integrators	No practical provision for skills transfer
	Acquisitions of beachhead positions in new industries as a base	Cross-business-unit committees, task forces, and other forums to serve as focal points for capturing and transferring skills	Ignoring the fact that the industry structure is not attractive
Integrative strategy	Activities in existing units can be shared with new business units to gain competitive advantage	Objective of line managers is to include skill transfer	Sharing for its own sake rather than for competitive advantage
	Benefits of sharing outweigh costs	Incentives based in part on corporate results	Assuming sharing will occur naturally without management playing an active role.
	Both start-ups and acquisitions as entry vehicles	Strategic business units that encouraged to share expertise and activitiest	Ignoring the fact that the industry structure is not attractive
	Ability to overcome organizational resistance to business unit collaboration	An active planning role at group, sector, and corporate levels	
	Fostering collaboration, not competition	High-level corporate staff members who see their roles primarily as integrators	
		Incentives based heavily on group and corporate results	

Fig. 3.2.17 Summary of strategies.

common approach may be somewhere between the two ways of approaching the program. They will be presented separately here, however.

3.2.6.2.1 Internal skills development

Most companies do not have all of the skills required to secure a market with a new product. Therefore, as part of the development program there must be sufficient planning to upgrade corporate capabilities. This means time and funds must be appropriated accordingly. This is a difficult thing to do because most teams are anxious to get started and often understate the time and expense of integrating and assimilating new technology. One need look no further than the time it takes to

integrate a new piece of software advertised to cut development time!!! It takes development time to learn!!!

Many technologies require capital equipment support and that requires funding and time. It is best to accurately assess and state the requirements and load them into the program schedule. In this way you will be on the right technology platform and be in a technical position to actually execute the development on time.

3.2.6.2.2 Management development

The lack of experience may not lie with the core development capabilities alone. It may lie with management. Consider a group whose best efforts are put forward in developing software for a product. Each week some progress is made; however, the end is not in sight and lingering bugs keep cropping up. A change in the code to fix one problem introduces another problem. In this case, the development group is too close to the problem resolution and is feeling the pressure to perform. They may have lost objectivity. Even though they are working flat-out to resolve and complete the software, an experienced manager may have to intervene and guide the group to a different way of completing the program. Painful as this may be, the manager must have the experience to recognize the technical impasse and navigate a way around it. Not all mangers have this capability or experience. Failure to have the experience will result in missed deadlines and potential program failure.

3.2.6.2.3 Funding

As we will see later, there are two important aspects to funding. Obvious as they may seem, they are often minimized and can cause trouble later. The first is the actual amount of funding. This includes the time and money to implement the program. Complete funding is often not budgeted or thought through at the beginning. Management may take a conservative approach by metering out a little funding and watching what happens. This lack of complete appropriation places the program at risk for follow-on funding because another opportunity make look sweeter later on. Get complete appropriation at the beginning, with phased draws, and segregate it from other programs.

The second aspect is where senior management may place one group in competition with another and fund the perceived winner. The cold truth about these situations is that the group does not have the confidence of management to drive the program to a successful completion. If there is no confidence, there will be a lot of waste with this approach. Strive to get complete funding or at least an unencumbered pathway toward it.

3.2.6.3 External means to product development

3.2.6.3.1 Model for brand label, joint venture, and acquisitions

3.2.6.3.1.1 Perspectives on partnering

A brand label, joint venture, or acquisition for purposes of new business development can be a successful approach. These relationships allow the company to gain access to markets quickly with products that are on target.

Partnerships, however, do not last forever. They are simply a means to an end. Consequently, initial plans in negotiating a venture should include plans for how the company will absorb the technology, buy the company, or continue the partnership past the term. Failure to adequately plan this at the outset will result in consequences at the end of the term, with minimized alternatives and options. After all the effort to initiate a development and create a business, it can be devastating to lose control of the business because of a lack of transition planning at the beginning of the negotiation.

This section will present a model for brand labels, joint ventures, and acquisitions from a product development perspective. It can be used as a planning guide for initial discussions and as a check-off list for implementation.

3.2.6.3.1.2 Levels of investigation and discussion

There are four basic levels of investigation required to understand another partner's business. Depending on the framework of the partnership (i.e., brand label, joint venture, or acquisition), the more in-depth the investigation should be. An acquisition will require more information than will a brand label arrangement, for example.

The levels of investigation are therefore arranged in increasing depth into the partner organization. Level 1 is cursory, whereas level 4 delves into the core of the company—its finances and structure. Summary of the levels are discussed in the following.

3.2.6.3.1.2.1 Level 1

General. The general concept behind level 1 is to investigate if it is feasible to create a deal between the two companies. This section allows a summary of the opportunity to be documented.

Background. The background outlines the market driving forces and events that led up to the identified opportunity.

Market information (secondary). The market information cited here is secondary and is obtained from trade literature, periodicals, and industry-published data. It is differentiated from primary market data by the fact that it is information obtained indirectly rather than

through primary customer contact. It is useful as trend data and initial substantiation, but cannot be used as verifiable market data. That must come from primary customer contact.

1. Size, demographics
2. Market position
3. Market plan

Contacts (Who are the company contacts?)

1. Hierarchy (what is the organizational hierarchy?)
2. Key players

Structure of the deal

1. Diagramatic
2. Product flow
3. Support

3.2.6.3.1.3 Level 2

Relation to overall business development plan

Establish how the proposed partnership arrangement fits into your company's overall strategic plan. Is this a means to gain expertise and immediate entry to a marketplace? If so, what is the plan for after the term of the agreement?

Commercial

Take care in outlining the commercial implications of the arrangement. If a careless analysis is performed, the entire objective of the program could be missed.

1. Costs and pricing
   Who are the key players in both companies and who are the deal-makers and -breakers? What is the basic structure of the deal? This should be constrained to a single-page explanation that is understandable to all concerned. Diagram the relationship between the two companies and show what goes where, and where everything comes from. If this cannot be done easily and understandably, recheck the thinking of the opportunity and the partnering relationship. Something is not right. Establish where the product comes from and how it gets through the infrastructure to the customer. Place notations on the diagram to note key points, areas of concern, and potential failure areas. Show how the product will be supported during the relationship. If you can complete this in a relatively short amount of time without major conflicts, the basic tenets of a deal are in place to build on. What are the projected costs and prices? Is there enough margin for each of the players? What will happen if the marketplace puts price pressure on products, creating margin pressure?
2. Feature/benefit mix
   Carefully outline the feature and benefit matrix and tie it to price. Do you have the properly configured platform to take to market with your partner? Is the product properly positioned in the marketplace at the time of product introduction?
3. Ramp-up plan, volumes (reasonable and achievable?)
   Check to make sure that the two firms are capable of the market-driven ramp-up. Is there a manufacturing infrastructure to support the projected business levels? Are the business-generated unit volumes required to justify the program reasonable and achievable from a manufacturing perspective?
4. Primary market information to check secondary market information
   Be diligent in performing primary market research to verify feature acceptance, pricing, value, and trends in the marketplace. Does the product need to be changed to fit the updated market criteria?
5. Best-case scenario and effects
   One of the best ways to determine feasibility and reasonableness is to outline a best-case scenario and the assumptions needed to achieve it. Also project this best-case success onto the organizations (both yours and the partner organization) to determine the effect it may have. Will success change the philosophy of the agreement and the principles? Will the contributing organization feel well-served in the transaction, or will they tend to regroup and exercise an alternative strategy?
6. Worst-case scenario and effects
   Likewise, create a worst-case analysis to determine downside effects. Also include the effect of missed opportunities because of time pressure. This can be used to create a sense of urgency in situations that may need arbitration between the parties. Also take time to examine effects of a worst-case analysis in terms of actions that may be taken by both companies. If the contributing company senses that the venture will not work out to plan, they may have a tendency to regroup and refocus effort elsewhere, which will further jeopardize the venture.
7. Most likely scenario and effect
   Finally, determine the most likely scenario that could occur. Hopefully, this scenario will generate the best-desired result for both parties. Take time to also project your plans and proposals for after the term of the agreement, based on the success. This can provide the backdrop for future negotiations.
8. Product line analysis: second-generation
9. Order flow

Technical

1. Meets functional description
2. Meets relevant standards
3. Meets brand-labeled documentation

4. Manufacturability study and cost reduction
5. Quality system assessment

It is also helpful to project the second-generation product analysis at this time. It forces a realistic, critical examination of the features and benefits to determine market fit. If a fit is lacking certain features at this time, the second-generation analysis will highlight it and allow a judgment to be made. At this time it is also beneficial to generate a flow chart of product and order flow. Design the process by which product orders will be entered and processed within the organizations and delivered to the customer. If there is partnership beyond design and manufacturing through to the sales channel, then it becomes necessary to outline the protocol and flow of product here.

This area deals with the technical aspect of the arrangement. There are a host of elements that must be orchestrated between the two firms to effectively meet the customer's requirements, and this section will outline those requirements. Does the product, as designed, meet the functional requirements of the marketplace? Does it meet the basic features and performance criteria at a deliverable cost? Does the product, as designed, meet the relevant standards that govern the marketplace, or is redesign necessary to incorporate measures to meet these standards? Has the required documentation been effectively prepared to be acceptable to the marketplace? Is it consistent with the corporate requirements of the brand labeler? How about the manufacturability of the product? Are the manufacturing strategy and product platform open-ended or close-ended from an enhancement and cost-reduction point of view? What is the level of quality system in the contributing and utilizing corporation? Are the two systems consistent (because this is required if one corporation is to design and the other is to produce)? Are the systems capable of continuous improvement and tracking each other? This is important to resolve at this level because it is virtually impossible to reconstruct one in the midst of a product problem or recall situation.

Risk analysis

Finally, as part of a level 2 investigation it is desirable to perform some numerical financial risk assessment in. Also, what is the qualitative risk in terms of reputation and positioning in the marketplace? Look at the partner company's income statement to see trends and business cycles. Cyclical activity generates cyclical performance and rearrangement of priorities over time. This reprioritization can affect progress on a project. The balance sheet can indicate the structure and financial strength of the potential partner. Lack of financial strength may cause delays and possibly require additional contribution of cash on your company's part to complete the program.

3.2.6.3.1.4 Level 3

Review partner's business plan in relation to original estimates and management track record. A level 3 assessment will evaluate the partner company's ability to set plans in motion and execute them to completion. It is important to know this perspective, since the proposed venture will depend on their ability to be decisive in their actions. As an individual manager, it is also a good idea to do a self-assessment at this time in order to evaluate your own company's ability to carry out a plan.

Detailed review of financials and performance data. A level 3 assessment will involve a detailed analysis of the financial performance of the company. This is the level where a joint venture would require a form of due diligence investigation. Because a joint venture would require significant deployment of intellectual and hard assets, your company must perform enough of an investigation to protect these assets.

1. Income statements
2. Balance sheets
3. Funds flow
4. Structure of organization, number of shareholders and who they are—the interrelationships, the length of investments, interviews, and profile of investors. Get a feel for the ownership structure of the firm. Who are the key decision makers on finances and policy? Determine if there will be any planned change of ownership within the term of the deal. Much of this information may not be located in a single package, neatly wrapped up for your review. However, by listening carefully and gathering secondary information through third-party services (e.g., Dun & Bradstreet and Standard & Poors), you can amass a fair amount of information.
5. Review accounting and sales order entry systems for accuracy, compatibility, and consistency. Are the expenses recognized similarly from year to year? Is the accounting system accurate? Does the inventory system reasonably track the value of the inventory without adjustments every year-end? Can the sales order entry system be aligned with your system with some means of internal controls? Are the two systems effectively compatible enough to support the deal and keep track of the required data? What is the measurement of the cost of quality, and is it consistent with your system? The thought here is that the two systems need to be compatible enough to measure performance of the venture to the satisfaction of both parties.
6. Investigate the cost reporting system and cost rollups (target versus actual). Investigate the specifics of the cost roll-up system so that product costs are kept accurate from year to year. This will ensure that there

will be no disagreement on profits or margins in case the agreement calls for shared margin or percent of profit. If possible, it would also be of value to your company to assess the original and actual product costs historically within the partner company, because this will demonstrate their track record in achieving a design-to-cost regimen.

3.2.6.3.1.5 Level 4

Company fitability: Yours, theirs, and overall plans of both. A level 4 investigation is a more comprehensive analysis of the partner organization from an acquisition perspective. It is part and parcel of the total valuation of the company that will be purchased. It is the measurement of hard assets, know-how, and the ability to please the customer and widen its presence in the marketplace. However, a company is not a thing to be weighed and bought by the pound, so to speak. It is a living, dynamic organization composed of nonlinear, emotionally driven, and often unpredictable people whose diligence and effort contribute to the numbers being evaluated dispassionately. As such, people have hopes and dreams and goals. How are these transferred into the organization? What are the goals of the organization and how do they compare with your collection of contributors? Knowing this is essential to effecting a good marriage between the two companies. This effort must transcend the new product development portion of the agreement and permeate the relationship.

3.2.6.3.2 Valuation of company

The decision has been made to purchase the company, and now it becomes necessary to evaluate its worth. In many ways, worth is in the eye of the purchaser, and desired value on behalf of the seller is based on circumstance. In the simplest terms, there are a minimum of five ways to value a going concern: owner's investment value, a balance sheet analysis, a rate of return, income statement analysis, buyer investment valuation, a liquidation alternative, or a replacement value calculation. Each of these circumstantial ways may affect the perceived value of the following items; however, it is still necessary to have a starting point for each of these.

3.2.6.3.2.1 Assets, cash property, plants, and equipment

These hard assets comprise the visible portion of the balance sheet where a buyer can touch tangible things. Although these items are procured and depreciated over time to a very low value, they still represent a significant portion of the business's value at the time of acquisition. Watch out for overstated intangible assets in a deal.

3.2.6.3.2.2 Liabilities

These represent the obligations of the business. They are the bills and expenses that your company will incur after the transition.

3.2.6.3.2.3 Goodwill

This intangible represents the reputation of the organization in the marketplace (e.g., brand names and longevity of relationships with customers), and is often used to accumulate dollars beyond hard, identifiable assets in an acquisition. This is true where the purchase price exceeds hard asset value alone.

3.2.6.3.2.4 Orders

This is valuable from a study point of view to see where the orders are coming from and what the orders consist of. By examining the orders of an organization, several useful things can be learned. Elements such as number of orders, line items per order, purchasing authorization, and pricing and discount classifications all tell a story about the business in question.

3.2.6.3.2.5 Backlog

The backlog is useful in determining the inertia of the company. It provides the stabilizing influence during the transition. A large backlog gives the acquiring company some breathing room in making changes and absorbing the acquisition.

3.2.6.3.2.6 Receivables

This represents the monies owed to the company. A close examination of the accounts receivable records will indicate how the company does business. Are terms extended to customers as an inducement to offset product shortcomings? Are late payers overlooked as an appeasement for product issues? Are the receivable days out of step with the industry norm?

3.2.6.3.2.7 Inventory

By and large, the inventory is one of the most uncertain aspects of an acquisition. Of all of the tangible assets, the inventory is the only one that is recoverable out of operations. Inventory is of very low value on the street or to anyone other than the company. It must be called out by a clean set of product documentation and be useable as delivered without significant addition of labor to recover its cost. Aftermarket inventory support levels may be large due to obsolescence and growing aftermarket business. This is cause for further investigation.

3.2.6.3.2.8 Headcount

How does the headcount compare to industry norms? What is the dollar contribution for both direct and indirect employees? What are the trends in headcount, past and projected? What is the average longevity of the employees? A lot of information is available by researching this area thoroughly.

3.2.6.3.2.9 Trends analysis and pro forma projection

Finally, what does the trend analysis say about each of these measurements? Does it foretell a growing, vital organization or a stagnant one? What has the organization been through recently, and how are they positioned for change?

3.2.6.3.3 Structure the valuations with risk analysis

As a purchasing decision is being made, a summary of the valuations and options and their substantiating data needs to be reviewed. With each one, a risk analysis needs to be incorporated as part of the review. This risk analysis will shade the valuation with a bit of realism in pricing each alternative.

3.2.6.3.4 Comparative analysis with industry averages

Every company is different and must be evaluated based on its own merit and future contribution to your company. It is therefore only a measurement and not a valuation to compare industry averages between companies. In the acquisition arena, when the reason for the acquisition is to generate a new product, the comparative analysis and industry averages assessment can only be used as a guide. As mentioned earlier, there are five basic ways to value a business. They are described in the following.

3.2.6.3.4.1 Set the price five ways

Consider the following:

1. Owner's investment value
2. Rate of return
3. Buyer investment value
4. Liquidation value
5. Replacement value

Each of these has a justification for the valuation and each can be used to modify a valuation; however, there is an additional consideration that must be given when acquiring a business—when the primary driving force is to obtain intellectual property, trade secrets, or know-how.

If additional new business is to be built on the existing business, then the complexion of the original will change. Investments may have to be made, and profit structures and cost structures may be different. In these cases, it becomes necessary to perform a more in-depth analysis of the business and the pro forma investment in it, not only from an acquisition perspective but also from the incremental investment perspective.

3.2.6.3.4.2 Set terms

Finally, set the terms of the deal. If the partner company must generate a new product as part of the deal, then the sales price and terms of inflow of cash will alter accordingly because the intellectual value does not exist until these programs are well on their way to completion. When constructing deals in these areas of new product development, linkages and demonstrable results take on new importance as monies change hands.

3.2.6.3.4.3 Matrix of analysis

Figure 3.2.18 represents a matrix of analysis for the different product partnership arrangements discussed in this chapter. It is a matrix of essentials for investigation in negotiating a brand label, joint venture, or acquisition arrangement. Although this list is not presented in its entirety, it does serve as a useful tool in determining scope of investigation in an arrangement.

Matrix of essentials in a brand label, joint venture, or acquisition arrangement

	Brand label	Joint venture	Acquisition
Level of investigation			
Level 1	X	X	X
Level 2	X	X	X
Level 3	N/A	X	X
Level 4	N/A	N/A	X
Miscellaneous topics	X	X	X

Fig. 3.2.18 Matrix of analysis.

3.2.6.3.4.4 Miscellaneous topics to agree on

There are several miscellaneous topics that should form the basis for negotiations as a deal is put together. Although each is not a deal-maker or -breaker, they should be addressed from the perspective of completeness in that resolving the issues will make the deal go that much more smoothly. Each of the topics, along with a short explanation, is included here for your reference and should serve as a check-off list in negotiating a deal.

1. Exclusive rights
2. Domestic, North America, global, time frame
3. Payment for certifications
4. Product label—buyer, seller, both
5. Guaranteed sales volumes
6. Unit sales per month, default volumes, escape clauses
7. Monthly forecasts—production, advance notification of product changes
8. Salesforce (markets served)
9. Buyer's relationship to existing salesforce, channel management after the agreement
10. Buyer's channel profile
11. Seller's channel profile
12. Authorizations
13. Engineering changes, product evolution
14. Chain of command for interactions, organization chart
15. Returns based on final specifications
16. Payment responsibilities
17. Projects versus enhancements versus normal engineering support or applications engineering support
18. Individualized services
19. New products research and development
20. Field services, trouble calls, and authorizations for field returns
21. Travel costs
22. Warranty
23. Training—buyer's people, sales channel, field service
24. Research and development
25. All development projects through seller? Or, buyer? Or, combination?
26. Employee utilization and priority requirements if seller is a small firm
27. Projects tied to contract
28. Production—sellers
29. Quality control to be proved, meets minimum standards
30. First level of traceability
31. Gear-up time to manufacture
32. Training
33. Inventory levels
34. Product testing
35. Support needed from seller
36. Setup
37. Repair
38. Training
39. Lead times
40. Determine the lead time (in weeks) on 90-day projection
41. Schedule production with shipments starting on 91st day
42. Storage of units (warehousing, seller and buyer)
43. Production—buyer
44. Tied to quality requirements
45. Tied to quality specifications
46. Return authorization requirements
47. Default clauses
48. Sales
49. Payment
50. Quality
51. Manufacturing flow
52. Development
53. Cancellation clause by either party
54. Investment and stock exchange
55. Right of first refusal
56. Restrictions apply to transfer
57. Cap on controlling shares
58. Preemptive rights, month and year dilution of stock

3.2.6.3.4.5 Checklist for brand-label arrangements

The following is a checklist (Figure 3.2.19), that can be used for setting up a brand-label arrangement. It contains the necessary topics for discussion and resolution between two partner companies.

1. Diligence checklist for a joint venture
   - Figure 3.2.20 represents a diligence checklist that can be used when crafting a joint venture or even as an initial acquisition survey.
2. How to handle a forced arrangement
   - There are times when the new product manager may not be part of a brand label, joint venture, or acquisition decision but may be required to manage a new product development as part of it.
   - Often decisions are arrived at rather quickly without a lot of thought given to the details. If this is the case, there are several questions that you as the new product manager need to clarify as your company proceeds down this pathway.
   - There are five major areas of concern when a deal is forced into the organization. As a manager, it is in your best interest to resolve the issues defined in the five areas as part of the overall plan. The five areas were previously outlined in Figure 3.2.19 and the diligence checklist is presented in Figure 3.2.20.

Issues to resolve in a brand label arrangement				
Marketing				
	1	How was the marketing study done? Primary, secondary		
	2	What is the product cost structure?		
	3	Pricing, feature, and competitive comparison		
	4	Commercial feasibility		
	5	Fit with overall strategy		
Development				
	1	Profit and cost sensitivity		
	2	Functional testing		
	3	Quality track record		
	4	Technical feasibility/analysis		
	5	Standards testing and conformance		
	6	Design validation		
Channel Issues				
	1	Exclusive arrangements and noncompete		
	2	Reseller revenue structures		
	3	New market identification, development, and promotion		
	4	Intended route to market		
Production				
	1	Manufacturing plan		
	2	Manufacturing venue		
	3	Capital equipment		
	4	Manufacturing documentation		
	5	Change notices and support		
	6	Vendor linkages		
	7	Production testing		

Fig. 3.2.19 Brand label checklist.

Quality				
	1	Quality system and reporting		
	2	Corrective action system		
	3	Track record for similar products		
	4	Warranty costs		
	5	Field feedback system for improvements		
	6	Quality event chart		

Fig. 3.2.19 Cont'd.

Diligence check list					
Item	**Description**	**Responsibility**	**Substantiation**	**Date 1**	**Date 2**
1	Product opportunity fits scope of corporate strategy				
2	Consistent organizational system for evaluation				
3	Consistent criteria				
4	Documented and followed procedures				
5	Linking department objectives				
6	Sound market planning				
7	Positioning				
8	Segmentation				
9	Capitalization				
10	Pricing				
11	Paced development to capture opportunity				
12	Product distinctiveness				
13	Supporting evidence of opportunity				
14	Sizable market with research and forecast				
15	Precise product definition				
16	Solid design criteria and review				
17	Technical feasibility in the company				
18	Achieving target factory cost objectives				
19	Projected competitive response				
20	Stability of the team/company				

Fig. 3.2.20 Diligence checklist.

3.2.6.4 Keeping balance internally and externally

Managing a joint development program, or even a brand-label program, is an exercise in pleasing many principals. The marketplace has requirements that need to be met, the partner company has needs, and your company has needs. At times the demands of each may conflict with internal pressures, and the manager needs to arbitrate issues. Keeping these requirements in balance is essential to the program's success. The interrelationship between the issues and the individual factions needs to be identified at the outset and managed throughout the term of the deal. If one area fails to make its own share of the sacrifice, the relationship eventually will be in jeopardy.

Since you are charged with the responsibility of acting in your company's best interest, and the partner company is a participant, your role is one of chief negotiator; each issue must be resolved to the company's and market's satisfaction in a timely and decisive manner. Don't waste time getting to market by focusing on arbitration and trying to please everyone. Execute the deal and bring the product to market in a form that will be accepted by the market.

3.2.6.5 Dealing with entrepreneurs, licensing, and technology transfer

Dealing with technology in a raw, unprocessed form (as in the case of entrepreneurs) can be difficult. The challenge occurs when taking raw technology and embedding it into a manufacturing organization. There are differences between the entrepreneur and the manufacturer that cloud the perception of completeness and manufacturability. Inventors solve a customer's problem by creating "designs to suit your needs." The manufacturer serves a cross section of a market, builds unchanging product in volume, and resists change by its desire for uniformity demanded by its structure.

Both the entrepreneur and the manufacturer have similarities that must be preserved, such as desire for quality designs, commitment to customer satisfaction, and commitment to the venture. There are differences, however, that must be resolved, such as designs that must be frozen and manufacturable and problems that must be jointly and swiftly addressed. Both parties must diligently work toward equitable participation in the venture.

3.2.6.5.1 The benefits of short-term and long-term trade-offs

3.2.6.5.1.1 Double-edged sword

When sculpting a new company by executing a long-term growth plan, there is a balance that must be struck between short-term and long-term trade-offs. The long-term view requires investment and time not recognizable in the short term. This means that the manager and the organization must exercise patience in achieving the long-term objective and be decisive in preserving the investment during those times when it is financially tempting to terminate the expenses.

Negotiating and executing a brand label or joint venture agreement will require a certain amount of street smarts in navigating both companies through the relationship. Each of the company's allegiances, pressure points, sensitivities, and goals desired out of the deal must be recognized. If you watch where the monies change hands, it will be easier to understand the dynamics of the relationship.

3.2.6.5.1.2 Don't get held hostage

Administering a venture between two companies can leave your company vulnerable in that a commitment has been made to see the venture to completion. Alternatives are no longer as beneficial or available after you have locked onto the partner company. Consequently, you run a risk of being held hostage with their technology and know-how. Being held hostage can have long-lasting negative effects that transcend a single obstacle or crisis.

The only way to counteract an attempt by an employee or a partner company to hold you hostage is to have a good alternative waiting in the background to take over. In fact, you may want to provide for this at the outset in the contract clauses. In any event, the chances of being held hostage must be mitigated by having alternative means to implement the program and demonstrating the resolve to exercise it.

It is said over and over that a good deal is the result of a win–win situation for both parties. It is fundamentally based on the fact that if both parties do not benefit from the venture, sooner or later the deal will fall apart. Neither party will knowingly continue a relationship indefinitely without tangible benefit.

3.2.6.5.1.3 Making it last

Given that win–win is the preferred result, the best way to preserve a deal or to continue a deal that has come to term is to structure additional benefits for both parties. It takes a significant effort to make a deal work in the first place and, if one can be extended, additional work can be saved. Leveraging the ventures now adds to the worth of the company by continuing to achieve results without significant additional effort. In addition, the individuals involved begin to know each other and overall communication improves.

3.2.6.5.1.4 Benefits: quick access, experience, willing partners, and instant critical mass

To reiterate the benefits of a venture: your company gains quick access to a market and benefits from the partner's experience in the market or the technology, or both.

Sales channel contention						
Channel	Partner cost	Your cost	With your markup	End user's price	Partner's margin	Your margin
Partner	100.00	200.00	285.71	285.71	100.00	85.71
Direct	100.00	Bypass	Bypass	200.00	100.00	0
			"Partner's"	–85.71		
			Advantage			

Fig. 3.2.21 Channel contention.

With a willing and cooperative partner throughout the term of the venture, you can add critical mass to your company for a fraction of the investment cost.

3.2.6.5.1.5 Familiarity and market contention: exposing your position

If the partner company is bringing a product technology to the deal, and if your company is providing the marketing and channel, remember the "camel in the tent" syndrome. This describes a competitive situation whereby your company exposes its customer base to the partner company, who has the means to take the customers away from you.

It is not unheard of to have the partner company secure customers directly after a time and effectively remove you from the transaction, thus allowing the "camel in your tent." It is most dangerous when your company is not the manufacturer.

In this case, the manufacturer has a built-in financial advantage based on the minimum margin your company would use in selling the product. By bypassing your channel, they can afford to sell at your cost and still secure the same margin as if they were selling to you for resale. Figure 3.2.21 illustrates the issue.

As can be seen in Figure 3.2.21, the partner can bypass your company to get to the customer and effectively deliver the same goods at an advantage over your company, which preserves their manufactured margin. When negotiating a venture with a partner company, legislate steps in the agreement to ensure they remain actual partners.

3.2.6.5.1.6 Plan for the future

A final note on the subject of brand labels and joint ventures: it is important to keep in mind that, with the exception of the acquisition case, a partnership deal doesn't last forever. Two companies will eventually go their separate ways after a venture because each will develop different strategies, needs, and goals. The two companies may have originally partnered as a result of expediency, and that expediency will not continue forever. Therefore, it is wise to plan for the future by absorbing the technology and manufacturing know-how (or whatever the partner company brought to the deal) to position yourself for after the term of the venture. Participate in the venture at 100% effort and make it work for both parties, then plan for the future. Remember the objective is to create long-term growth and value for your company, so it's incumbent on the venture manager to provide this continuity.

3.2.7 Summary

This chapter's objective was to familiarize the reader with the basic concept of market opportunity. A market opportunity was related to the solving of a customer's problem and to making a profitable business out of the opportunity while satisfying the customer need.

Evaluation of ideas within the strategic framework of the business and competitive activities was presented. Tactics and strategy differences between large and small firms drive different results, and means for accomplishing those results through partnerships were presented as an alternative to executing a long-range plan. Finally, perspectives on partnering within the framework of new product development was discussed.

The reader should now have a basic, practical understanding of how a new product idea relates to a business and how it must relate to sustain the business in the long term. You should also have a working knowledge of partnership arrangements, their caveats, and benefits.

This serves as the basis for the next section, where the product opportunity and market opportunity will be refined into a product.

This page is intentionally left blank

Section **Four**

Design and materials

This page is intentionally left blank

Chapter 4.1

4.1

Materials and the environment

Michael Ashby

4.1.1 Introduction and synopsis

All human activity has some impact on the environment in which we live. The environment has some capacity to cope with this, so that a certain level of impact can be absorbed without lasting damage. But it is clear that current human activities exceed this threshold with increasing frequency, diminishing the quality of the world in which we now live and threatening the well-being of future generations. The manufacture and use of products, with their associated consumption of materials and energy, are among the culprits. The position is dramatized by the following statement: at a global growth rate of 3% per year we will mine, process, and dispose of more "stuff" in the next 25 years than in the entire history of mankind. *Design for the environment* is generally interpreted as the effort to adjust our present design methods to correct known, measurable, environmental degradation; the timescale of this thinking is 10 years or so, an average product's expected life. *Design for sustainability* is the longer view: that of adaptation to a lifestyle that meets present needs without compromising the needs of future generations. The timescale here is less clear — it is measured in decades or centuries — and the adaptation required is much greater. This chapter focuses on the role of materials and processes in achieving design for the environment. Sustainability requires social and political changes.

4.1.2 The material life cycle

The nature of the problem is brought into focus by examining the materials life cycle, sketched in Figure 4.1.1. Ore and feedstock, most of them non-renewable, are processed to give materials; these are manufactured into products that are used, and, at the end of their lives, disposed, a fraction perhaps entering a recycling loop, the rest committed to incineration or land-fill. Energy and materials are consumed at each point in this cycle (we shall call them "phases"), with an associated penalty of CO_2 and other emissions — heat, and gaseous, liquid, and solid waste.

The problem, crudely put, is that the sum of these unwanted by-products now often exceeds the capacity of the environment to absorb them. Some of the injury is local, and its origins can be traced and remedial action taken. Some is national, some global, and here remedial action has wider social and organizational prerequisites. Much present environmental legislation aims at modest reductions in the damaging activity; a regulation requiring 20% reduction in — say — the average gasoline consumption of passenger cars is seen by car-makers as a major challenge.

Sustainability requires solutions of a completely different kind. Even conservative estimates of the adjustment needed to restore long-term equilibrium with the environment envisage a reduction in the flows of Figure 4.1.1 by a factor of 4 or more. Population growth and the growth of the expectations of this population more than cancel any modest savings that the developed nations may achieve. It is here that the challenge is greatest, requiring difficult adaptation, and one for which no generally agreed solutions yet exist. But it remains the long-term driver of eco-design, to be retained as background to any creative thinking.

Materials Selection in Mechanical Design; ISBN: 9780750661683
Copyright © 1999 Elsevier Ltd. All rights of reproduction, in any form, reserved.

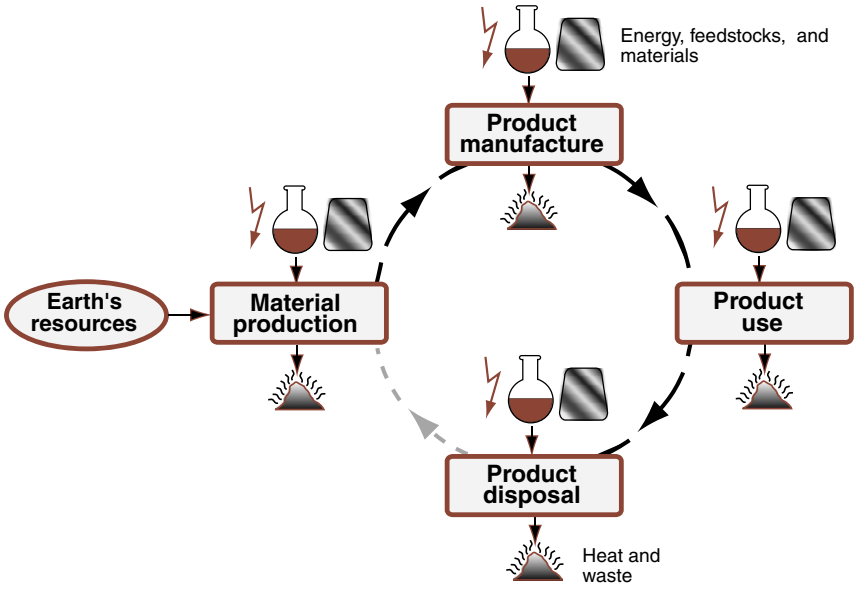


Fig. 4.1.1 The material life cycle. Ore and feedstock are mined and processed to yield a material. These are manufactured into a product that is used and at the end of its life, discarded or recycled. Energy and materials are consumed in each phase, generating waste heat and solid, liquid, and gaseous emissions.

4.1.3 Material and energy-consuming systems

It would seem that the obvious ways to conserve materials is to make products smaller, make them last longer, and recycle them when they finally reach the end of their lives. But the seemingly obvious can sometimes be deceptive. Materials and energy form part of a complex and highly interactive system, of which Figure 4.1.2 is a cartoon. Here primary catalysts of consumption such as *new technology, planned obsolescence, increasing wealth and education,* and *population growth* influence aspects of product use and through these, the consumption of materials and energy and the by-products that these produce. The connecting lines indicate influences; a red line suggests positive, broadly desirable influence; black line suggests negative, undesirable influence and red-black suggests that the driver has the capacity for both positive and negative influence.

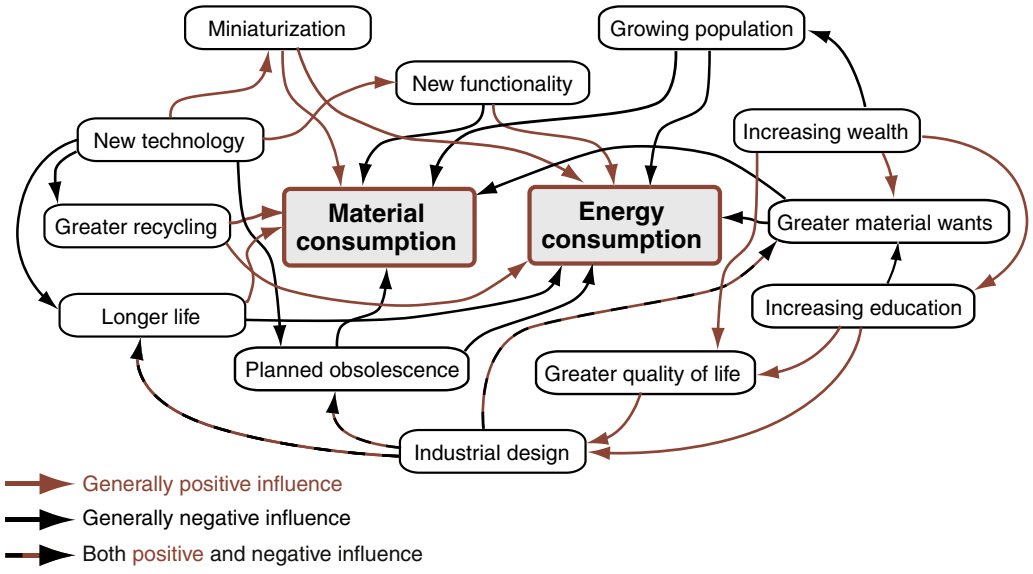


Fig. 4.1.2 The influences on consumption of materials and energy. It is essential to see eco-design as a systems problem, not solved by simple choosing "good" and avoiding "bad" materials, but rather that of matching the material to the system requirements.

The diagram brings out the complexity. Follow, for instance the lines of influence of *new technology* and its consequences. It offers more material and energy-efficient products, but by also offering new functionality it creates obsolescence and the desire to replace a product that has useful life left in it. Electronic products are prime examples of this: 80% are discarded while still functional. And observe, even at this simple level, the consequences of *longer life* — a seemingly obvious measure. It can certainly help to conserve materials (a positive influence) but, in an era in which new technology delivers more energy-efficient products (particularly true of cars, electronics and household appliances today), extending the life of old products can have a negative influence on energy consumption.

As a final example, consider the bi-valent influence of industrial design — the subject of Chapter 4.2. The lasting designs of the past are evidence of its ability to create products that are treasured and conserved. But today it is frequently used as a potent tool to stimulate consumption by deliberate obsolescence, creating the perception that "new" is desirable and that even slightly "old" is unappealing.

Table 4.1.1 Use matrix of product classes

	High load factor	Modest load factor	Low load factor	
Primary power-consuming	Family car Train set Aircraft	Television Freezer	Coffee maker Vacuum cleaner Washing machine	Energy Intensive
Secondary power-consuming	Housing (heat, light)	Car-park (light)	Household dishes Clothing (washing)	
Non-power-consuming	Bridges Roads	Furniture Bicycle	Canoe Tent	Material intensive
	High impact ⟷		Low impact	

4.1.3.1 The use-patterns of products

Table 4.1.1 suggests a matrix of product use-patterns. Those in the first row require energy to perform their primary function. Those in the second could function without energy but, for reasons of comfort, convenience, or safety, consume energy to provide a secondary function. Those in the last row provide their primary function without any need for energy other than human effort. The load factor across the top is an approximate indicator of the intensity of use — one that will, of course, vary widely.

The choice of materials and processes influences all the phases of Figure 4.1.1: *production*, through the drainage of resources and the undesired by-products of refinement; *manufacture*, through the level of efficiency and cleanness of the shaping, joining, and finishing processes; *use*, through the ability to conserve energy through light-weight design, higher thermal efficiency and lower energy consumption; and (finally) *disposal* through a greater ability to allow reuse, disassembly, and recycling.

It is generally true that one of the four phases of Figure 4.1.1 dominates the picture. Simplifying for a moment, let us take *energy consumption* as a measure of both the inputs and undesired by-products of each phase and use it for a character-appraisal of use-sectors. Figure 4.1.3 presents the evidence, using this measure. It has two significant features, with important implications. First, one phase almost always dominates, accounting for 80% or more of the energy consumed during life — often much more. If large changes are to be achieved, it is this phase that must be the target; a reduction by a factor of 2, even of 10, in any other makes little significant difference to the total. The second: when differences are as great as those of Figure 4.1.3, precision is not the issue — an error of a factor of 2 changes very little. It is the nature of people who *measure* things to wish to do so with precision, and precise data must be the ultimate goal. But it is possible to move forward without it: precise judgments can be drawn from imprecise data. This is an important consideration: much information about eco-attributes is imprecise.

4.1.4 The eco-attributes of materials

4.1.4.1 Material production: energy and emissions

Most of the energy consumed in the four phases of Figure 4.1.1 is derived from fossil fuels. Some is consumed in that state — as gas, oil, coal, or coke. Much is first converted to electricity at a European average conversion efficiency of about 30%. Not all electricity is generated from fossil fuels — there are contributions from hydroelectric, nuclear, and wind/wave generation. But with the exception of Norway (70% hydro) and France (80% nuclear) the predominant energy sources are fossil fuels; and since the national grids of European countries are linked, with power flowing from one to

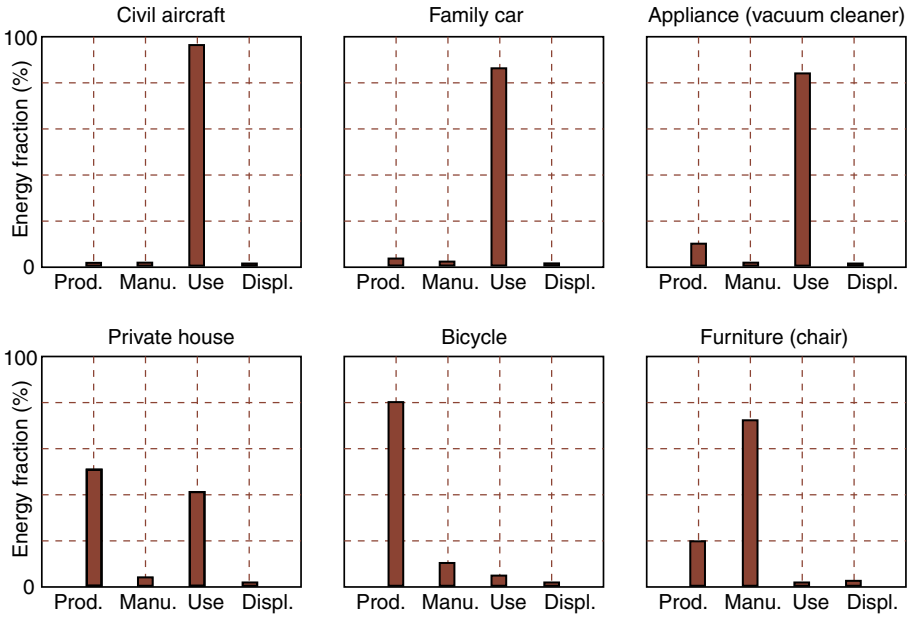


Fig. 4.1.3 Approximate values for the energy consumed at each phase of Figure 4.1.1 for a range of products.

another as needed, it is a reasonable approximation to speak of a European average fossil-fuel energy per kilowatt of delivered electrical power.

The fossil-fuel energy consumed in making one kilogram of material is called its *production energy*. Some of the energy is stored in the created material and can be reused, in one sense or another, at the end of life. Polymers made from oil (as most are) contain energy in another sense — that of the oil that enters the production as a primary feedstock. Natural materials such as wood, similarly, contain "intrinsic" or "contained" energy, this time derived from solar radiation absorbed during growth. Views differ on whether the intrinsic energy should be included in the production energy or not. There is a sense in which not only polymers and woods, but also metals, carry intrinsic energy that could — by chemical reaction or by burning the metal in the form of finely divided powder — be recovered, so omitting it when reporting production energy for polymers but including it for metals seems inconsistent. For this reason we will include intrinsic energy from non-renewable resources in reporting production energies, which generally lie in the range 25–250 MJ/kg, though a few are much higher. The existence of intrinsic energy has another consequence: that the energy to recycle a material is sometimes much less than that required for its first production, because the intrinsic energy is retained. Typical values lie in the range 10–100 MJ/kg.

The production of 1 kg of material is associated with *undesired gas emissions*, among which CO_2, NO_x, SO_x, and CH_4 cause general concern (global warming, acidification, ozone-layer depletion). The quantities can be large — each kilogram of aluminum produced by using energy from fossil fuels creates some 12 kg of CO_2, 40 g of NO_x and 90 g of SO_x. Production is generally associated with other undesirable outputs, particularly toxic wastes and particulates, but these can, in principle, be dealt with at source.

Wood, bamboo, and other plant-based materials, too, contain intrinsic energy, but unlike man-made materials it derives from sunlight, not from non-renewable resources. The production energy data for these materials does not include this intrinsic energy, and the emissions take account of the CO_2 absorbed during their growth. Because of this, woods have a near-neutral energy balance, and a negative value for CO_2 emissions.

4.1.4.2 Material processing energies at 30% efficiency

Many processes depend on casting, evaporation or deformation. It is helpful to have a feel for the approximate magnitudes of energies required by these.

4.1.4.2.1 Melting

To melt a material, it must first be raised to its melting point, requiring a minimum input of the heat $C_p(T_m - T_0)$, and then caused to melt, requiring the latent heat of melting, L_m

$$H_{min} = C_p(T_m - T_0) + L \qquad (4.1.1)$$

where H_{min} is the minimum energy per kilogram for melting, C_p is the specific heat, T_m is the melting point,

T_0 is the ambient temperature. A close correlation exists between L_m and C_pT_m

$$L_m \approx 0 : 4C_pT_m \tag{4.1.2}$$

and for metals and alloys $T_m > T_0$ giving

$$H_{min} \approx 1 : 4C_pT_m \tag{4.1.3}$$

Assuming efficiency of 30%, the estimated energy to melt 1 kg, H_m, is

$$H_m^* \approx 4.2C_pT_m \tag{4.1.4}$$

the asterisk recalling that it is an estimate. For metals and alloys, the quantity H_m^*, lies in the range 0.4–4 MJ/kg.

4.1.4.2.2 Vaporization

As a rule of thumb the latent heat of vaporization, L_v, is larger than that for melting, L_m, by a factor of 24 ± 5, and the boiling point T_b is larger than the melting point, T_m, by a factor 2.1 ± 0.5. Using the same assumptions as before, we find an estimate for the energy to evaporate 1 kg of material (as in PVD processing) to be

$$H_v^* \approx 38C_pT_m \tag{4.1.5}$$

again assuming an efficiency of 30%. For metals and alloys, the quantity H_v lies in the range 3–30 MJ/kg.

4.1.4.2.3 Deformation

Deformation processes like rolling or forging generally involve large strains. Assuming an average flow-strength of $(\sigma_y + \sigma_{uts})/2$, a strain of $\varepsilon = 90\%$ and an efficiency factor of 30% we find, for the work of deformation per kilogram to be

$$W_D^* \approx 1.5(\sigma_y + \sigma_{uts})\varepsilon = 1.35(\sigma_y + \sigma_{uts}) \tag{4.1.6}$$

where σ_y is the yield strength and σ_{uts} is the tensile strength. For metals and alloys, the quantity W_D lies in the range 0.01–1 MJ/kg.

We conclude that casting or deformation require processing energies that are small compared to the production energy of the material being processed, but the larger energies required for vapor-phase processing may become comparable with those for material production.

4.1.4.3 End of life

Quantification of material *recycling* is difficult. Recycling costs energy, and this energy carries its burden of gases. But the *recycle energy* is generally small compared to the initial production energy, making recycling — when it is possible at all — an energy-efficient propositions. It may not, however, be one that is cost efficient; that depends on the degree to which the material has become dispersed. In-house scrap, generated at the point of production or manufacture, is localized and is already recycled efficiently (near 100% recovery). Widely distributed "scrap" — material contained in discarded products — is a much more expensive proposition to collect, separate and clean. Many materials cannot be recycled, although they may still be reused in a lower-grade activity; continuous-fiber composites, for instance, cannot be re-separated economically into fiber and polymer in order to recycle them, though they can be chopped and used as fillers. Most other materials require an input of virgin material to avoid build-up of uncontrollable impurities. Thus the fraction of a material production that can ultimately re-enter the cycle of Figure 4.1.1 depends both on the material itself and on the product into which it has been incorporated. Despite this complexity, some data for the fraction re-entering the cycle of Figure 4.1.1 are available. More usually, the position is characterized by indicating simply whether the material can or cannot be *recycled*, *down-cycled*, *biodegraded*, *incinerated*, or committed to *landfill*.

4.1.4.3.1 Bio-data

Some materials are toxic, creating potential problems during production, during use, and particularly in their disposal.

4.1.4.4 Aggregated measures: eco-indicators

Table 4.1.2 shows an example of the eco-attributes of a material, here a 1000 series aluminum alloy. A designer, seeking to cope with many interdependent decisions that any design involves, finds it hard to know how best to use these data. How are production energies or CO_2 and SO_x burdens to be balanced against resource depletion, toxicity or ease of recycling? This perception has lead to efforts to condense the eco-attributes of a material into a single measure or *indicator*, giving the designer a simple, numeric ranking.

To do this, four steps are necessary. They are shown in Figure 4.1.4. The first is that of *classification* of the data listed in Table 4.1.2 according to the impact each causes (global warming, ozone depletion, acidification, etc.), giving an eco-profile of the material. The second step is that of *normalization* to remove the units (of which there are several in Table 4.1.2) and reduce them to a common scale (0–100, for instance). The third step is that of *weighting* to reflect the perceived seriousness of each impact, based on the classification of Step 1: thus global warming might be seen as more serious than resource depletion, giving it a larger weight. In the final step, the

Table 4.1.2 The eco-attributes of one grade of aluminum alloy

Wrought 1200 Grade Aluminum			
Material production: energy and emissions			
Production energy		1.9e2–2.1e2	MJ/kg
Carbon dioxide	*	12–13	kg/kg
Nitrogen oxides	*	72–79	g/kg
Sulfur oxides	*	1.2e2–1.4e2	g/kg
Indicators for principal component			
Eco indicator		7.4e2–8.2e2	millipoints/kg
Material processing energy at 30% efficiency			
Min. energy to melt		3.5–3.8	MJ/kg
Min. energy to vaporization		29–32	MJ/kg
Min. energy to 90% deform.		0.04–0.044	MJ/kg
End of life			
Recycle	Yes		
Downcycle	Yes		
Biodegrade	No		
Incinerate	No		
Landfill	Acceptable		
Recycling energy	*	23–26	MJ/kg
Recycle fraction		34–38	%
Bio-data			
Toxicity rating	Non-toxic		

weighted, normalized measures are *summed* to give the indicator.

The use of a single-valued indicator is criticized by some. The grounds for criticism are that there is no agreement on normalization or weighting factors, that the method is opaque since the indicator value has no simple physical significance, and that defending design decisions based on a measurable quantity like energy consumption or CO_2 generation carries more conviction than doing so with an indicator. At the time of writing there is no general agreement on how best to use eco-data in design. But on one point there is international agreement (the Kyoto Protocol of 1997): that the developed nations should progressively reduce CO_2 emissions — a considerable challenge in times of industrial growth, increasing affluence and growing population. Thus there is a certain logic in using CO_2 emission as the "indicator", though it is more usual at present to use energy.

4.1.5 Eco-selection

To select materials to minimize the impact on the environment we must first ask — as we did in Section 4.1.2 — which phase of the life cycle of the product under consideration makes the largest contribution? The answer guides the choice of strategy to ameliorate it (Figure 4.1.5). The strategies are described below.

4.1.5.1 The material production phase

If material production is the dominant phase of life it is this that we must target. Drink containers (Figure 4.1.6) provide an example: they consume materials and energy during material extraction and production, but (apart

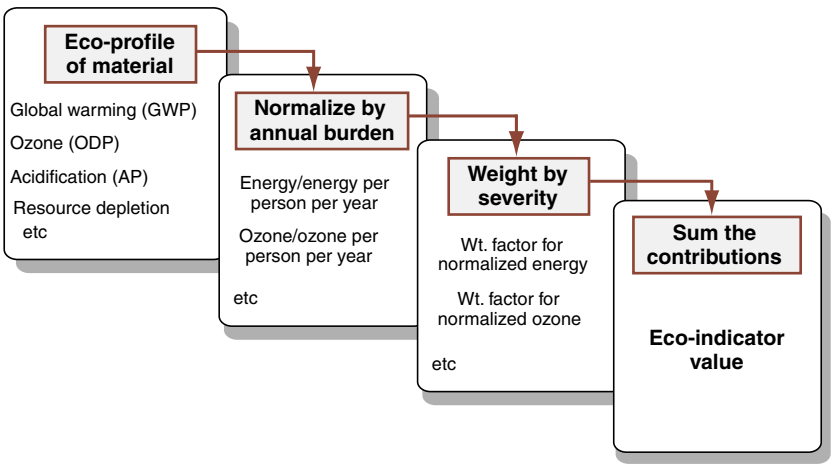


Fig. 4.1.4 The steps in calculating an eco-indicator. Difficulty arises in step 3: there is no agreement on how to choose the weight factors.

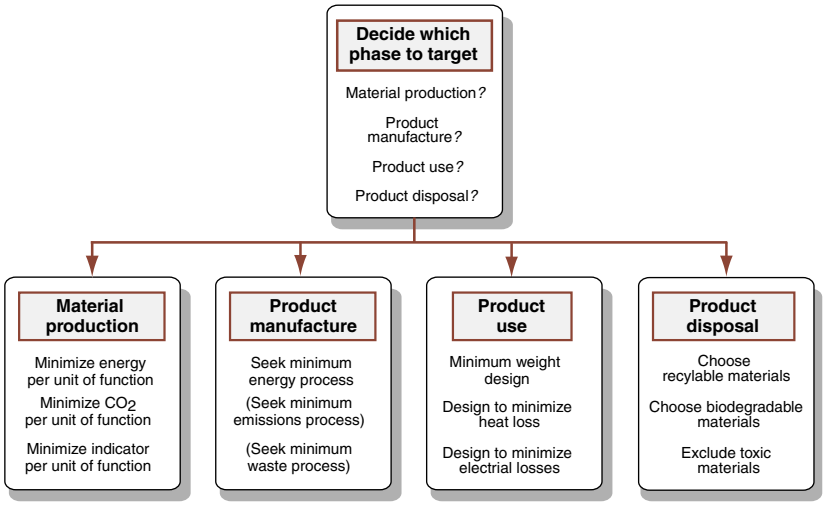


Fig. 4.1.5 Rational design for the environment starts with an analysis of the phase of life to be targeted. This decision then guides the method of selection to minimize the impact of the phase on the environment.

from transport, which is minor) not thereafter. We use the energy consumed in extracting and refining the material (the production energy of Table 4.1.2) as the measure; CO_2, NO_x, and SO_x emissions are related to it, although not in a simple way. The energy associated with the production of 1 kg of a material is H_p, that per unit volume is $H_p\rho$ where ρ is the density of the material. The bar-charts of Figures 4.1.7(a) and (b) show these two quantities for ceramics, metals, polymers, and composites. On a "per kg" basis (upper chart) glass, the material of the first container, carries the lowest penalty. Steel is higher. Polymer production carries a much higher burden than does steel. Aluminum and the other light alloys carry the highest penalty of all. But if these same materials are compared on a "per m^3" basis (lower chart) the conclusions change: glass is still the lowest, but now commodity polymers such as PE and PP carry a *lower* burden than steel; the composite GFRP is only a little higher. But is comparison "per kg" or "per m^3" the right way to do it? Rarely. To deal with environmental impact at the production phase properly we must seek to minimize the energy, the CO_2 burden or the eco-indicator value *per unit of function*.

Material indices that include energy content are derived in the same way as those for weight or cost. An example: the selection of a material for a beam that must meet a stiffness constraint at minimum energy content. The best materials to minimize production energy of a beam of specified stiffness and length are those with large values of the index

$$M = \frac{E^{1/2}}{H_p\rho} \tag{4.1.7}$$

where E is the modulus of the material of the beam. The stiff tie of minimum energy content is best made of a material of high $E/H_p\rho$; the stiff plate, of a material with high $E^{1/3}/H_p\rho$ and so on.

Fig. 4.1.6 Liquid containers: glass, polyethylene, PET, and aluminum. All can be recycled. Which carries the lowest production energy penalty?

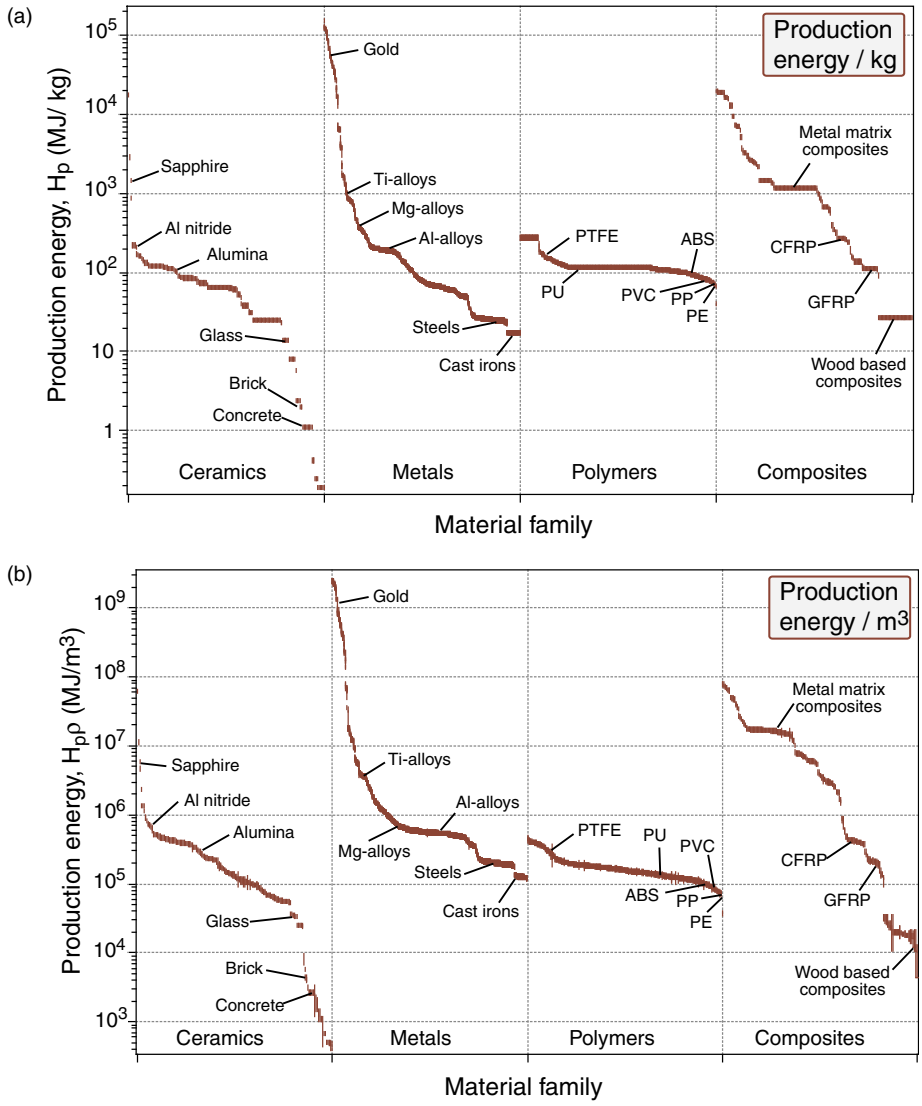


Fig. 4.1.7 (a and b). The energy per unit mass and per unit volume associated with material production.

Strength works the same way. The best materials for a beam of specified bending strength and minimum energy content are those with large values of

$$M = \frac{\sigma^{2/3}}{H_p \rho} \tag{4.1.8}$$

where σ_f is the failure strength of the beam-material. Other indices follow in a similar way.

Figures 4.1.8 and 4.1.9 are a pair of materials selection charts for minimizing production energy H_p per unit of function (similar charts for CO_2 burden can be made using the CES 4 software). The first show modulus E plotted against $H_p\rho$; the guidelines give the slopes for three of the commonest performance indices. The second shows strength σ_f plotted against $H_p\rho$; again, guidelines give the slopes. The two charts give a survey data for minimum energy design. They are used in exactly the same way as the $E - \rho$ and $\sigma_f - \rho$ charts for minimum mass design.

Most polymers are derived from oil. This leads to statements that they are energy intensive, with implications for their future. The two charts show that, per unit of function in bending (the commonest mode of loading), most polymers carry a lower energy penalty than primary aluminum, magnesium, or titanium, and that several are competitive with steel. Most of the energy consumed in the production of metals such as steel, aluminum, or magnesium is used to reduce the ore to the elemental metal, so that these materials, when recycled, require much less energy. Efficient collection and recycling makes important contributions to energy saving.

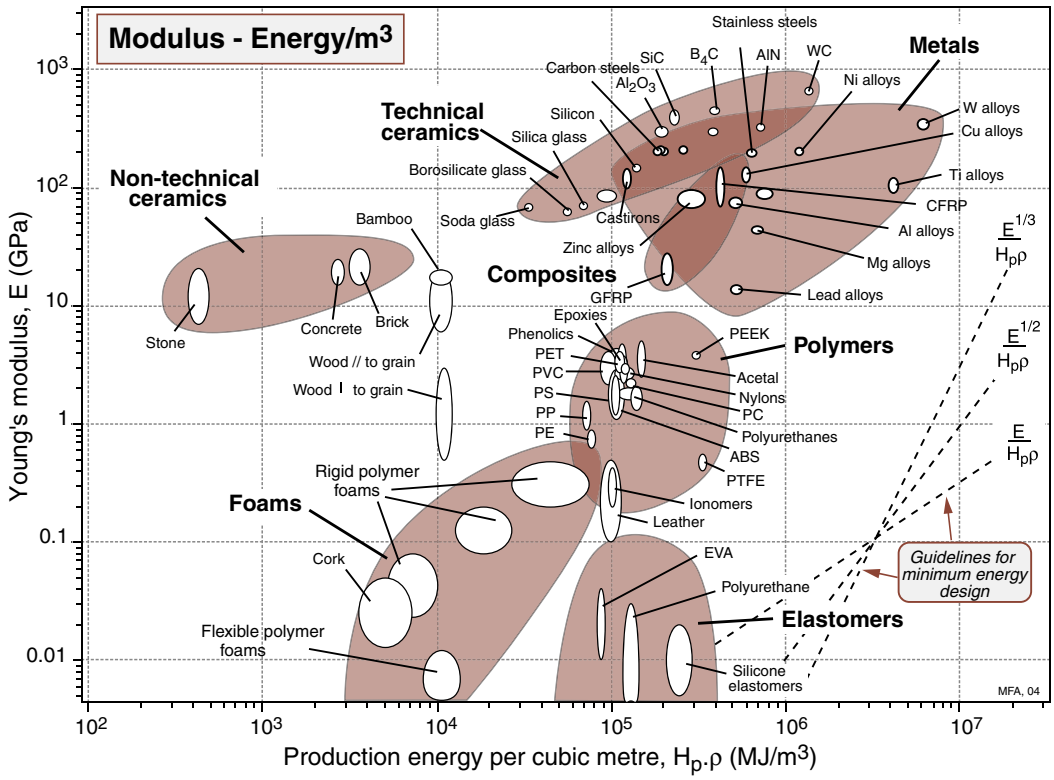


Fig. 4.1.8 A selection chart for stiffness with minimum production energy.

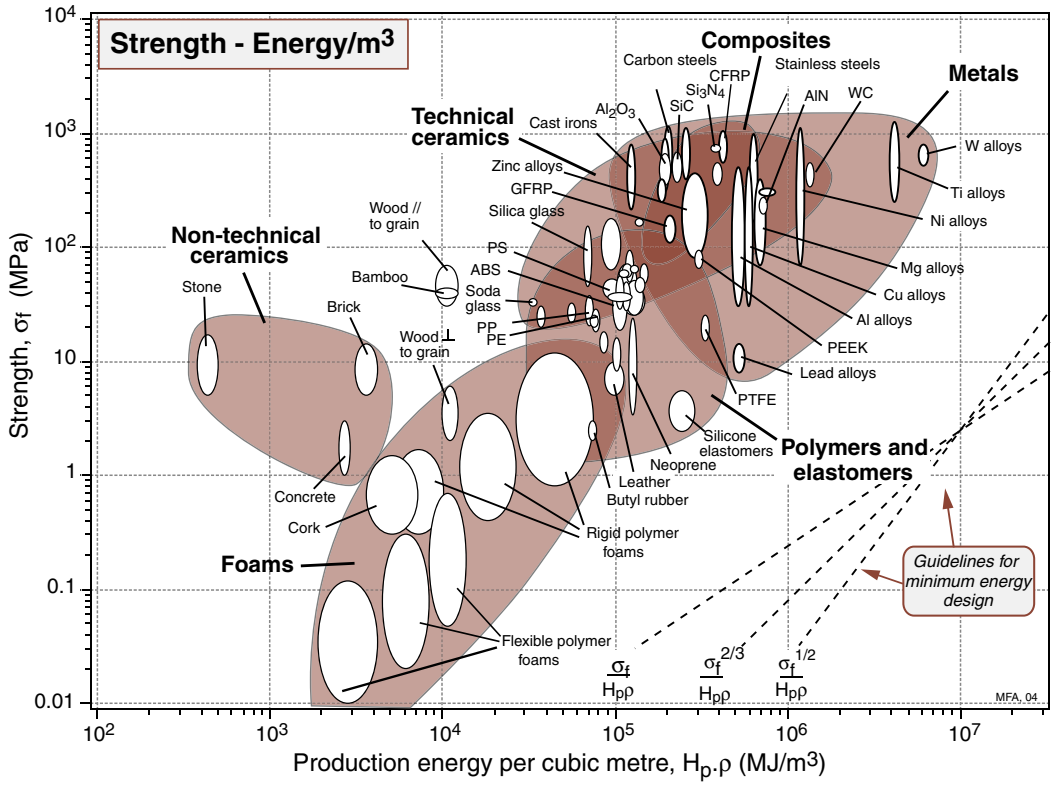


Fig. 4.1.9 A selection chart for strength with minimum production energy.

Table 4.1.3 Examples of indices to minimize impact in the production phase

Function	Maximize*
Minimum energy content for given tensile stiffness	$E/H_p\rho$
Minimum CO_2 emissions for given tensile strength	$\sigma_y/[CO_2]\ \rho$
Minimum energy for given bending stiffness (beam)	$E^{1/2}/H_p\rho$
Minimum energy for given bending stiffness (panel)	$E^{1/3}/H_p\rho$
Minimum CO_2 for given bending strength (beam)	$\sigma_f^{2/3}/[CO_2]\rho$
Minimum CO_2 for given bending strength (panel)	$\sigma_f^{1/3}/[CO_2]\rho$
Minimum eco-indicator points for given thermal conduction	$\lambda/$le P

*H_p = production energy content per kg; $[CO_2] = CO_2$ production per kg; le = eco-indicator per kg.

Table 4.1.3 contains examples of other indices for eco-selection to minimize impact during the production phase of life.

4.1.5.2 The product-manufacture phase

The obvious message of Section 4.1.3 is that vapor-forming methods are energy intensive; casting, and deformation processing are less so. Certainly it is important to save energy in production. But higher priority often attaches to the *local* impact of emissions and toxic waste during manufacture, and this depends crucially on local circumstances. Paper-making (to take an example) uses very large quantities of water. Historically the waste water was heavily polluted with alkalis and particulates, devastating the river systems into which it was dumped. Today, the best paper mills discharge water that is as clean and pure as it was when it entered. Production sites of the former communist-block countries are terminally polluted; those producing the same materials elsewhere, using best-practice methods, have no such problems. Clean manufacture is the answer here.

4.1.5.3 The use phase

The eco-impact of the use phase of energy-consuming products has nothing to do with the energy content of the materials themselves — indeed, minimizing this may frequently have the opposite effect on use-energy. Use-energy depends on mechanical, thermal and electrical efficiencies; it is minimized by maximizing these efficiencies. Fuel efficiency in transport systems (measured, say, by MJ/km) correlates closely with the mass of the vehicle itself; the objective becomes that of minimizing mass. Energy efficiency in refrigeration or heating systems is achieved by minimizing the heat flux into or out of the system; the objective is then that of minimizing thermal conductivity or thermal inertia. Energy efficiency in electrical generation, transmission, and conversion is maximized by minimizing the ohmic losses in the conductor; here the objective is to minimize electrical resistance while meeting necessary constraints on strength, cost, etc.

4.1.5.4 The product disposal phase

The environmental consequences of the final phase of product life have many aspects. The aggregation of these into the single "indicator" does not appear to be a helpful path to follow. Limited data are available for recycle energies, and for the fraction of current supply currently met by recycling. The simple Boolean (yes/no) classification, shown in Table 4.1.3, signals that a given material can be recycled, reused in a lower grade activity, bio-degraded, incinerated, or committed to landfill.

4.1.6 Case studies: drink containers and crash barriers

The methods are illustrated below by case studies.

4.1.6.1 The energy content of containers

4.1.6.1.1 The problem

The containers of Figure 4.1.6 are examples of products for which the first and second phases of life — material production and product manufacture — are ones that consume energy. Thus material selection to minimize energy and consequent gas and particle emissions focuses on these. Table 4.1.4 summarizes the requirements.

The masses of five competing container-types, the material of which they are made, and the specific energy content of each are listed Tables 4.1.5a and 4.1.5b. Their production involves molding or deformation; approximate energies for each are listed. All

Table 4.1.4 Design requirements for the containers

Function	Container for cold drink
Constraints	Must be recyclable
Objective	Minimize production energy per unit capacity
Free variables	Choice of material

five of the materials can be recycled. Which container-type carries the lowest overall energy penalty per unit of fluid contained?

4.1.6.1.2 The method and results

A comparison of the energies in Tables 4.1.5(a) and 4.1.5(b) show that the energy to shape the container is always less than that to produce the material in the first place. Only in the case of glass is the forming energy significant. The dominant phase is that of material production. Summing the two energies for each material and multiplying by the container-mass per liter of capacity gives the ranking shown in the second last column of Table 4.1.5(a). The steel can carry the lowest energy penalty, glass and aluminum the highest.

4.1.6.2 Crash barriers

4.1.6.2.1 The problem

Barriers to protect driver and passengers of road vehicles are of two types: those that are static — the central divider of a freeway, for instance — and those that move — the fender of the vehicle itself (Figure 4.1.10). The static type line tens of thousands of miles of road. Once in place they consume no energy, create no CO_2 and last a long time. The dominant phases of their life in the sense of Figure 4.1.1 are those of material production and manufacture. The fender, by contrast, is part of the vehicle; it adds to its weight and thus to its fuel consumption. The dominant phase here is that of use. This means that, if eco-design is the objective, the criteria for selecting materials for the two sorts of barrier will differ. We take as a criterion for the first that of maximizing the energy that the barrier can absorb per unit of production energy; for the second we take absorbed energy per unit mass. Table 4.1.6 summarizes.

Table 4.1.5a Details of the containers

Container type	Material	Mass (g)	Mass/liter (g)	Energy/liter (MJ/l)
PET 400 ml bottle	PET	25	62	5.4
PE 1 milk bottle	High density PE	38	38	3.2
Glass 750 ml bottle	Soda glass	325	433	8.2
Al 440 ml can	5000 series Al alloy	20	45	9.0
Steel 440 ml can	Plain carbon steel	45	102	2.4

Table 4.1.5b Data for the materials of the containers

Material	Production energy (MJ/kg)	Forming method	Forming energy (MJ/kg)
PET	84	Molding	3.1
PE	80	Molding	3.1
Soda glass	14	Molding	4.9
5000 series Al alloy	200	Deep drawing	0.13
Plain carbon steel	23	Deep drawing	0.15

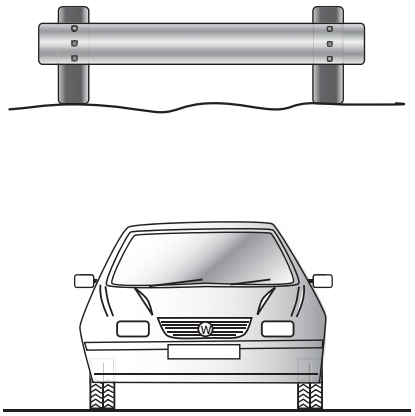


Fig. 4.1.10 Two crash barriers, one static, the other – the fender – attached to something that moves. Different eco-criteria are needed for each.

4.1.6.2.2 The method and results

The energy absorbed per unit volume by a material when stretched until it fractures is approximately

$$U = \int_0^{\varepsilon_f} \sigma d\varepsilon \approx \sigma_y \varepsilon_f \ \mathrm{MJ/m^3}$$

Here σ_y is the yield strength and ε_f is the elongation to fracture. Crash barriers are loaded in more complex ways

Table 4.1.6 Design requirements for the crash barriers

Function	Energy-absorbing crash barriers
Constraints	Must be recyclable
Objective	• Maximize energy absorbed per unit production energy, or • Maximize energy absorbed per unit mass
Free variables	Choice of material

that this, but when the part that is being stretched fractures, the barrier looses effectiveness. The materials with the largest U per unit production energy are those with large values of

$$M = \frac{\sigma_f \varepsilon_f}{H_p \rho} \tag{4.1.9}$$

Those that do so per unit mass are those with large

$$M = \frac{\sigma_f \varepsilon_f}{\rho} \tag{4.1.10}$$

Figures 4.1.11(a) and (b) compare metals and polymers on this basis. The first guides selection for static barriers. Among metals, carbon steels are the best choice. Barriers made of polyethylene or polypropylene could be equally effective. The second figure guides selection for barriers that move. Among metals, titanium alloys offer the lightest solution; steels and light alloys are almost as good and much cheaper. But here polymers become particularly attractive: a polymeric fender, potentially, is much lighter than a metal one.

4.1.7 Summary and conclusions

Rational selection of materials to meet environmental objectives starts by identifying the phase of product-life that causes greatest concern: production, manufacture, use, or disposal. Dealing with all of these requires data not only for the obvious eco-attributes (energy, CO_2 and other emissions, toxicity, ability to be recycled and the like) but also data for mechanical, thermal, electrical, and chemical properties. Thus if material production is the phase of concern, selection is based on minimizing production energy or the associated emissions (e.g. CO_2 production). But if it is the use-phase that is of concern, selection is based instead on light weight, excellence as a thermal insulator, or as an electrical conductor (while meeting other constraints on stiffness, strength, cost, etc.). This chapter develops methods to deal with these. The methods are most effective when implemented in software. The Edu version of the CES 4 system has limited eco data; the more specialized Eco-selector has much more.

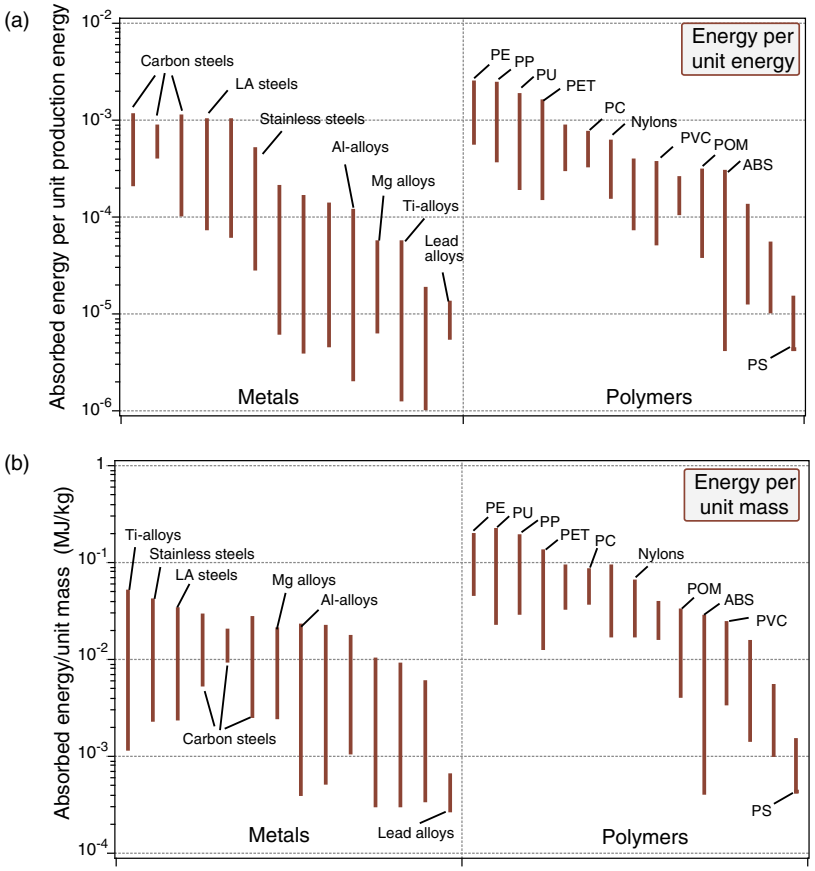


Fig. 4.1.11 (a) The energy absorbed by a material in plastic deformation, per unit of production energy and (b) the energy absorbed per unit mass.

Further reading

CES'05 (2005) *The Cambridge Engineering Selector*, Granta Design, Cambridge (*The material selection platform now has an optional eco-design module.* www.grantadesign.com)

Eternally Yours (2001) www.eternally-yours.nl *(The Eternally Yours Foundation seeks to analyze why people surround themselves more and more with products they feel less and less attached to, and to explore the design of products that retain their appeal over long periods of time.)*

Fuad-Luke, A. (2002) *The Eco-Design Handbook*, Thames and Hudson, London, UK. ISBN 0-500-28343-5. *(A remarkable source-book of examples, ideas and materials of eco-design.)*

Goedkoop, M.J., Demmers, M. and Collignon, M.X. (1995) *Eco-Indicator '95, Manual*, Pré Consultants, and the Netherlands Agency for Energy and the Environment, Amersfort, Holland. ISBN 90-72130-80-4.

Goedkoop, M., Effting, S. and Collignon, M. (2000) The Eco-Indicator 99: *A Damage Oriented Method for Life Cycle Impact Assessment, Manual for Designers. (PRe Consultants market a leading life-cycle analysis tool and are proponents of the eco-indicator method.* http://www.pre.nl)

ISO 14001 (1996) and ISO 14040 (1997, 1998, 1999) International Organisation for Standardisation (ISO), Environmental management system-specification with guidance for use, Geneva, Switzerland.

Kyoto Protocol (1997) United Nations, Framework Convention on Climate Change. Document FCCC/CP1997/7/ADD.1 Environmental management – life cycle assessment and subsections, Geneva, Switzerland. *(An agreement between the developed nations to limit their greenhouse gas emissions relative to the levels emitted in 1990. The US agreed to reduce emissions from 1990 levels by 7% during the period 2008–2012; European nations adopted a more stringent agreement.)*

Lovins, L.H., von Weizsäcker, E. and Lovins, A.B. (1998) *Factor Four: Doubling Wealth, Halving Resource Use*, Earthscan publications, London, UK. ISBN 1-853834-068. *(An influential book arguing that resource productivity can and should grow fourfold, that is, the wealth extracted from one unit of natural resources can quadruple allowing the world to live twice as well yet use half as much.)*

Mackenzie, D. (1997) Green design: design for the environment, 2nd edition, Lawrence King Publishing, London, UK. ISBN 1-85669-096-2. *(A lavishly produced compilation of case studies of eco design in architecture, packaging, and product design.)*

Meadows, D.H., Meadows, D.L., Randers, J. and Behrens, W.W. (1972) *The Limits to Growth – 1st Report of the Club of Rome*, Universe Books, New York, USA. *(A pivotal publication, alerting the world to the possibility of resource depletion, undermined by the questionable quality of the data used for the analysis, but despite that, the catalyst for subsequent studies and for views that are now more widely accepted.)*

Schmidt-Bleek, F. (1997) How much environment does the human being need – factor 10 – the measure for an ecological economy, Deutscher Taschenbuchverlag, Munich, Germany. *(The author argues that true sustainability requires a reduction in energy and resource consumption by the developed nations by a factor of 10.)*

Wenzel, H. Hauschild, M. and Alting, L. (1997) *Environmental Assessment of Products*, Vol. 1, Chapman and Hall, London, UK. *(Professor Alting leads Danish research in eco-design.)*

This page is intentionally left blank

Chapter 4.2

4.2

Materials and industrial design

Michael Ashby

4.2.1 Introduction and synopsis

Good design works. Excellent design also gives pleasure.

Pleasure derives from form, color, texture, feel, and the associations that these invoke. Pleasing design says something about itself; generally speaking, honest statements are more satisfying than deception, though eccentric or humorous designs can be appealing too.

Materials play a central role in this. A major reason for introducing new materials is the greater freedom of design that they allow. Metals, in the past century, allowed structures which could not have been built before: cast iron, the Crystal Palace; wrought iron, the Eiffel Tower; drawn steel, the Golden Gate Bridge, all undeniably beautiful. Polymers lend themselves to bright colors, satisfying textures, and great freedom of form; they have opened new styles of design, of which some of the best examples are found in the household appliance sector: kitchen equipment, radio and CD-players, hair dryers, telephones, and vacuum cleaners make extensive and imaginative use of materials to allow styling, weight, feel, and form which give pleasure.

Those who concern themselves with this aesthetic dimension of engineering are known, rather confusingly, as "industrial designers". This chapter introduces some of the ideas of industrial design, emphasizing the role of materials. It ends with case-studies. But first a word of caution.

Systematic ways of choosing material and processes are discussed in Section 4 of this book. "Systematic" means that if *you* do it and *I* do it, following the same procedure, we will get the same result, and that the result, next year, will be the same as it is today. Industrial design is not, in this sense, systematic. Success, here, involves sensitivity to fashion, custom and educational background, and is influenced (manipulated, even) by advertising and association. The views of this chapter are partly those of writers who seem to me to say sensible things, and partly my own. You may not agree with them, but if they make you think about designing to give pleasure, the chapter has done what it should.

4.2.2 The requirements pyramid

The pen with which I am writing cost $5 (Figure 4.2.1, upper image). If you go to the right shop you can find a pen that costs well over $1000 (lower image). Does it

Fig. 4.2.1 Pens, inexpensive and expensive. The chosen material — acrylic in the upper picture, gold, silver, and enamel in the lower one — create the aesthetics and the associations of the pens. They are perceived differently, one pair as utilitarian, the other as something rare and crafted. (Lower figure courtesy of David Nishimura of Vintagepens.com.) The perceptions have much to do with the materials of which they are made.

Materials Selection in Mechanical Design; ISBN: 9780750661683
Copyright © 1999 Elsevier Ltd. All rights of reproduction, in any form, reserved.

write 200 times better than mine? Unlikely; mine writes perfectly well. Yet there is a market for such pens. Why?

A product has a *cost* — the outlay in manufacture and marketing it. It has a *price* — the sum at which it is offered to the consumer. And it has a *value* — a measure of what the consumer thinks it is worth. The expensive pens command the price they do because the consumer perceives their value to justify it. What determines value?

Functionality, provided by sound technical design, clearly plays a role. The requirements pyramid of Figure 4.2.2 has this as its base: the product must work properly, be safe and economical. Functionality alone is not enough: the product must be easy to understand and operate, and these are questions of *usability*, the second tier of the figure. The third, completing the pyramid, is the requirement that the product gives *satisfaction*: that it enhances the life of its owner.

The value of a product is a measure of the degree to which it meets (or exceeds) the expectation of the consumer in all three of these — functionality, usability, and satisfaction. Think of this as the character of the product. It is very like human character. An admirable character is one who functions well, interacts effectively, and is rewarding company. An unappealing character is one that does none of these. An odious character is one that does one or more of them in a way so unattractive that you cannot bear to be near him.

Products are the same. All the pens in Figure 4.2.1 function well and are easy to use. The huge difference in price implies that the lower pair provides a degree of satisfaction not offered by the upper ones. The most obvious difference between them is in the materials of which they are made — the upper pair of molded acrylic, the lower pair of gold, silver, and enamel. Acrylic is the material of tooth-brush handles, something you throw away after use. Gold and silver are the materials of precious jewelry, they have associations of craftsmanship, of heirlooms passed from one generation to the next. Well, that's part of the difference, but there is more. So — the obvious question — how do you create product character?

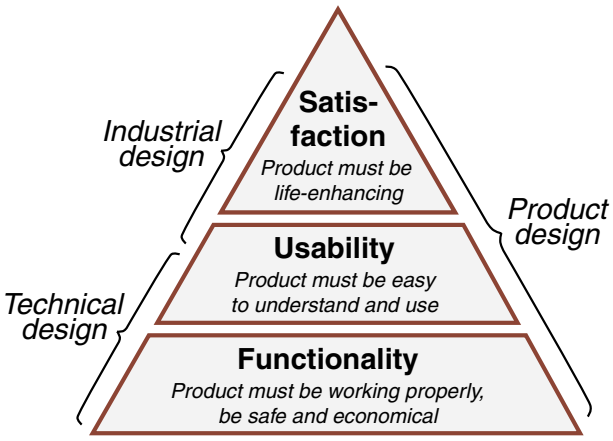


Fig. 4.2.2 The requirements pyramid. The lower part of the pyramid tends to be called "technical design", the upper part, "Industrial design" suggesting that they are separate activities. It is better to think of all three tiers as part of a single process that we shall call "product design".

4.2.3 Product character

Figure 4.2.3 shows a way of dissecting product character. It is a map of the ideas we are going to explore; like all maps there is a lot of detail, but we need it to find our way. In the center is information about the PRODUCT itself: the basic design requirements, its function, its features. The way these are thought through and developed is conditioned by the *context*, shown in the circle above it. The context is set by the answers to the questions: *Who? Where? When? Why?* Consider the first of these: *Who?* A designer seeking to create a product attractive to women will make choices that differ from those for a product intended for children, or for elderly people, or for sportsmen. *Where?* A product for use in the home requires a different choice of material and form than one to be used — say — in a school or hospital. *When?* One intended for occasional use is designed in a different way than one that is used all the time; one for formal occasions differs from one for informal use. *Why?* A product that is primarily utilitarian involves different design decisions than one that is largely a life-style statement. The context influences and conditions all the decisions that the designer takes in finding a solution. It sets the *mood*.[1]

On the left of PRODUCT lies information about the MATERIALS and the PROCESSES used to shape, join, and finish it. Each illustrates the library, so to speak, from which the choices can be made and — of course — they are the families, classes, and members that we have already met. Choosing these to provide functionality — the bottom of the pyramid — was the subject of this book. Material and process give the product its tangible form, its flesh, and bones so to speak; they create the *product physiology*.

On the right of Figure 4.2.3 are two further packages of information. The lower one — USABILITY — characterizes the ways in which the product communicates with the user: the interaction with their sensory, cognitive, and motor

[1] Many designers, working on a project, assemble a *mood-board* with images of the sort of people for whom the product is intended, the surroundings in which they suppose it will be used, and other products that the intended user group might own, seeking to capture the flavor of their life style.

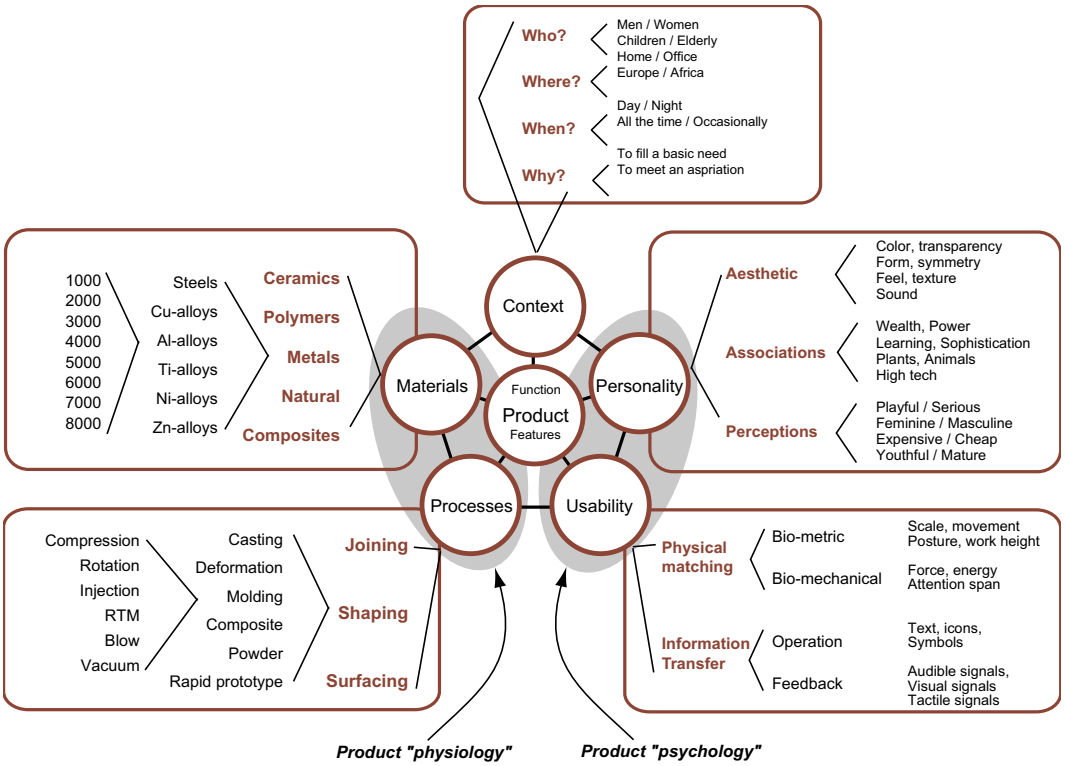


Fig. 4.2.3 The dissection of product character. Context defines the intentions or "mood"; materials and processes create the flesh and bones; the interface with the user determines usability, and the aesthetics, associations, and perceptions of the product create its personality. These terms are explained more fully in the text.

functions. Products success requires a mode of operation that, as far as possible, is intuitive, does not require taxing effort, and an interface that communicates the state of the product and its response to user action by visible, acoustic, or tactile response. It is remarkable how many products fail in this, and by doing so, exclude many of their potential users. There is, today, an awareness of this, giving rise to research on *inclusive design*: design to make products that can be used by a larger spectrum of the population.

One circle on Figure 4.2.3 remains: the one labeled PERSONALITY. Product personality derives from *aesthetics*, *associations*, and *perceptions*.

Anaesthetics dull the senses. Aesthetics do the opposite: they stimulate the five senses: sight, hearing, touch, taste, smell, and via them, the brain. The first row of the personality box elaborates: we are concerned here with color, form, texture, feel, smell, and sound — think of the smell of a new car and the sound of its door closing. They are no accident. Car makers spend millions to make them as they are.

A product also has associations — the second row of the box. Associations are the things the product reminds you of, the things it suggests. The Land Rover and other SUVs have forms and (often) colors that mimic those of military vehicles. The streamlining of American cars of the 1960s and 1970s carried associations of aerospace. It may be an accident that the VW Beetle has a form that suggests the insect, but the others are no accident; they were deliberately chosen by the designer to appeal to the consumer group (the *Who?*) at which the product was aimed.

Finally, the most abstract quality of all, perceptions. Perceptions are the reactions the product induces in an observer, the way it makes you *feel*. Here there is room for disagreement; the perceptions of a product change with time and depend on the culture and background of the observer. Yet in the final analysis it is the perception that causes the consumer, when choosing between a multitude of similar models, to prefer one above the others; it creates the "must have" feeling. Table 4.2.1 lists some perceptions with their opposites, in order to sharpen the meaning. They derive from product reviews and magazines specializing in product design; they are a part of a vocabulary, one that is used to communicate views about product character.

4.2.4 Using materials and processes to create product personality

Do materials, of themselves, have a personality? There is a school of thinking that holds as a central tenant that

Table 4.2.1 Some perceived attributes of products, with opposites

Perceptions (with opposites)	
Aggressive — Passive	Extravagant — Restrained
Cheap — Expensive	Feminine — Masculine
Classic — Trendy	Formal — Informal
Clinical — Friendly	Hand-made — Mass produced
Clever — Silly	Honest — Deceptive
Common — Exclusive	Humorous — Serious
Decorated — Plain	Informal — Formal
Delicate — Rugged	Irritating — Lovable
Disposable — Lasting	Lasting — Disposable
Dull — Sexy	Mature — Youthful
Elegant — Clumsy	Nostalgic — Futuristic

Table 4.2.2 Some aesthetic attributes of materials

Sense	Attribute	Sense	Attribute
Touch	Warm Cold Soft Hard Flexible Stiff	Hearing	Muffled Dull Sharp Resonant Ringing High pitched Low pitched
Sight	Optically Clear Transparent Translucent Opaque Reflective Glossy Matte Textured	Taste Smell	Bitter Sweet

materials must be used "honestly". By this they mean that deception and disguise are unacceptable — each material must be used in ways that expose its intrinsic qualities and natural appearance. It has its roots in the tradition of craftsmanship — the potters' use of clays and glazes, the carpenters' use of woods, the skills of silversmiths and glass makers in crafting beautiful objects that exploit the unique qualities of the materials with which they work, an integrity to craft and material.

This is a view to be respected. But it is not the only one. Design integrity is a quality that consumers value, but they also value other qualities: humor, sympathy, surprise, provocation, even shock. You do not have to look far to find a product that has one of these, and often it is achieved by using materials in ways that deceive. Polymers are frequently used in this way — their adaptability invites it. And, of course, it is partly a question of definition — if you say that a characterizing attribute of polymers is their ability to mimic other materials, then using them in this way *is* honest.

4.2.4.1 Materials and the senses: aesthetic attributes

Aesthetic attributes are those that relate to the senses: touch, sight, hearing, taste, and smell (Table 4.2.2). Almost everyone would agree that metals feel 'cold'; that cork feels 'warm'; that a wine glass, when struck, 'rings'; that a pewter mug sounds "dull", even "dead". A polystyrene water glass can look indistinguishable from one made of glass, but pick it up and it feels lighter, warmer, less rigid; tap it and it does not sound the same. The impression it leaves is so different from glass that, in an expensive restaurant, it would be completely unacceptable. Materials, then, have certain characterizing aesthetic attributes. Let us see if we can pin these down.

4.2.4.1.1 Touch: soft-hard/warm-cold

Steel is "hard"; so is glass; diamond is harder the either of them. Hard materials do not scratch easily; indeed they can be used to scratch other materials. They generally accept a high polish, resist wear, and are durable. The impression that a material is hard is directly related to its Vickers hardness H. Here is an example of a sensory attribute that relates directly to a technical one.

"Soft" sounds like the opposite of "hard" but it has more to do with modulus E than with hardness H. A soft material deflects when handled, it gives a little, it is squashy, but when it is released it returns to its original shape. Elastomers (rubbers) feels soft; so do polymer foams. Both have moduli that are 100 – 10,000 lower than ordinary 'hard' solids; it is this that makes them feel soft. Soft to hard is used as one axis of Figure 4.2.4. It uses the quantity $\sqrt{EH}$ as a measure.

A material feels "cold" to the touch if it conducts heat away from the finger quickly; it is "warm" if it does not. This has something to do with its thermal conductivity λ but there is more to it than that — it depends also on its specific heat C_p. A measure of the perceived coldness or warmth of a material is the quantity $\sqrt{\lambda C_p \rho}$. It is shown as the other axis of Figure 4.2.4, which nicely displays the tactile properties of materials. Polymer foams and low-density woods are warm and soft; so are balsa and cork. Ceramics and metals are cold and hard; so is glass. Polymers and composites lie in between.

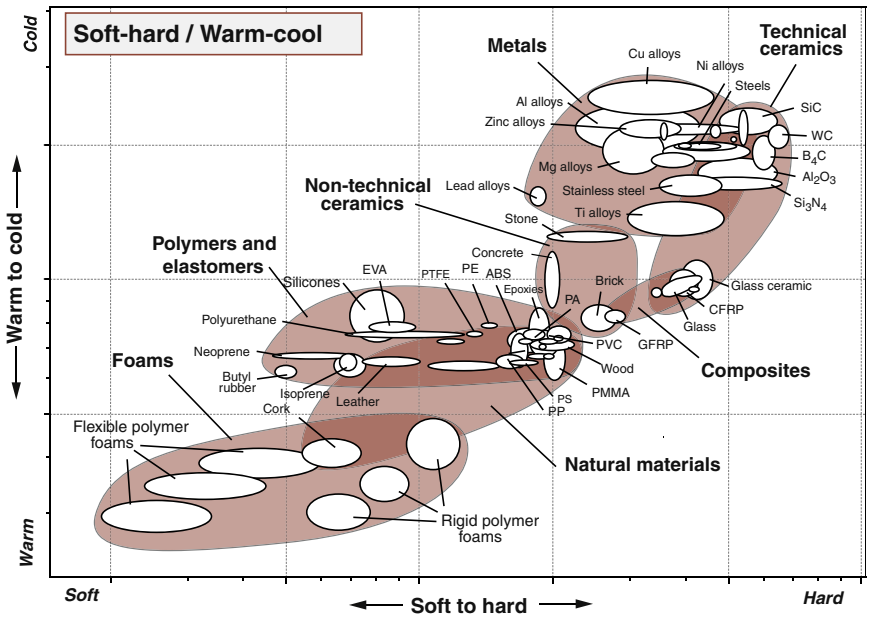


Fig. 4.2.4 Tactile qualities of materials. Foams and many natural materials are soft and warm; metals, ceramics, and glasses are hard and cold. Polymers lie in-between.

4.2.4.1.2 Sight: transparency, color, reflectivity

Metals are opaque. Most ceramics, because they are polycrystalline and the crystals scatter light, are either opaque or translucent. Glasses, and single crystals of some ceramics, are transparent. Polymers have the greatest diversity of optical transparency, ranging from transparency of optical quality to completely opaque.

Transparency is commonly described by a 4-level ranking that uses easily-understood everyday words: "opaque", "translucent", "transparent", and "water-clear". Figure 4.2.5 ranks the transparency of common materials. In order to spread the data in a useful way, it is plotted against cost. The cheapest materials offering optical-quality transparency ("water-clarity") are glass,

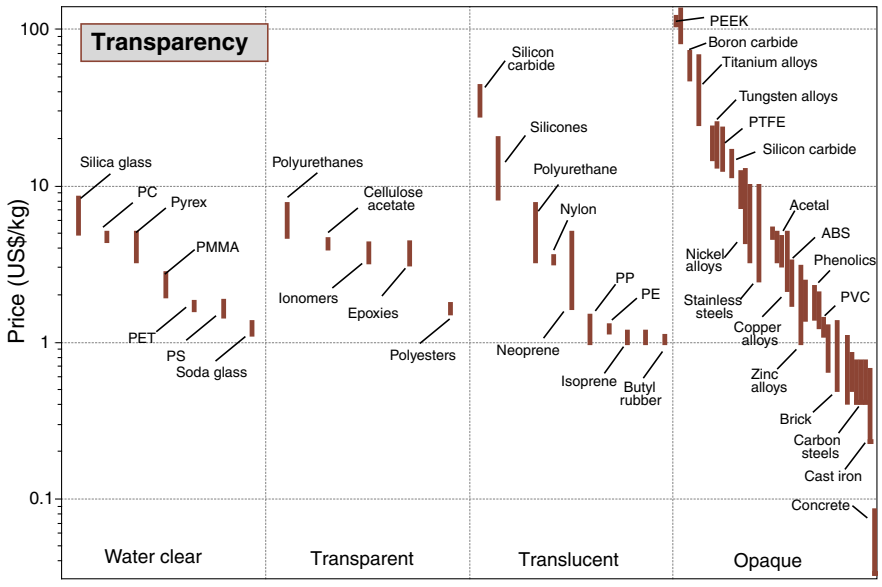


Fig. 4.2.5 Here transparency is ranked on a four-point scale, form water-clear to opaque. Water-clear materials are used for windows, display cases, and lenses. Transparent and translucent materials transmit light but diffuse it in doing so. Opaque materials absorb light.

PS, PET, and PMMA. Epoxies can be transparent but not with water clarity. Nylons are, at best, translucent. All metals, most ceramics and all carbon-filled or reinforced polymers are opaque.

Color can be quantified by analyzing spectra but this — from a design standpoint — does not help much. A more effective method is one of color matching, using color charts such as those provided by Pantone[2]; once a match is found it can be described by the code that each color carries. Finally there is reflectivity, an attribute that depends partly on material and partly on the state of its surface. Like transparency, it is commonly described by a ranking: dead matte, eggshell, semi-gloss, gloss, mirror.

4.2.4.1.3 Hearing: pitch and brightness

The frequency of sound (pitch) emitted when an object is struck relates to its material properties. A measure of this pitch, $\sqrt{E/\rho}$, is used as one axis of Figure 4.2.6. Frequency is not the only aspect of acoustic response — the other has to do with the damping or loss coefficient, η. A highly damped material sounds dull and muffled; one with low damping rings. Acoustic brightness — the inverse of damping — is used as the other axis of Figure 4.2.6. It groups materials that have similar acoustic behavior.

Bronze, glass, and steel ring when struck, and the sound they emit has — on a relative scale — a high pitch; they are used to make bells; alumina, on this ranking, has bell-like qualities. Rubber, foams, and many polymers sound dull, and, relative to metals, they vibrate at low frequencies; they are used for sound damping. Lead, too, is dull and low-pitched; it is used to clad buildings for sound insulation.

The three figures show that each material class has a certain recognizable aesthetic character. Ceramics are hard, cold, high-pitched, and acoustically bright. Metals, too, are relatively hard and cold but although some, like bronze, ring when struck, others — like lead — are dull. Polymers and foams are most nearly like natural materials — warm, soft, low-pitched, and muffled, though some have outstanding optical clarity and almost all can be colored. But their low hardness means that they scratch easily, loosing their gloss.

These qualities of a material contribute to the product personality. The product acquires some of the attributes of the material from which it is made, an effect that designers recognize and use when seeking to create a personality. A stainless steel facia, whether it be in a car or on a hi-fi system, has a different personality than one of polished wood or one of leather, and that in part is because the product has acquired some of the aesthetic qualities of the material.

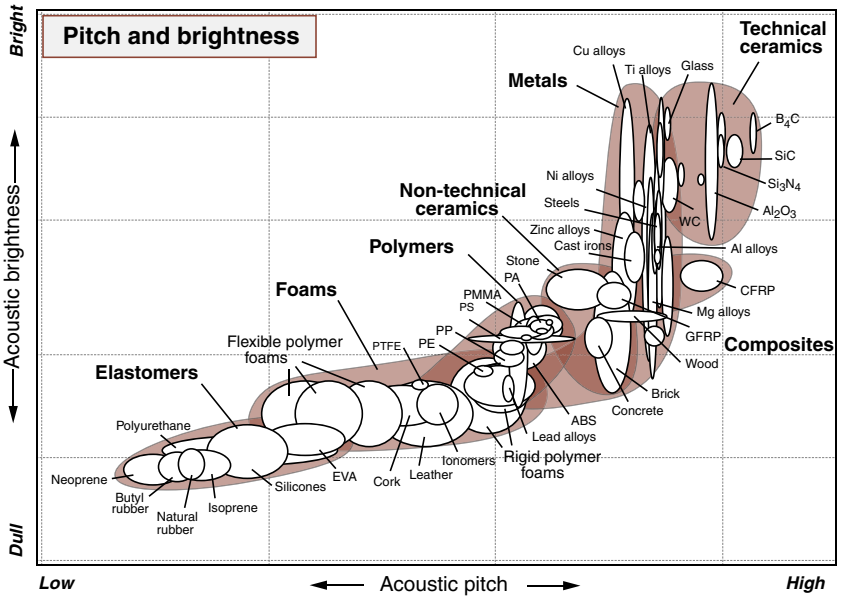


Fig. 4.2.6 Acoustic properties of materials. The "ring" of a wine glass is because glass in an acoustically bright material with a high natural pitch; the dull "ping" of a plastic glass is because polymers are much less bright and — in the same shape — vibrate at a lower frequency. Materials at the top right make good bells; those at the bottom left are good for damping sound.

[2] Pantone (www.pantone.com.) provide detailed advice on color selection, including color-matching chart and good descriptions of the associations and perceptions of color.

4.2.4.2 Materials and the mind: associations and perceptions

So a material certainly has aesthetic qualities — but can it be said to have a personality? At first sight, no — it only acquires one when used in a product. Like an actor, it can assume many different personalities, depending on the role it is asked to play. Wood in fine furniture suggests craftsmanship, but in a packing case, cheap utility. Glass in the lens of a camera has associations of precision engineering, but in a beer bottle, that of disposable packaging. Even gold, so often associated with wealth and power, has different associations when used in microcircuits: that of technical functionality.

But wait. The object in Figure 4.2.7 has its own sombre association. It appears to be made of polished hardwood — the traditional material for such things. If you had to choose one, you would probably not have any particular aversion to this one — it is a more or less typical example. But suppose I told you it was made of *plastic* — would you feel the same? Suddenly it becomes like a bin, a wastebasket, inappropriate for its dignified purpose. Materials, it seems, *do* have personality.

4.2.4.2.1 Expression through material

Think of wood. It is a natural material with a grain that has a surface texture, pattern, color, and feel that other materials do not have. It is tactile — it is perceived as warmer than many other materials, and seemingly softer. It is associated with characteristic sounds and smells. It has a tradition; it carries associations of craftsmanship. No two pieces are exactly alike; the wood-worker selects the piece on which he will work for its grain and texture. Wood enhances value: the interior of cheap cars is plastic, that of expensive ones is burr-walnut and calves leather. And it ages well, acquiring additional character with time; objects made of wood are valued more highly when they are old than when they are new. There is more to this than just aesthetics; there are the makings of a personality, to be brought out by the designer, certainly, but there none the less.

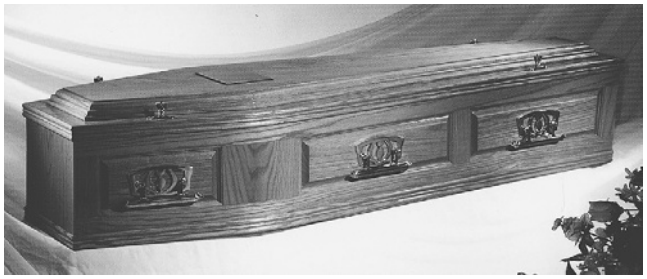


Fig. 4.2.7 A coffin. Wood is perceived to be appropriate for its sombre, ceremonial function, plastic inappropriate.

And metals…Metals are cold, clean, precise. They ring when struck. They reflect light — particularly when polished. They are accepted and trusted: machined metal looks strong, its very nature suggests it has been *engineered*. They are associated with robustness, reliability, permanence. The strength of metals allows slender structures — the cathedral-like space of railway stations or the span of bridges. They can be worked into flowing forms like intricate lace or cast into solid shapes with elaborate, complex detail. The history of man and of metals is intertwined — the titles "bronze age" and "iron age" tell you how important these metals were — and their qualities are so sharply defined that they have become ways of describing human qualities — an iron will, a silvery voice, a golden touch, a leaden look. And, like wood, metals can age well, acquiring a patina that makes them more attractive than when newly polished — think of the bronze of sculptures, the pewter of mugs, the lead of roofs.

Ceramics and glass? They have an exceptionally long tradition — think of Greek pottery and Roman glass. They accept almost any color; this and their total resistance to scratching, abrasion, discoloration, and corrosion gives them a certain immortality, threatened only by their brittleness. They are — or were — the materials of great craft-based industries: the glass of Venice, the porcelain of Meissen, the pottery of Wedgwood, valued at certain times more highly than silver. But at the same time they can be robust and functional — think of beer bottles. The transparency of glass gives it an ephemeral quality — sometimes you see it, sometimes you do not. It interacts with light, transmitting it, refracting it, reflecting it. And ceramics today have additional associations — those of advanced technology: kitchen stovetops, high pressure/high temperature valves, space shuttle tiles…materials for extreme conditions.

And finally polymers. "A cheap, plastic imitation", it used to be said — and that is a hard reputation to live down. It derives from the early use of plastics to simulate the color and gloss of Japanese handmade pottery, much valued in Europe. Commodity polymers *are* cheap. They are easily colored and molded (that is why they are called "plastic"), making imitation easy. Unlike ceramics, their gloss is easily scratched, and their colors fade — they do not age gracefully. You can see where the reputation came from. But is it justified? No other class of material can take on as many characters as polymers: colored, they look like ceramics; printed, they can look like wood or textile; metalized, they look exactly like metal. They can be as transparent as glass or as opaque as lead, as flexible as rubber or as stiff — when reinforced — as aluminum. Plastics emulate precious stones in jewelry, glass in drinking glasses and glazing, wood in counter tops, velvet and fur in clothing, even grass. But despite this chameleon-like

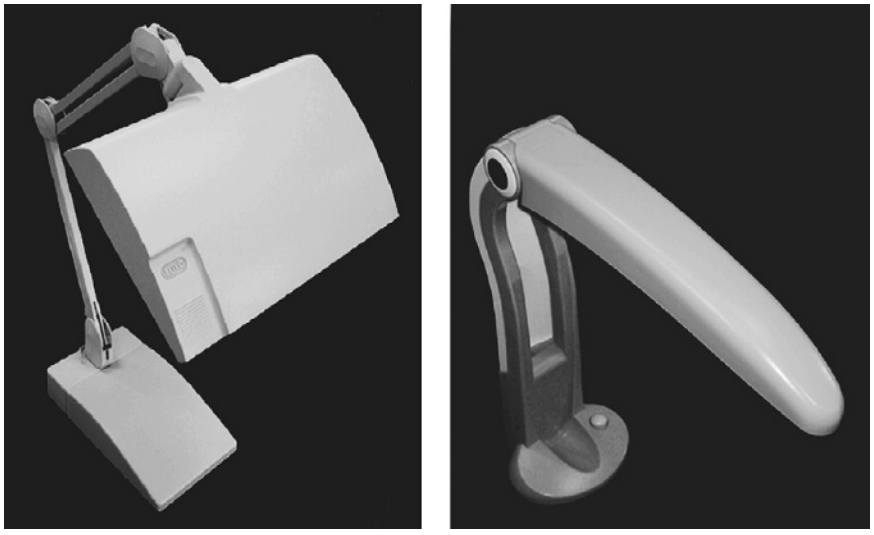


Fig. 4.2.8 Lamps. Both have the same technical rating, but differ completely in their personalities. Materials, processes, form, weight, and color have all contributed to the personality.

behavior they *do* have a certain personality: they feel warm — much warmer than metal or glass. They are adaptable — that is part of their special character; and they lend themselves, particularly, to brightly colored, light-hearted, even humorous, design. But their very cheapness creates problems as well as benefits: our streets, county-side and rivers are littered with discarded plastic bags and packaging that decay only very slowly.

The ways in which material, processes, usability and personality combine to create a product character tuned to the context or "mood" are best illustrated by examples. Figure 4.2.8 shows the first. The lamp on the left is designed for the office. It is angular, functional, creamy-gray, and it is heavy. Its form and color echo those of computer consoles and keyboards, creating associations of contemporary office technology. Its form and weight transmit the ideas of stability, robustness, efficiency, and fitness for task — but for tasks in the workplace, not in the bedroom. Materials and processes have been chosen to reinforce these associations and perceptions. The enameled frame is pressed and folded sheet steel, the base-weight is cast iron, the reflector is stainless steel set in a high-impact ABS enclosure.

The lamp on the right of Figure 4.2.8 has the same technical rating of that on the left; the same functionality and usability. But there the resemblance ends. This product is not designed for the busy executive but for children (and adults that still enjoy being children), to be used in the playroom or bedroom. It has a contoured form, contrasting translucent colors, and it is very light. It is made of colored acrylic in translucent and opaque grades, so that the outside of the lamp glows like a neon sign when it is lit. Its form is partly derived from nature, partly from cartoons and comic strips, giving it a light-hearted character. I perceive it as playful, funny, cheerful, and clever — but also as eccentric and easily damaged. You may perceive it in other ways — perception is a personal thing; it depends where you are coming from. Skilled designers manipulate perception to appeal to the user-group they wish to attract.

Figure 4.2.9 shows a second example. Here are two contrasting ways of presenting electronic home entertainment systems. On the left: a music center aimed at successful professionals with disposable income, comfortable with (or addicted to) advanced technology, for whom only the best is good enough. The linear form, the use of primitives (rectangles, circles, cylinders, cones) and the matt silver and black proclaim that this product has not just been *made*, it has been *Designed* (big *D*). The formal geometry and finish suggest precision instruments, telescopes, electron microscopes and the shapes resemble those of organ pipes (hence associations of music, of culture). The perception is that of quality cutting-edge technology, a symbol of discriminating taste. The form has much to do with this associations and perceptions, but so too do the materials: brushed aluminum, stainless steel and black enamel — these are not materials you choose for a cuddly toy.

On the right: electronics presented in another way. This is a company that has retained market share, even increased it, by *not* changing, at least as far as appearance is concerned (I had one 40 years ago that looked exactly like this). The context? Clearly, the home, perhaps aimed at consumers who are uncomfortable with modern technology (though the electronics in these radios is modern enough), or who simply feel that it clashes with the home environment. Each radio has a simple form, it is

Fig. 4.2.9 Consumer electronics. The products on the left is aimed at a different consumer group than those on the right. The personalities of each (meaning the combination of aesthetics, associations, and perceptions) have been constructed to appeal to the target group. Materials play a central role here in creating personality. (Figure on left courtesy of Bang & Olufsen.)

pastel-colored, it is soft and warm to touch. It is the materials that make the difference: these products are available in suede or leather in six or more colors. The combination of form and material create associations of comfortable furniture, leather purses, and handbags (hence, luxury, comfort, style), the past (hence, stability) and perceptions of solid craftsmanship, reliability, retro-appeal, traditional but durable design.

So there is a character hidden in a material even before it has been made into an recognizable form — a sort of embedded personality, a shy one, not always obvious, easily concealed or disguised, but one that, when appropriately manipulated, imparts its qualities to the design. It is for this reason that certain materials are so closely linked to certain *design styles*. A *style* is a short-hand for a manner of design with a shared set of aesthetics, associations and perceptions. The Early Industrial style (1800–1890)[3] embraced the technologies of the industrial revolution, using cast iron, and steel, often elaborately decorated to give it a historical façade. The Arts and Crafts movement (1860–1910) rejected this, choosing instead natural materials and fabrics to create products with the character of traditional hand-crafted quality. Art Nouveau (1890–1918), by contrast, exploited the fluid shapes and durability made possible by wrought iron and cast bronze, the warmth and textures of hard wood and the transparency of glass to create products of flowing, organic character. The Art Deco movement (1918–1935) extended the range of materials to include for the first time plastics (Bakelite and Catalin) allowing production both of luxury products for the rich and also mass-produced products for a wider market. The simplicity and explicit character of Bauhaus designs (1919–1933) is most clearly expressed by the use of chromed steel tubing, glass and moulded plywood. Plastics first reach maturity in product design in the cheeky iconoclastic character of the Pop-Art style (1940–1960). Since then the range of materials has continued to increase, but their role in helping to mould product character remains.

4.2.5 Summary and conclusions

What do we learn? The element of satisfaction is central to contemporary product design. It is achieved through an integration of good technical design to provide functionality, proper consideration of the needs of the user in the design of the interface, and imaginative industrial design to create a product that will appeal to the consumers at whom it is aimed.

[3] The dates are, of course, approximate. Design styles do not switch on an off on specific dates, they emerge as a development of, or reaction to, earlier styles with which they often co-exist, and they merge into the styles that follow.

Materials play a central role in this. Functionality is dependent on the choice of proper material and process to meet the technical requirements of the design safely and economically. Usability depends on the visual and tactile properties of materials to convey information and respond to user actions. Above all, the aesthetics, associations, and perceptions of the product are strongly influenced by the choice of the material and its processing, imbuing the product with a personality that, to a greater or lesser extent, reflects that of the material itself.

Consumers look for more than functionality in the products they purchase. In the sophisticated market places of developed nations, the "consumer durable" is a thing of the past. The challenge for the designer no longer lies in meeting the functional requirements alone, but in doing so in a way that also satisfies the aesthetic and emotional needs. The product must carry the image and convey the meaning that the consumer seeks: timeless elegance, perhaps; or racy newness. One Japanese manufacturer goes so far as to say: "*Desire* replaces *need* as the engine of design".

Not everyone, perhaps, would wish to accept that. So we end with simpler words — the same ones with which we started. Good design works. Excellent design also gives pleasure. The imaginative use of materials provides it.

If you found this chapter interesting and would like to read more, you will find the ideas it contains developed more fully in the first book listed in the 'Further reading' section.

Further reading

Ashby, M.F. and Johnson, K. (2002) *Materials and Design — The Art and Science of Materials Selection in Product Design*, Butterworth-Heinemann, Oxford, UK. ISBN0-7506-5554-2. (*A book that develops further the ideas outlined in this chapter.*)

Clark, P. and Freeman, J. (2000) *Design, a Crash Course*, The Ivy Press Ltd, Watson-Guptil Publications, BPI Communications Inc. New York, NY, USA. ISBN 0-8230-0983-1. (*An entertainingly-written scoot through the history of product design from 5000BC to the present day.*)

Dormer, P. (1993) *Design since 1945*, Thames and Hudson, London UK. ISBN 0-500-20269-9. (*A well-illustrated and inexpensive paperback documenting the influence of industrial design in furniture, appliances and textiles — a history of contemporary design that complements the wider-ranging history of Haufe (1998), q.v.*)

Forty, A. (1986) *Objects of Desire — Design in Society Since 1750*, Thames and Hudson, London, UK. ISBN 0-500-27412-6. (*A refreshing survey of the design history of printed fabrics, domestic products, office equipment and transport system. The book is mercifully free of eulogies about designers, and focuses on what industrial design does, rather than who did it. The black and white illustrations are disappointing, mostly drawn from the late 19th or early 20th centuries, with few examples of contemporary design.*)

Haufe, T. (1998) *Design, a Concise History*, Laurence King Publishing, London, UK (originally in German). ISBN 1-5669-134-9. (*An inexpensive soft-cover publication. Probably the best introduction to industrial design for students (and anyone else). Concise, comprehensive, clear and with intelligible layout and good, if small, color illustrations.*)

Jordan, P.S. (2000) *Designing Pleasurable Products*, Taylor and Francis, London, UK. ISBN 0-748-40844-4. (*Jordan, Manager of Aesthetic Research and Philips Design, argues that products today must function properly, must be usable, and must also give pleasure. Much of the book is a description of market-research methods for eliciting reactions to products from users.*)

Julier, G. (1993) *Encyclopedia of 20th Century Design and Designers*, Thames & Hudson, London, UK. ISBN 0-500-20261-3. (*A brief summary of design history with good pictures and discussions of the evolution of product designs.*)

Manzini, E. (1989) *The Material of Invention*, The Design Council, London UK. ISBN 0-85072-247-0 (*Intriguing descriptions of the role of material in design and in inventions. The translation from Italian to English provides interesting — and often inspiring — commentary and vocabulary that is rarely used in traditional writings about materials.*)

McDermott, C. (1999) *The Product Book*, D & AD in association with Rotovison, UK. (*50 essays by respected designers who describe their definition of design, the role of their respective companies and their approach to product design.*)

Norman, D.A. (1988) The Design of Everyday Things, Doubleday, New York, USA. ISBN 0-385-26774-6. (*A book that provides insight into the design of products with particular emphasis on ergonomics and ease of use.*)

Chapter 4.3

4.3

Forces for change

Michael Ashby

4.3.1 Introduction and synopsis

If there were no forces for change, everything would stay the same. Things are changing faster now than ever before. The evolving circumstances of the world in which we live change the boundary conditions for design, and with them those for selecting materials and processes.

These changes are driven by a number of forces. First, there is the *market-pull*: the demand from industry for materials that are lighter, stiffer, stronger, tougher, cheaper, and more tolerant of extremes of temperature and environment, and that offer greater functionality. Then there is the *science-push*: the curiosity-driven researches of materials experts in the laboratories of universities, industries, and government. There is the driving force of what might be called *mega-projects*: historically, the Manhattan Project, the development of nuclear power, the space-race and various defense programs; today, one might think of alternative energy technologies, the problems of maintaining an aging infrastructure of drainage, roads, bridges and aircraft, and the threat of terrorism. There is the trend to miniaturization while at the same time increasing the functionality of products. There is legislation regulating product safety and there is the increased emphasis on liability established by recent legal precedent.

This chapter examines forces for change and the directions in which they push materials and their deployment. Figure 4.3.1 sets the scene.

4.3.2 Market-pull and science-push (extreme right and left of Figure 4.3.1)

4.3.2.1 The power of the market

The end-users of materials are the manufacturing industries. They decide which material they will purchase, and adapt their designs to make best use of them. Their decisions are based on the nature of their products. Materials for large civil structures (which might weigh 10,000 tonnes or more) must be cheap; economy is the overriding consideration. By contrast, the cost of the materials for high-tech products (sports equipment, military hardware, space projects, biomedical applications) plays a less important role: for an artificial heart valve, for instance, material cost is almost irrelevant. Performance, not economy, dictates the choice.

The market price of a product has several contributions. One is the cost of the materials of which the product is made, but there is also the cost of the research and development that went into its design, the cost of manufacture and marketing and the perceived value associated with fashion, scarcity, lack of competition and such like. When the material costs are a large part of the market value (50%, say) — that is, when the value added to the material is small — the manufacturer seeks to economize on material usage to increase profit or market share. When, by contrast, material costs are a tiny fraction of the market value (1%, say), the manufacturer

Materials Selection in Mechanical Design; ISBN: 9780750661683
Copyright © 1999 Elsevier Ltd. All rights of reproduction, in any form, reserved.

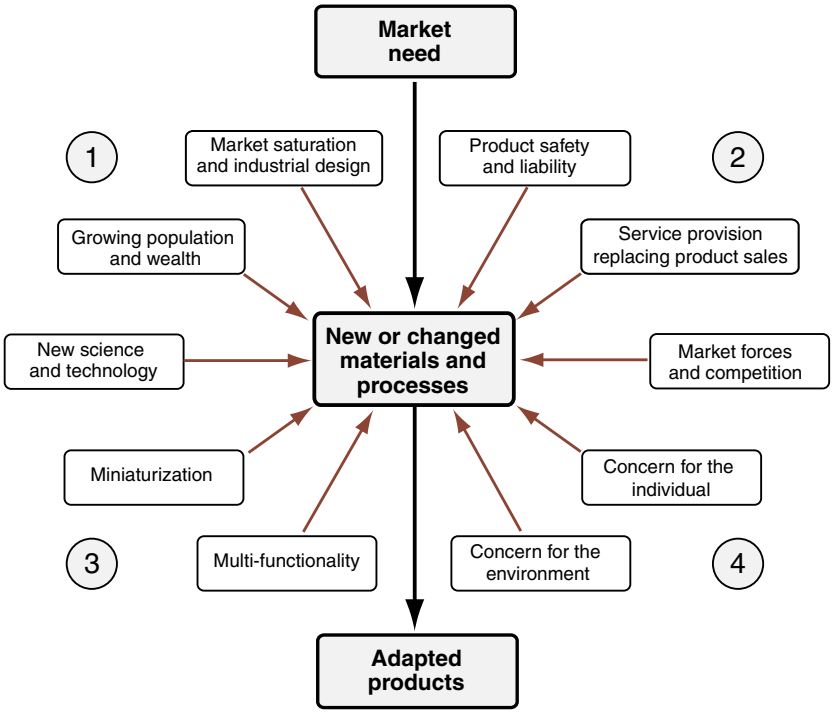


Fig. 4.3.1 Forces for change. Each of the influences exerts pressure to change the choice of material and process, and stimulates efforts to develop new ones.

seeks the materials that will most improve the performance of the produce with little concern for their cost.

With this background, examine Figures 4.2.2 and 4.2.3. The vertical axis is the price per unit weight ($/kg) of materials and of products: it gives a common measure by which materials and products can be compared. The measure is a crude one but has the great merit that it is unambiguous, easily determined, and bears some relationship to value-added. A product with a price/kg that is only two or three times that of the materials of which it is made is material-intensive and is sensitive to material costs; one with a price/kg that is 100 times that of its materials is insensitive to material costs, and is probably performance-driven rather than cost-driven. On this scale the price per kg of a contact lens differs from that of a glass bottle by a factor of 10^5, even though both are made of almost the same glass; the cost per kg of a heart valve differs from that of a plastic bottle by a similar factor, even though both are made of polyethylene. There is obviously something to be learned here.

Look first at the price per unit weight of materials (Figure 4.3.2). The bulk "commodity" materials of construction and manufacture lie in the shaded band; they all cost between $0.05 and $20/kg. Construction materials like concrete, brick, timber, and structural steel lie at the lower end; high-tech materials like titanium alloys lie at the upper. Polymers span a similar range: polyethylene at the bottom, polytetrafluoroethylene (PTFE) near the top. Composites lie higher, with GFRP at the bottom and CFRP at the top of the range Engineering ceramics, at present, lie higher still, though this will change as production increases. Only the low-volume "exotic" materials lie much above the shaded band.

The price per kg of products (Figure 4.3.3) shows a different distribution. Eight market sectors are shown, covering much of the manufacturing industry. The shaded band on this figure spans the cost of commodity materials, exactly as on the previous figure. Sectors and their products within the shaded band have the characteristic that material cost is a major fraction of product price: about 50% in civil construction, large marine structures and some consumer packaging, falling to perhaps 20% as the top of the band is approached (family car — around 25%). The value added in converting material to product in these sectors is relatively low, but the market volume is large. These constraints condition the choice of materials: they must meet modest performance requirements at the lowest possible cost. The associated market sectors generate a driving force for improved processing of conventional materials in order to reduce cost without loss of performance, or to increase reliability at no increase in cost. For these sectors, incremental improvements in well-tried materials are far more important than revolutionary research-findings. Slight improvements in steels, in precision

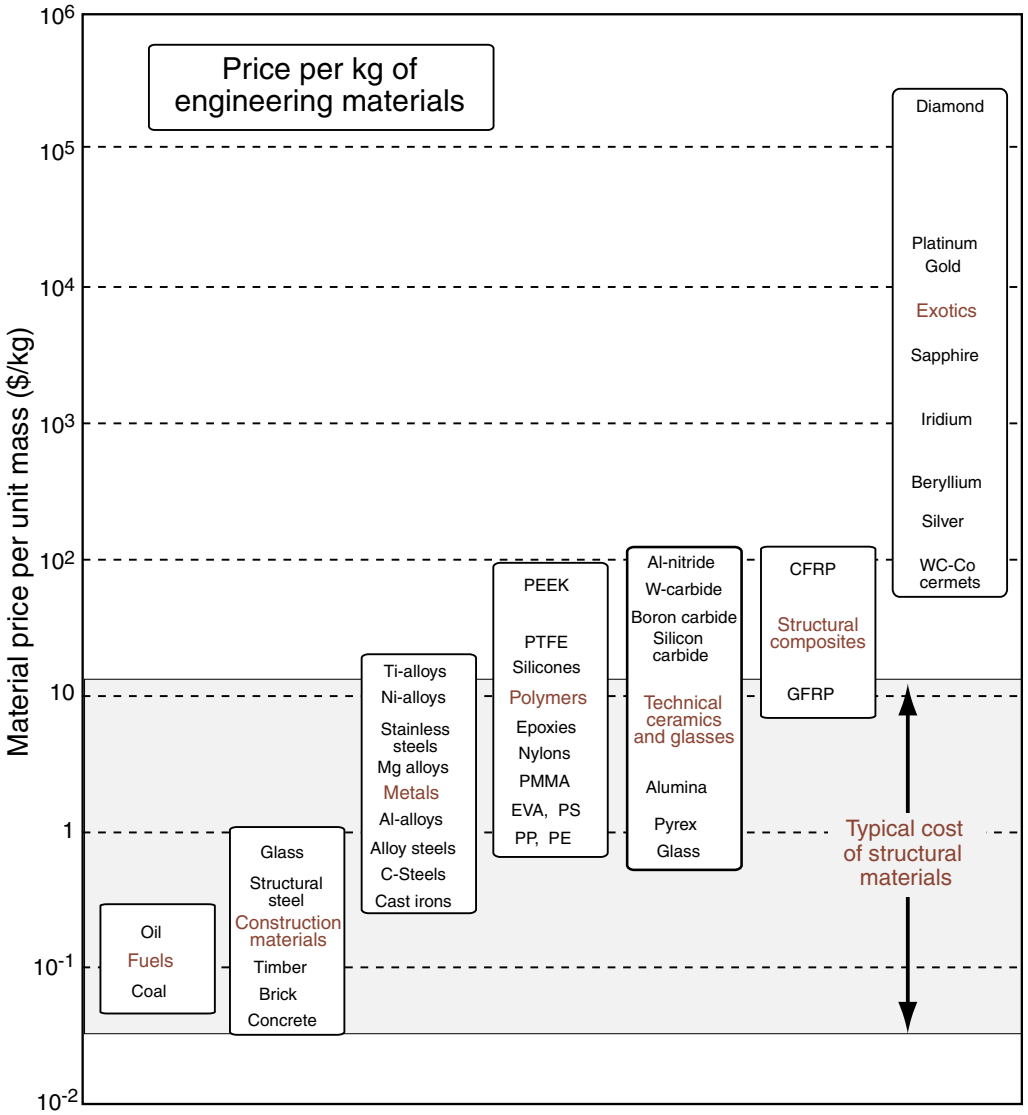


Fig. 4.3.2 The price per unit weight diagram for materials. The shaded band spans the range in which lies the most widely used commodity material of manufacture and construction.

manufacturing methods, or in lubrication technology are quickly assimilated and used.

The products in the upper half of the diagram are technically more sophisticated. The materials of which they are made account for less than 10% — sometimes less than 1% — of the price of the product. The value added to the material during manufacture is high. Product competitiveness is closely linked to material performance. Designers in these sectors have greater freedom in their choice of material and there is a readier acceptance of new materials with attractive property-profiles. The market-pull here is for performance, with cost as a secondary consideration. These smaller volume, higher value-added sectors drive the development of new or improved materials with enhanced performance: materials that are lighter, or stiffer, or stronger, or tougher, or expand less, or conduct better — or all of these at once.

The sectors have been ordered to form an ascending sequence, prompting the question: what does the horizontal axis measure? Many factors are involved here, one of which can be identified as "information content". The accumulated knowledge involved in the production of a contact lens or a heart valve is clearly greater than that in a beer-glass or a plastic bottle. The sectors on the left make few demands on the materials they employ; those on the right push materials to their limits, and at the same time demand the highest reliability. These features make them information-intensive. But there are also other factors: market size, competition (or lack of it), perceived value, fashion and taste, and so on. For this reason the diagram should not be over-interpreted:

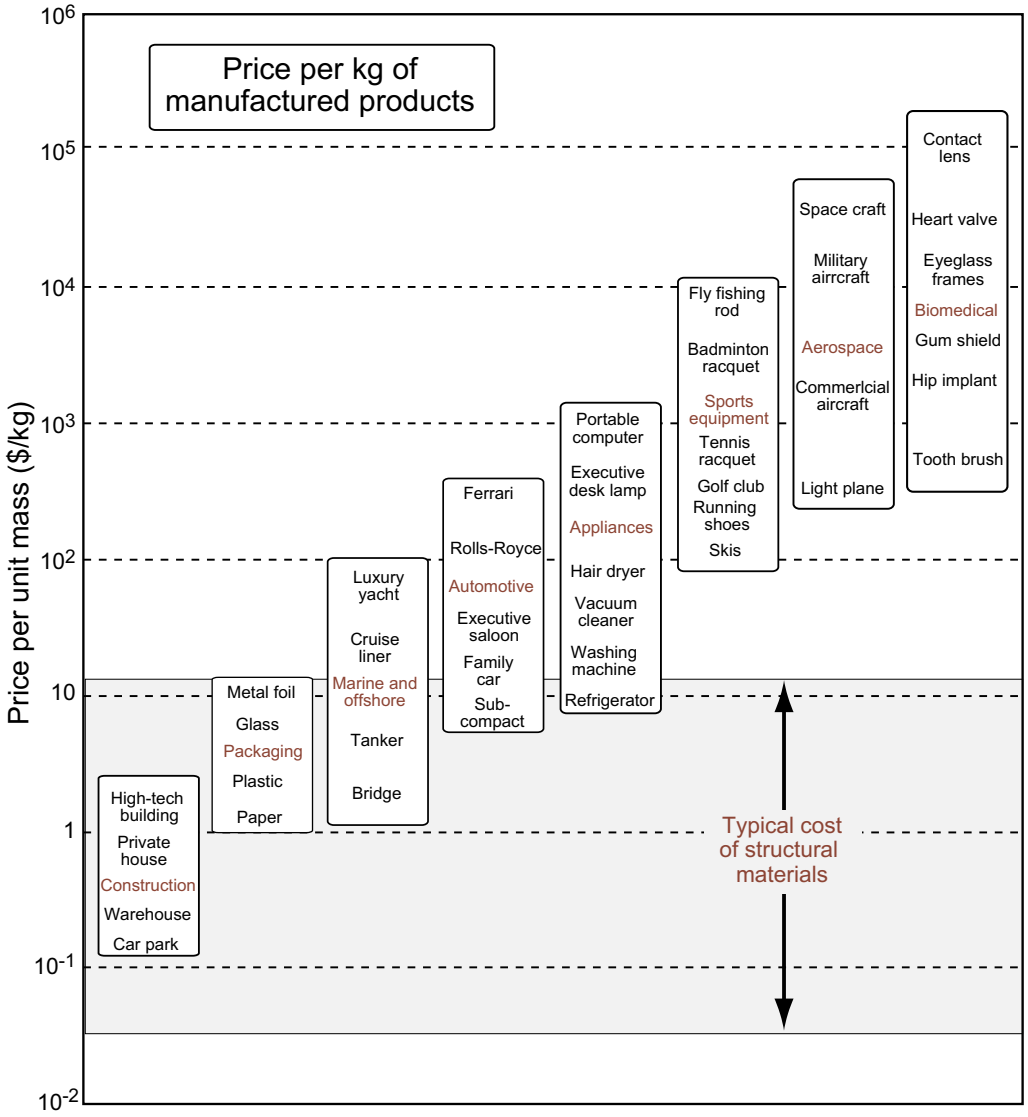


Fig. 4.3.3 The price per unit weight diagram for products. The shaded band spans the range in which lies most of the material of which they are made. Products in the shaded band are material-intensive; those above it are not.

it is a help in structuring information, but it is not a quantitative tool.

The manufacturing industry, even in times of recession, has substantial resources; and it is in the interests of governments to support their needs. The market pull is, ultimately, the strongest force for change.

4.3.2.2 New science: curiosity-driven research

Compulsive curiosity, in children, may be a tribulation, but it is the oxygen of innovative engineering. Technically advanced countries sustain the flow of new ideas by supporting research in three kinds of organization: universities, government laboratories, and industrial research laboratories. Some of the scientists and engineers working in these institutions are encouraged to pursue ideas that may have no immediate economic objective, but that can evolve into the materials and manufacturing methods of future decades. Numerous now-commercial materials started in this way. Aluminum, in the time of Napoleon III, was a scientific wonder — he commissioned a set of aluminum spoons for which he paid more than those of solid silver. Aluminum was not, at that time, a commercial success; now it is. Titanium, more recently, has had a similar history. Amorphous (=non-crystalline) metals, now important in transformer technology and in recording-heads of tape decks, were, for years, of only academic interest. It seems improbable that superconductors or semiconductors would have been discovered in response to

market forces alone; it took long-term curiosity-driven research to carry them to the point that they became commercially attractive. Polyethylene was discovered by chemists studying the effect of pressure on chemical reactions, not by the sales or marketing departments of multi-national corporations. History is dotted with examples of materials and processes that have developed from the inquisitiveness of individuals.

What new ideas are churning in the minds of the materials scientists of today? There are many, some already on the verge of commercialization, others for which the potential is not yet clear. Some, at least, will provide opportunities for innovation; the best may create new markets.

Monolithic ceramics, now produced in commercial quantities, offer high hardness, chemical stability, wear resistance, and resistance to extreme temperatures. Their use as substrates for microcircuits is established; their use in wear-resistant applications is growing, and their use in heat engines is being explored. The emphasis in the development of *composite materials* is shifting toward those that can support loads at higher temperatures. Metal–matrix composites (example: the aluminum containing particles or fibers of silicon–carbide and intermetallic–matrix composites (titanium–aluminide or molybdenum–disilicide containing silicon–carbide, for instance) can do this. So, potentially, can ceramic-matrix composites (alumina with silicon carbide fibers) though the extreme brittleness of these materials requires new design techniques. Metallic foams, up to 90% less dense than the parent metal, promise light, stiff sandwich structures competing with composites.

New *bio-materials*, designed to be implanted in the human body, have structures onto which growing tissue will bond without rejection. New *polymers* that can be used at temperatures up to 350°C allow plastics to replace metals in even more applications — the inlet manifold of the automobile engine is an example. New *elastomers* are flexible but strong and tough; they allow better seals, elastic hinges, and resilient coatings. Techniques for producing *functionally-graded materials* can give tailored gradients of composition and structure through a component so that it could be corrosion resistant on the outer surface, tough in the middle and hard on the inner surface. *Nano-structured materials* promise unique mechanical and electrical properties. *"Intelligent" materials* that can sense and report their condition (via embedded sensors) allow safety margins to be reduced.

Developments in *rapid prototyping* now permit single, complex, parts to be made quickly without dies or moulds, from a wide range of materials. *Micron-scale fabrication methods* create miniature electro-mechanical systems (MEMS). New techniques of *surface engineering* allow the alloying, coating or heat treating of a thin surface layer of a component, modifying its properties to enhance its performance. They include laser hardening, coatings of well-adhering polymers and ceramics, ion implantation, and even the deposition of ultra-hard carbon films with a structure and properties like those of diamond. New *adhesives* displace rivets and spot-welds; the glue-bonded automobile is a real possibility. And new techniques of *mathematical modeling* and *process control* allow much tighter control of composition and structure in manufacture, reducing cost and increasing reliability and safety.

All these and many more are now realities. They have the potential to enable new designs and to stimulate redesign of product that already have a market, increasing their market share. The designer must stay alert.

4.3.3 Growing population and wealth, and market saturation (sector 1, Figure 4.3.1)

The world's population is growing in number. Much of this population is growing also in wealth. Whether or not it is wise, they consume more products, and as wealth increases they want still more. But the growth of the world's product-producing capacity has grown faster than either of these, with the result that, in developed and developing countries, product markets are saturated. If you want a product — a mobile phone, a refrigerator, a car — you do not have to stand in line with the hope of getting the only one in the store (as was the case within living memory throughout Europe). Instead, a sales person descends on you to guide you through the task of selecting, from an array of near-identical products with near-identical prices, the one you think you want.

This has certain consequences. One is the massive and continuing growth in consumption of energy and material resources. This, as a force for change in material selection, was the subject of Chapter 4.1. Another arises because, in a saturated market, the designer must seek new ways to attract the consumer. The more traditional reliance on engineering qualities to sell a product are replaced (or augmented) by visual qualities, associations, and carefully managed perceptions created by industrial design. This, too, influences material and process choice, and it, too, was the subject of Chapter 4.2.

4.3.4 Product liability and service provision (sector 2, Figure 4.3.1)

Legislation now requires that, if a product has a fault, it must be recalled and the fault rectified. Product recalls are exceedingly expensive and they damage the company's image. The higher up the chart of Figure 4.3.3

we go, the more catastrophic a fault becomes. When you buy a packet of six ball-point pens, it is no great tragedy if one does not write properly. The probability ratio 1:6, in a Gaussian distribution, is one standard deviation or "1-sigma". But would you fly in a plane designed on a "1-sigma" basis? Safety critical systems today are designed on a "6-sigma" reliability rating, meaning that the probability that the component or assembly fails to meet specification is less than one part in 10^9.

This impacts the way materials are produced and processed. Reliability and reproducibility requires sophisticated process control and that every aspect of their production is monitored and verified. Clean steels (steels with greatly reduced inclusion-account), aluminum alloys with tight composition control, real-time feedback control of processing, new non-destructive testing for quality and integrity, and random sample testing all help ensure that quality is maintained.

This pressure on manufacturers to assume full-life responsibility for their products causes them to consider taking maintenance and replacement out of the hands of the consumer and doing it themselves. In this way they can monitor product use, and replace it, taking back the original, not when it is worn out but when it is optimal for them to recondition and re-deploy it. The producer no longer sells the product (it remains his); he sells a service. In common with many others, my university does not own copying machines, though we have many. Instead it has a contract with a provider to guarantee a certain copying-capacity — the provision of "pages per week", you might call it.

This changes the economic boundary conditions for the designer of the copier. The objective is no longer that of building copiers with the most features at the lowest price, but rather those that will, over the long term, provide copies at the lowest cost per page. The result: a priority to design for ease of replacement, reconditioning, and standardization of materials and components between models. The best copier is no longer the one that lasts the longest; it's the one that is the cheapest for the maker to recondition, upgrade, or replace. This is not an isolated example. Aircraft engine makers now adopt a strategy of providing "power by the hour". It is the same as that of the makers of copiers: they own the engines, monitor their use, and replace them on the plane at the point at which the trade-off between cost, reliability and power-providing capacity is optimized.

4.3.5 Miniaturization and multi-functionality (sector 3, Figure 4.3.1)

Today we make increasing use of precision mechanical engineering at the micron scale. CD and DVD players require the ability to position the read-head with micron precision. Hard disk drives are yet more impressive: think of storing and retrieving a gigabyte of information on a few square millimeters. MEMS, now universal as the accelerometers that trigger automobile air-bags, rely on micron-thick cantilever beams that bend under inertial forces in a (very) sudden stop. The MEMS technology promises much more. Now, there are video projectors that work, not by projecting light through an LCD screen but by reflecting it from arrays of electrostatically actuated mirrors, each a few microns across, allowing much higher light intensities. There are even studies of mechanical micro-processors using bi-stable mechanical switching that could complete with semiconductor switching for information processing.

The smaller you make the components of a device, the greater the functionality you can pack into it. Think of mobile phones, PDAs, mini-disk players and, above all, portable computers so small that they fit in a jacket pocket without causing a bulge. All these involve components with mechanical functions: protection, positioning, actuation, sensing. Micro-engineering no longer means Swiss watches. It means almost all the devices with which we interact continuously throughout a working day.

Miniaturization imposes new demands on materials. As devices become smaller, it is mechanical failures that become design-limiting. In the past the allowable size and weight of casings, connectors, keypads, motors, actuators gave plenty of margin of safety on bending stiffness, strength, wear-rates, and corrosion-rates. But none of these scales linearly with size. If we measure scale by a characterizing length L, bending stiffness scales as L^3 or L^4, strength as L^2 or L^3, and wear and corrosion, if measured by fractional loss of section per unit time, scales as $1/L$. Thus the smaller the device, the greater are the demands placed on the materials of which it is made.

That is the bad news. The better news is that, since the device is smaller, the amount of material needed to make it is less. Add to that the fact that consumers want small size with powerful functionality, and will pay more for it. The result: expensive, high-performance materials that would not merit consideration for bulky products become viable for this new, miniaturized, generation. Think of the armature of an air-bag actuator: it weighs about 1 mg of almost any material — even gold — costs, in its raw state, almost nothing.

So, it seems, material cost no longer limits choice. It is processing that imposes restrictions. Making micron-scale products with precision presents new processing challenges. The watchmakers of Switzerland perfected tools to meet their needs, but the digital watch decimated the mechanical watch business — their methods were not transferable to the new form of miniaturization, which requires mass production at low cost. The makers of micro-processors, on the other hand, had already

coped with this sort of scaling-down in non-mechanical systems — their manufacturing methods, rather than those of the watchmaker, have been adapted to make micro-machines. But in adopting them the menu of materials is drastically curtailed. In their present form these processes can shape silicon, silicon oxides, nitride and carbide, thin films of copper and gold, but little else. The staples of large-scale engineering — carbon and alloy steels, aluminum alloys, polyolefins, glass — do not appear. The current challenge is to broaden the range of micro-fabrication processes to allow a wider choice of materials to be manipulated.

4.3.6 Concern for the environment and for the individual (sector 4, Figure 4.3.1)

We live in a carbon-burning economy. Energy from fossil fuels has given the developed countries high standards of living and made them wealthy. Other, much more populous, countries aspire to the same standards of life and wealth and are well advanced in acquiring them. Yet the time is fast approaching when burning hydrocarbons will no longer meet energy needs, and even if it could, the burden it places on the natural environment will force limits to be set. Short-term measures to address this and the implications these have for material selection were the subject of Chapter 4.1. The changes necessary to allow longer term sustainable development will have to be much further-reaching, changing the ways in which we manufacture, transport ourselves and our goods, and live, with major impact on the way materials, central to all of these, are used.

It is one thing to point out that this force for change is a potent one, but quite another to predict what its consequences will be. It is possible to point to fuel-cell technology, energy from natural sources and safe nuclear energy (all with material challenges) as a way forward, but — in the sense we are discussing here — even these are short-term solutions. Remember the figure cited in the introduction to Chapter 4.1: at a global growth rate of 3% per year we will mine, process, and dispose of more "stuff" in the next 25 years than in the entire history of mankind. The question that has no present answer is: how can a world population growing at about 3% per year, living in countries with economic growth rates of between 2% and 20% per year, continue to meet their aspirations with zero — or even negative — growth in consumption of energy and materials?

Wealth has another dimension: it enables resources to be allocated to health care, for which, with a population with increasing lifespan, there is growing demand. A surprisingly large fraction of the organs in a human body perform predominantly mechanical functions: teeth for cutting and grinding, bone to support structural loads, joints to allow articulation, the heart to pump blood and artery to carry it under pressure to the extremities of the body, muscle to actuate, skin to provide flexible protection. Aging or accidental damage frequently causes one or more of these to malfunction. Being mechanical, it is possible, in principle, to replace them. One way to do so is to use real body parts, but the shear number of patients requiring such treatment and the ethical and other difficulties of using human replacements drives efforts to develop of artificial substitutes. Man-made teeth and bone implants, hip, and knee joint replacements, artificial artery and skin, even hearts exist already and are widely used. But, at present, they are exceedingly expensive, limiting their availability, and, for the most part, they are only crude substitutes for the real thing. There is major incentive for change here, stimulating research into affordable materials for artificial organs of all types.

Studies show that, in a world that is aging, the usability of many products and services is unnecessarily challenging to people, whether they are young or old, able-bodied, or less able. Until recently the goal of many designers was to appeal to the youth market (meaning the 15–35 age group) as the mainstay of their business, thereby developing products that, often, could not be used by others. Astute companies have perceived the problem and have sought to redesign their products and services to make them accessible to a wider clientele. The US company OXO has seen sales of its Good Grips range of kitchen and garden tools grow by 50% year on year, and the UK supermarket Tesco has added thousands of internet customers by getting rid of clever but confusing graphics on its website and instead making it quick to download and simple and intuitive to navigate, thereby including a wider spectrum of people in on-line shopping. There is a powerful business case for *inclusive design* — design to ensure that the products and services address the needs of the widest possible audience. The social case is equally compelling — at present many people are denied use of electronic and other products because they are unable to understand how to use them or lack the motor skills to do so.

All products exclude some users, sometimes deliberately — the childproof medication bottle for instance — but more usually the exclusion is unintended and unnecessary. Many services rely on the use of products for their delivery, so unusable products deny people access to these too. Video recorders are the standing joke of poor usability, regularly derided in reviews. Installing a VCR is now such a complex task that many retailers offer an installation service. Using them is little better — almost all users have at some point recorded the wrong channel, set the wrong time, recorded the

picture but not the sound, or simply given up in despair. A much larger number of contemporary products exclude those with disability — limited sight, hearing, physical mobility or strength, or mental acuity. Countering design exclusion is a growing priority of government and is increasingly seen as important by product makers. This thinking influences material choice, changing the constraints and objectives for their selection. The use of materials that, through color or feel, communicate the function of a control, materials to give good grip, materials to insulate or protect becomes a priority.

4.3.7 Summary and conclusions

Powerful forces drive the development of new and improved materials, encourage substitution, and modify the way in which materials are produced and used. Market forces and military prerogatives, historically the most influential, remain the strongest. The ingenuity of research scientists, too, drives change by revealing a remarkable spectrum of new materials with exciting possibilities, though the time it takes to develop and commercialize them is long: typically 15 years from laboratory to market place.

Today additional new drivers influence material development and use. Increasing wealth creates markets for ever more sophisticated products. The trend to products that are smaller in size, lighter in weight, and have greater functionality makes increasing demands on the mechanical properties of the materials used to make them. Greater insistence on product reliability and safety, with the manufacture held liable for failures or malfunctions, requires materials with consistently reproducible properties and processes that are tightly controlled. Concern for the impact of industrial growth on the natural environment introduces the new objective of selecting materials in such a way as to minimize it. And the perception that many products exclude users by their complexity and difficulty of use drives a reassessment of the way in which they are designed and the choice of materials to make them.

The result is that products that were seen as optimal yesterday are no longer optimal today. There is always scope for reassessing designs and the choice of materials to implement them.

Further reading

Defence and aerospace materials and structures, National advisory committee (NAC) Annual report 2000. http://www.iom3.org/foresight/nac/html

Delivering the Evidence: Defra's Science and Innovation Strategy (2003–06), Defra. May 2003. http://www.defra.gov.uk/science/S_IS

Energy White Paper: Our energy future — creating a low carbon economy, DTI (Cm 5761). TSO London, February 2003. http://www.dti.gov.uk/energy/whitepaper/index.shtml

Keates, S. and Clarkson, J. (2004) *Countering Design Exclusion — an Introduction to Inclusive Design*, Springer Verlag, London, UK. ISBN 1–85233–769–9. *(An in-depth study of design exclusion and ways to overcome it.)*

van Griethuysen, A.J. (ed.) (1987) *New Applications of Materials*, Scientific, and Technical Publications Ltd, The Hague, The Netherlands. *(A Dutch study of material evolution.)*

Williams, J.C. (1998) Engineering Competitive Materials, *The Bridge* 28(4). *(The Bridge is the journal of the US National Academy of Engineering. Dr. Williams has long experience in the aerospace industry and in academia.)*

Section **Five**

New materials

This page is intentionally left blank

Chapter 5.1

5.1

Advanced alloys

R.E. Smallman and A.H.W. Ngan

5.1.1 Introduction

In this chapter we will outline some of the developments and properties of modern metallic alloys. Crucial to these materials have been the significant developments that have taken place in manufacturing, made possible by a more detailed understanding of the manufacturing process itself and of the behavior of the material during both processing and in-service performance. Casting techniques in particular have advanced much over the past decade and now provide reliable clean material with precision. Process modeling is developing to the extent that the process designer is able to take the microstructural specification for a given composition, which controls the properties of the material, and define an optimum manufacturing route to provide the desired material and performance. Modern alloys therefore depend on the proper integration of alloy composition and structure with processing to produce the desired properties and performance.

5.1.2 Commercial steels

5.1.2.1 Plain carbon steels

Carbon is an effective, cheap, hardening element for iron and hence a large tonnage of commercial steels contains very little alloying element. They may be divided conveniently into low-carbon (<0.3% C), medium-carbon (0.3–0.7% C) and high-carbon (0.7–1.7% C) steels. Figure 5.1.1 shows the effect of carbon on the strength and ductility. The low-carbon steels combine moderate strength with excellent ductility and are used extensively for their fabrication properties in the annealed or normalized condition for structural purposes, i.e. bridges, buildings, cars and ships. Even above about 0.2% C, however, the ductility is limiting for deep-drawing operations, and brittle fracture becomes a problem, particularly for welded thick sections. Improved low-carbon steels (<0.2% C) are produced by deoxidizing or 'killing' the steel with Al or Si, or by adding Mn to refine the grain size. It is now more common, however, to add small amounts (<0.1%) of Nb, which reduces the carbon content by forming NbC particles. These particles not only restrict grain growth, but also give rise to strengthening by precipitation hardening within the ferrite grains. Other carbide formers, such as Ti, may be used but because Nb does not deoxidize, it is possible to produce

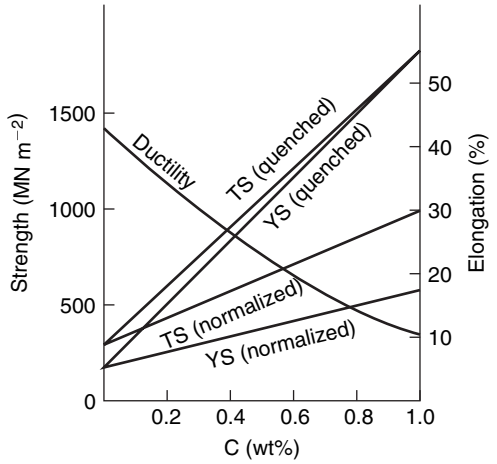


Fig. 5.1.1 Influence of carbon content on the strength and ductility of steel.

Physical Metallurgy and Advanced Materials; ISBN: 9780750669061
Copyright © 2007 Elsevier Ltd. All rights of reproduction, in any form, reserved.

a semi-killed steel ingot which, because of its reduced ingot pipe, gives increased tonnage yield per ingot cast.

Medium-carbon steels are capable of being quenched to form martensite and tempered to develop toughness with good strength. Tempering in higher-temperature regions (i.e. 350–550 °C) produces a spheroidized carbide which toughens the steel sufficiently for use as axles, shafts, gears and rails. The process of ausforming can be applied to steels with this carbon content to produce even higher strengths without significantly reducing the ductility. The high-carbon steels are usually quench hardened and lightly tempered at 250 °C to develop considerable strength with sufficient ductility for springs, dies and cutting tools. Their limitations stem from their poor hardenability and their rapid softening properties at moderate tempering temperatures.

5.1.2.2 Alloy steels

In low/medium alloy steels, with total alloying content up to about 5%, the alloy content is governed largely by the hardenability and tempering requirements, although solid solution hardening and carbide formation may also be important. Some of these aspects have already been discussed, the main conclusions being that Mn and Cr increase hardenability and generally retard softening and tempering; Ni strengthens the ferrite and improves hardenability and toughness; Cu behaves similarly but also retards tempering; Co strengthens ferrite and retards softening on tempering; Si retards and reduces the volume change to martensite, and both Mo and V retard tempering and provide secondary hardening.

In larger amounts, alloying elements either open up the austenite phase field, as shown in Figure 5.1.2a, or close the γ-field (Figure 5.1.2b). 'Full' metals with atoms like hard spheres (e.g. Mn, Co, Ni) favor close-packed structures and open the γ-field, whereas the stable bcc transition metals (e.g. Ti, V, Cr, Mo) close the field and form what is called a γ-loop. The development of austenitic steels, an important class of ferrous alloys, is dependent on the opening of the γ-phase field. The most common element added to iron to achieve this effect is Ni, as shown in Figure 5.1.2a. From this diagram the equilibrium phases at lower temperatures for alloys containing 4–40% Ni are ferrite and austenite. In practice, it turns out that it is unnecessary to add the quantity of Ni to reach the γ-phase boundary at room temperature, since small additions of other elements tend to depress the γ/α transformation temperature range, so making the γ metastable at room temperature. Interstitial C and N, which most ferrous alloys contain, also expand the γ-field because there are larger interstices in the fcc than the bcc structure. The other common element which expands the γ-field is Mn. Small amounts (<1%) are usually present in most commercial steels to reduce the harmful effect of FeS. Up to 2% Mn may be added to replace the more expensive Ni, but additions in excess of this concentration have little commercial significance until 12% Mn is reached.

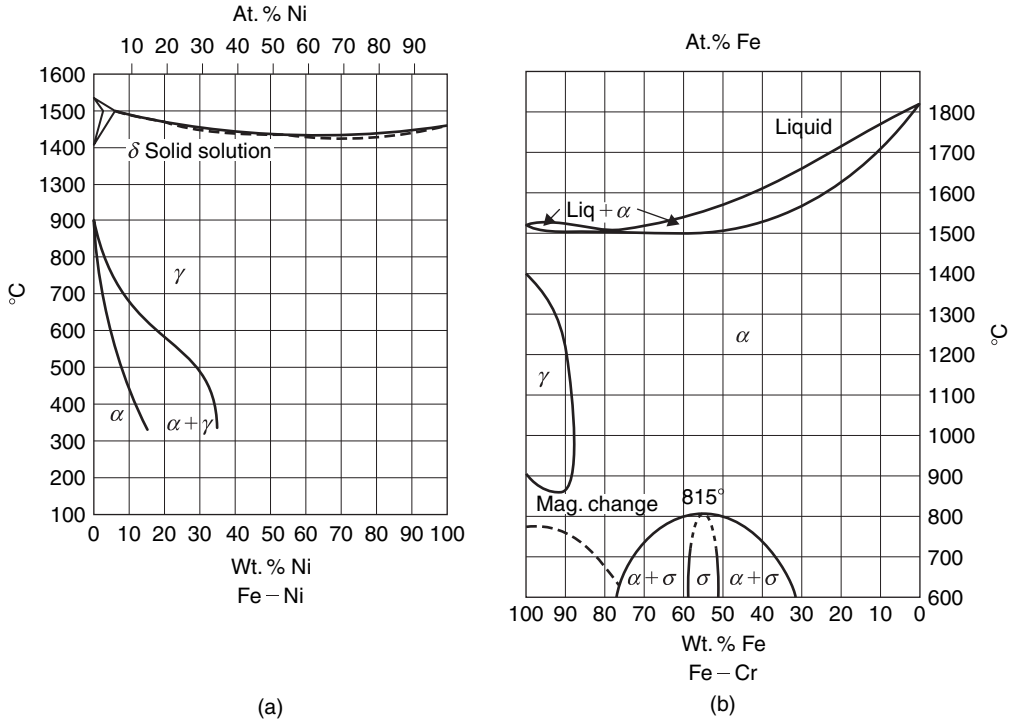


Fig. 5.1.2 Effect of Ni (a) and Cr (b) on γ-field (from Smithells, 1967).

Hadfield's steel contains 12–14% Mn and 1% C, is noted for its toughness, and is used in railway points, drilling machines and rock crushers. The steel is water-quenched to produce austenite. The fcc structure has good fracture resistance and, having a low stacking-fault energy, work-hardens very rapidly. During the abrasion and work hardening the hardening is further intensified by a partial strain transformation of the austenite to martensite; this principle is also used in the sheet-forming of stainless steels (see below).

To make the austenitic steels resistant to oxidation and corrosion the element Cr is usually added in concentrations greater than 12%. Chromium closes the γ-field, however, and with very low carbon contents single-phase austenite cannot be produced with the stainless (>12%) composition. These alloys form the stainless (ferritic) irons and are easily fabricated for use as furnace components. Increasing the carbon content expands the γ-loop and in the medium-carbon range Cr contents with good stainless qualities ($\approx$15–18%) can be quench-hardened for cutlery purposes, where martensite is required to give a hard, sharp cutting edge. The combination of both Cr and Ni (i.e. 18/8) produces the metastable austenitic stainless steel which is used in chemical plant construction, kitchenware and surgical instruments because of its ductility, toughness and cold-working properties. Metastable austenitic steels have good press-forming properties because the strain-induced transformation to martensite provides an additional strengthening mechanism to work hardening, and moreover counteracts any drawing instability by forming martensite in the locally thinned, heavily deformed regions.

High-strength transformable stainless steels with good weldability to allow fabrication of aircraft and engine components have been developed from the 0.05–0.1% C/12% Cr/stainless steels by secondary hardening addition (1.5–2% Mo; 0.3–0.5% V). Small additions of Ni or Mn (2%) are also added to counteract the ferrite-forming elements Mo and V to make the steel fully austenitic at the high temperatures. Air quenching to give α followed by tempering at 650 °C to precipitate Mo_2C produces a steel with high yield strength (0.75 GN m^{-2}), high TS (1.03 GN m^{-2}), and good elongation and impact properties. Even higher strengths can be achieved with stainless (12–16% Cr; 0.05% C) steels which, although austenitic at room temperature (5% Ni; 2% Mn), transform on cooling to −78 °C. The steel is easily fabricated at room temperature, cooled to control the transformation and finally tempered at 650–700 °C to precipitate Mo_2C.

5.1.2.3 Maraging steels

A serious limitation in producing high-strength steels is the associated reduction in fracture toughness. Carbon is one of the elements which mostly affects the toughness and hence in alloy steels it is reduced to as low a level as possible, consistent with good strength. Developments in the technology of high-alloy steels have produced high strengths in steels with very low carbon contents (<0.03%) by a combination of martensite and age hardening, called maraging. The maraging steels are based on an Fe–Ni containing between 18% and 25% Ni to produce massive martensite on air cooling to room temperature. Additional hardening of the martensite is achieved by precipitation of various intermetallic compounds, principally Ni_3Mo or Ni_3(Mo, Ti), brought about by the addition of roughly 5% Mo and 8% Co as well as small amounts of Ti and Al; the alloys are solution heat-treated at 815 °C and aged at about 485 °C. Many substitutional elements can produce age hardening in Fe–Ni martensites, some strong (Ti, Be), some moderate (Al, Nb, Mn, Mo, Si, Ta, V) and other weak (Co, Cu, Zr) hardeners. There can, however, be rather strong interactions between elements such as Co and Mo, in that the hardening produced when these two elements are present together is much greater than if added individually. It is found that A_3B-type compounds are favored at high Ni or (Ni + Co) contents and A_2B Laves phases at lower contents.

In the unaged condition maraging steels have a yield strength of about 0.7 GN m^{-2}. On aging this increases up to 2.0 GN m^{-2} and the precipitation strengthening is due to an Orowan mechanism according to the relation $\sigma = \sigma_0 + (\alpha\mu b/L)$, where σ_0 is the matrix strength, α a constant and L the interprecipitate spacing. The primary precipitation-strengthening effect arises from the (Co + Mo) combination, but Ti plays a double role as a supplementary hardener and a refining agent to tie up residual carbon. The alloys generally have good weldability, resistance to hydrogen embrittlement and stress corrosion, but are used mainly (particularly the 18% Ni alloy) for their excellent combination of high strength and toughness.

5.1.2.4 High-strength low-alloy steels

The requirement for structural steels to be welded satisfactorily has led to steels with lower C (<0.1%) content. Unfortunately, lowering the C content reduces the strength and this has to be compensated for by refining the grain size. This is difficult to achieve with plain C-steels rolled in the austenite range, but the addition of small amounts of strong carbide-forming elements (e.g. <0.1% Nb) causes the austenite boundaries to be pinned by second-phase particles and fine grain sizes (<10 μm) to be produced by controlled rolling. Nitrides and carbonitrides as well as carbides, predominantly fcc and mutually soluble in each other, may feature as suitable

grain refiners in high-strength low-alloy (HSLA) steels; examples include AlN, Nb(CN), V(CN), (NbV)CN, TiC and Ti(CN). The solubility of these particles in the austenite decreases in the order VC, TiC, NbC while the nitrides, with generally lower solubility, decrease in solubility in the order VN, AlN, TiN and NbN. Because of the low solubility of NbC, Nb is perhaps the most effective grain size controller. However, Al, V and Ti are effective in high-nitrogen steels, Al because it forms only a nitride, V and Ti by forming V(CN) and Ti(CN), which are less soluble in austenite than either VC or TiC.

The major strengthening mechanism in HSLA steels is grain refinement, but the required strength level is usually obtained by additional precipitation strengthening in the ferrite. VC, for example, is more soluble in austenite than NbC, so if V and Nb are used in combination, then on transformation of the austenite to ferrite, NbC provides the grain refinement and VC precipitation strengthening; Figure 5.1.3 shows a stress–strain curve from a typical HSLA steel.

Solid–solution strengthening of the ferrite is also possible. Phosphorus is normally regarded as deleterious due to grain boundary segregation, but it is a powerful strengthener, second only to carbon. In car construction, where the design pressure is for lighter bodies and energy saving, HSLA steels, rephosphorized and bake-hardened to increase the strength further, have allowed sheet gauges to be reduced by 10–15% while maintaining dent resistance. The bake hardening arises from the locking of dislocations with interstitials during the time at the temperature of the paint-baking stage of manufacture.

5.1.2.5 Dual-phase steels

Much research into the deformation behavior of speciality steels has been aimed at producing improved strength while maintaining good ductility. The conventional means of strengthening by grain refinement, solid-solution additions (Si, P, Mn) and precipitation hardening by V, Nb or Ti carbides (or carbonitrides) have been extensively explored and a conventionally treated HSLA steel would have a lower yield stress of 550 MN m^{-2}, a TS of 620 MN m^{-2} and a total elongation of about 18%. In recent years an improved strength–ductility relationship has been found for low-carbon, low-alloy steels rapidly cooled from an annealing temperature at which the steel consisted of a mixture of ferrite and austenite. Such steels have a microstructure containing principally low-carbon, fine-grained ferrite intermixed with islands of fine martensite and are known as DP steels. Typical properties of this group of steels would be a TS of 620 MN m^{-2}, a 0.2% offset flow stress of 380 MN m^{-2} and a 3% offset flow stress of 480 MN m^{-2}, with a total elongation ≈28%.

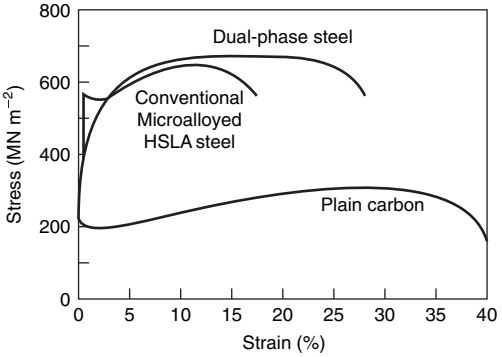


Fig. 5.1.3 Stress–strain curves for plain carbon, HSLA and DP steels.

The implications of the improvement in mechanical properties are evident from an examination of the nominal stress–strain curves, shown in Figure 5.1.3. The DP steel exhibits no yield discontinuity but work-hardens rapidly so as to be just as strong as the conventional HSLA steel when both have been deformed by about 5%. In contrast to ferrite–pearlite steels, the work-hardening rate of DP steel increases as the strength increases. The absence of discontinuous yielding in DP steels is an advantage during cold-pressing operations and this feature, combined with the way in which they sustain work hardening to high strains makes them attractive materials for sheet-forming operations. The flow stress and tensile strength of DP steels increase as the volume fraction of hard phase increases with a corresponding decrease in ductility; about 20% volume fraction of martensite produces the optimum properties.

The DP is produced by annealing in the $(\alpha + \gamma)$ region, followed by cooling at a rate which ensures that the γ-phase transforms to martensite, although some retained austenite is also usually present, leading to a mixed martensite–austenite (M–A) constituent. To allow air cooling after annealing, microalloying elements are added to low-carbon–manganese–silicon steel, particularly vanadium or molybdenum and chromium. Vanadium in solid solution in the austenite increases the hardenability but the enhanced hardenability is due mainly to the presence of fine carbonitride precipitates which are unlikely to dissolve in either the austenite or the ferrite at the temperatures employed and thus inhibit the movement of the austenite/ferrite interface during the post-anneal cooling.

The martensite structure found in DP steels is characteristic of plate martensite having internal microtwins. The retained austenite can transform to martensite during straining, thereby contributing to the increased strength and work hardening. Interruption of the cooling, following intercritical annealing, can lead to stabilization of the austenite with an increased strength on subsequent

deformation. The ferrite grains (≈ 5 μm) adjacent to the martensite islands are generally observed to have a high dislocation density resulting from the volume and shape change associated with the austenite to martensite transformation. Dislocations are also usually evident around retained austenitic islands due to differential contraction of the ferrite and austenite during cooling.

Some deformation models of DP steels assume both phases are ductile and obey the Ludwig relationship, with equal strain in both phases. Measurements by several workers have, however, clearly shown a partitioning of strain between the martensite and ferrite, with the mixed (M–A) constituent exhibiting no strain until deformations well in excess of the maximum uniform strain. Models based on the partitioning of strain predict a linear relationship between yield stress, TS and volume fraction of martensite, but a linear relationship is not sensitive to the model. An alternative approach is to consider the microstructure as approximating to that of a dispersion-strengthened alloy. This would be appropriate when the martensite does not deform and still be a good approximation when the strain difference between the two phases is large. Such a model affords an explanation of the high work-hardening rate arising from the interaction of the primary dislocations with the dense 'tangle' of dislocations generated in the matrix around the hard phase islands.

Several workers have examined DP steels to determine the effect of size and volume fraction of the hard phase. Figure 5.1.4 shows the results at two different strain values and confirms the linear relationship between work-hardening rate ($d\sigma/d\varepsilon$) and $(f/d)^{1/2}$ predicted by the dispersion-hardening theory. Increasing the hard-phase volume fraction while keeping the island diameter constant increases the work-hardening rate, increases the TS but decreases the elongation. At constant volume fraction of hard phase, decreasing the mean island diameter produces no effect on the tensile strength but increases the work-hardening rate and the maximum uniform elongation (Figure 5.1.5). Thus, the strength is improved by increasing the volume fraction of hard phase, while the work hardening and ductility are improved by reducing the hard-phase island size. Although DP steels contain a complex microstructure it appears from their mechanical behavior that they can be considered as agglomerates of non-deformable hard particles, made up of martensite and/or bainite and/or retained austenite, in a ductile matrix of ferrite. Consistent with the dispersion-strengthened model, the Bauschinger effect, where the flow stress in compression is less than that in tension, is rather large in DP steels, as shown in Figure 5.1.6, and increases with increase in martensite content up to about 25%. The Bauschinger effect arises from the long-range back-stress exerted by the martensite islands, which adds to the applied stress in reversed straining.

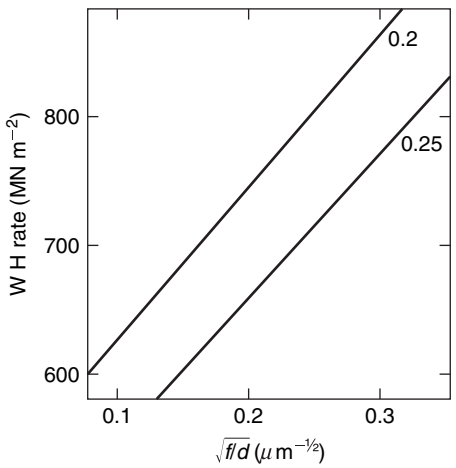


Fig. 5.1.4 Dependence of work-hardening rate on (volume fraction *f*/particle size)$^{1/2}$ for a DP steel at strain values of 0.2 and 0.25 (after Balliger and Gladman, 1981).

The ferrite grain size can give significant strengthening at small strains, but an increasing proportion of the strength arises from work hardening and this is independent of grain-size changes from about 3 to 30 μm. Solid–solution strengthening of the ferrite (e.g. by silicon) enhances the work-hardening rate; P, Mn and V are also beneficial. The absence of a sharp yield point must imply that the DP steel contains a high density of mobile dislocations. The microstructure exhibits such a dislocation density around the martensite islands, but why these remain unpinned at ambient temperature is still in doubt, particularly as strain aging is significant on aging between 423 and 573 K. Intercritical annealing allows a partitioning of the carbon to produce very low carbon ferrite, while aluminum- or silicon-killed steels have limited nitrogen remaining in solution. However, it is doubtful whether the concentration of interstitials is sufficiently low to prevent strain aging at low temperature; hence it is considered more likely that continuous yielding is due to the residual stress fields surrounding second-phase islands. Two possibilities then arise: (1) yielding can start in several regions at the same time rather than in one local region, which initiates a general yield process catastrophically, and (2) any local region is prevented from yielding catastrophically because the glide band has to overcome a high back-stress from the M–A islands. Discontinuous yielding on aging at higher temperatures is then interpreted in terms of the relaxation of these residual stresses, followed by classical strain-aging.

In DP steels the n value ≈ 0.2 gives the high and sustained work-hardening rate required when stretch formability is the limiting factor in fabrication. However, when fracture *per se* is limiting, DP steels probably

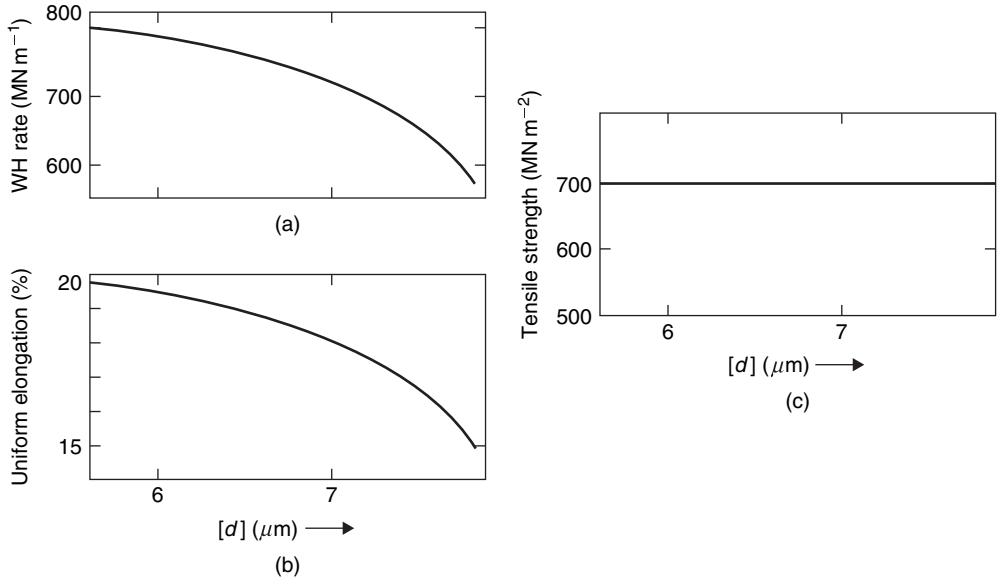


Fig. 5.1.5 Effect of second-phase particles size *d* at constant volume fraction *f* on: (a) work-hardening rate, (b) elongation and (c) tensile strength (after Balliger and Gladman, 1981).

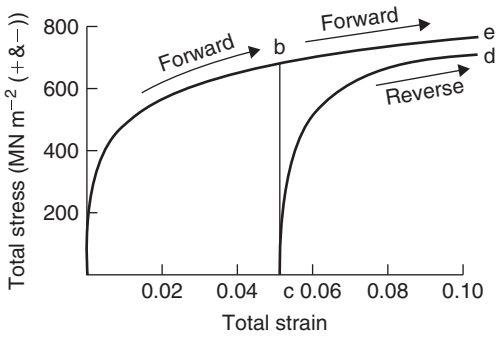


Fig. 5.1.6 Bauschinger tests for a 0.06% C–1.5% Mn–0.85% Si DP steel (courtesy of D. V. Wilson).

perform no better than other steels with controlled inclusion content. Tensile failure of DP steels is initiated either by decohesion of the martensite–ferrite interface or by cracking of the martensite islands. Improved fracture behavior is obtained when the martensite islands are unconnected, when the martensite–ferrite interface is free from precipitates to act as stress raisers, and when the hard phase is relatively tough. The optimum martensite content is considered to be 20%, because above this level void formation at hard islands increases markedly.

5.1.2.6 Mechanically alloyed steels

For strengthening at high temperatures, dispersion strengthening with oxide, nitride or carbide particles is an attractive possibility. Such dispersion-strengthened materials are usually produced by powder processing, a special form of which is known as mechanical alloying (MA).

Mechanical alloying is a dry powder, high-energy ball-milling process in which the particles of elemental or pre-alloyed powder are continuously welded together and broken apart until a homogeneous mixture of the matrix material and dispersoid is produced. Mechanical alloying is not simply mixing on a fine scale but one in which true alloying occurs. The final product is then consolidated by a combination of high temperature and pressure (i.e. extrusion of canned powder) or hot isostatic pressing (i.e. HIPing). Further processing is by thermomechanical processing (TMP) to produce either (1) fine equiaxed grains for good room-temperature strength and good fatigue strength or (2) coarser, elongated grains to give good high-temperature stress–rupture strength and thermal-fatigue resistance.

Various types of ferrous alloy have been made by mechanical alloying, including 17% Cr–7% Ni–1.2% Al precipitation-hardened austenitic martensitic steel and Fe–25Cr–6Al–2Y. However, the most highly developed material is the 20% Cr–4.5% Al ferritic stainless steel, dispersion strengthened with 0.5% Y_2O_3 (*MA 956*). *MA 956*, which has been made into various fabricated forms, has extremely good high-temperature strength (0.2% proof strength is 200 MN m^{-2} at 600 °C, 100 MN m^{-2} at 1000 °C and 75 MN m^{-2} at 1200 °C).

The high-strength capability is combined with exceptional high-temperature oxidation and corrosion resistance, associated with the formation of an aluminum oxide scale which is an excellent barrier to carbon. No carburization occurs in hydrogen–methane mixtures at 1000 °C. Sulfidation resistance is also good.

MA 956 was originally developed for use in sheet form in gas-turbine combustors but, with its combination of high strength up to 1300 °C, corrosion resistance and formability, the alloy has found many other applications in power stations, including oil and coal burners and swirlers, and fabricated tube assemblies for fluid-bed combustion.

5.1.2.7 Designation of steels

The original system for labeling wrought steels was devised in 1941 and used En numbers. This system was replaced in 1976 by the British Standard (BS) designation of steels, which uses a six-unit system. Essentially, it enables the code to express composition, steel type and supply requirements. The latter is shown by three letters: M means supply to specified mechanical properties, H supply to hardenability requirements and A supply to chemical analysis requirements. For convenience, steels are divided into types: namely, carbon and carbon–manganese steels, free-cutting steels, high-alloy steels and alloy steels. For example, carbon and carbon–manganese steels are designated by mean of Mn/letter/mean of C. Thus, 080H41 signifies 0.6–1.0 Mn/hardenability requirement/0.38–0.45 C. Free-cutting steels are designated by 200–240/letter/mean of C. Thus, 225M44 signifies free-cutting 0.2–0.3 S/mechanical properties requirement/0.4–0.48 C with 1.3–1.7 Mn. High-alloy steels include stainless and valve steels. The designation is similar to the AISI system and is given by 300–499/letters/variants 11–19. Thus, 304S15 (previously known as Type 304 as used by the AISI) signifies 0.06 max. C/8–11 Ni/17.5–19 Cr. Alloy steels are designated by 500–999/letter/mean of C. Thus, 500–519 are Ni steels, 520–539 Cr steels, 630–659 Ni–Cr steels, 700–729 Cr–Mo steels and 800–839 Ni–Cr–Mo steels. Typically, 530M40 signifies 0.36–0.44 C/0.9–1.2 Cr, supplied to mechanical properties.

Tables 5.1.1 and 5.1.2 give the compositions of typical carbon, alloy and stainless steels.

5.1.3 Cast irons

In the iron–carbon system, carbon is thermodynamically more stable as graphite than cementite. At the low carbon contents of typical steels, graphite is not formed, however, because of the sluggishness of the reaction to graphite. But when the carbon content is increased to that typical of cast irons (2–4% C), either graphite or cementite may separate depending on the cooling rate, chemical (alloy) composition and heat treatment (see Figure 5.1.7). When the carbon exists as cementite, the cast irons are referred to as white because of the bright fracture produced by this brittle constituent. In gray cast irons the carbon exists as flakes of graphite embedded in the ferrite–pearlite matrix and these impart a dull gray appearance to the fracture. When both cementite and graphite are present a 'mottled' iron is produced.

High cooling rates, which tend to stabilize the cementite, and the presence of carbide formers give rise to white irons. The addition of graphite-forming elements (Si, Ni) produces gray irons, even when rapidly cooled if the Si is above 3%. These elements, particularly Si, alter the eutectic composition which may be taken into account by using the carbon equivalent of the cast iron, given by [total % C + (% Si + %*P*)/3], rather than the true carbon content. Phosphorus is present in most cast irons as a low melting point phosphide eutectic, which improves the fluidity of the iron by lengthening the solidification period; this favors the decomposition of

Table 5.1.1 Compositions of some carbon and alloy steels

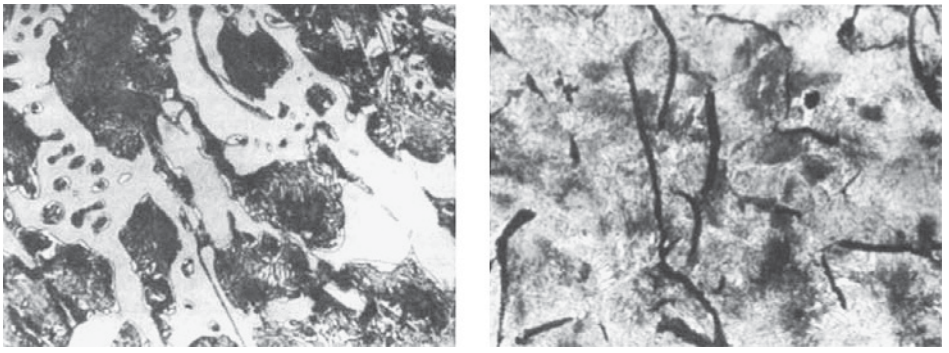
BS designation	AISI-SAE number*	% C	% Si	% Mn	% Ni	% Cr	% others
040A20	1020	0.18–0.23		0.30–0.50			
080A62	1060	0.60–0.65		0.70–0.90			
070A78	1080	0.75–0.82		0.60–0.80			
150M36	1340	0.32–0.40		1.30–1.70			
212M44	1140	0.40–0.48	0.25 max	1.00–1.40			0.12–0.2 S
527A19	5120	0.17–0.22		0.70–0.90		0.70–0.90	
665H20	4620	0.17–0.23	0.10–0.35	0.35–0.75	1.50–2.00		0.20–0.30 Mo
708M40	4140	0.36–0.44	0.10–0.35	0.70–1.00		0.90–1.20	0.15–0.25 Mo

* Approximately equivalent composition.

Table 5.1.2 Compositions and properties of some stainless steels

Steel	BS designation	% C	% Cr	% Ni	Others	Tensile strength (MN m^{-2})	Yield strength (MN m^{-2})	% elongation	Condition
Austenitic									
304	304S15	0.08	18–20	8.0–10.5		517	207	30	Annealed
						1276	966	9	Cold worked
316	316S16	0.08	16–18	10–14	2–3% Mo	517	207	30	Annealed
321	321S12	0.08	17–19	9–12	Ti (5 × % C)	586	241	55	Annealed
347	347S17	0.08	17–19	9–13	Nb (10 × % C)	621	241	50	Annealed
Ferritic									
430	430S15	0.12	16–18			448	207	22	Annealed
Martensitk									
416	416S21	0.15	12–14		0.60% Mo	1241	966	18	Quenched and tempered
431	431S29	0.20	15–17	1.25–2.50		1379	1035	16	
Precipitation hardening									
17–4		0.07	16–18	3–5	0.15–0.45% Nb	1310	1172	10	Age hardened
17–7		0.09	16–18	6.5–7.8	0.75–1.25% Al	1655	1586	6	

cementite. Gray cast iron is used for a wide variety of applications because of its good strength/cost ratio. It is easily cast into intricate shapes and has good machinability, since the chips break off easily at the graphite flakes. It also has a high damping capacity and hence is used for lathe and other machine frames where vibrations need to be damped out. The limited strength and ductility of gray cast iron may be improved by small additions of the carbide formers (Cr, Mo), which reduce the flake size and refine the pearlite. The main use of white irons is as a starting material for malleable cast iron, in which the cementite in the casting is decomposed by annealing. Such irons contain sufficient Si (<1.3%) to promote the decomposition process during the heat treatment, but not enough to produce graphite flakes during casting. White-heart malleable iron is made by heating the casting in an oxidizing environment (e.g. hematite iron ore at 900 °C for 3–5 days). In thin sections the carbon is oxidized to ferrite and, in thick sections, ferrite at the outside gradually changes to graphite clusters in a ferrite–pearlite matrix near the inside. Black-heart malleable iron is made by annealing the white iron in a neutral packing (i.e. iron silicate slag) when the cementite is changed to rosette-shaped



(a) (b)

Fig. 5.1.7 Microstructure of cast irons: (a) white iron and (b) gray iron (×400). (a) Shows cementite (white) and pearlite; (b) shows graphite flakes, some ferrite (white) and a matrix of pearlite.

graphite nodules in a ferrite matrix. The deleterious cracking effect of the graphite flakes is removed by this process and a cast iron which combines the casting and machinability of gray iron with good strength and ductility, i.e. TS 350 MN m^{-2} and 5–15% elongation is produced. It is therefore used widely in engineering and agriculture, where intricate-shaped articles with good strength are required.

Even better mechanical properties (550 MN m^{-2}) can be achieved in cast irons, without destroying the excellent casting and machining properties, by the production of a spherulitic graphite. The spherulitic nodules are roughly spherical in shape and are composed of a number of graphite crystals, which grow radially from a common nucleus with their basal planes normal to the radial growth axis. This form of growth habit is promoted in an as-cast gray iron by the addition of small amounts of Mg or Ce to the molten metal in the ladle, which changes the interfacial energy between the graphite and the liquid. Good strength, toughness and ductility can thus be obtained in castings that are too thick in section for malleabilizing and can replace steel castings and forgings in certain applications.

Heat treating the ductile cast iron produces austempered ductile iron (ADI), with an excellent combination of strength, fracture toughness and wear resistance for a wide variety of applications in automotive, rail and heavy engineering industries. A typical composition is 3.5–4.0% C, 2–2.5% Si, 0.03–0.06% Mg, 0.015% maximum S and 0.06% maximum P. Alloying elements such as Cu and Ni may be added to enhance the heat treatability. Heat treatment of the cast ductile iron (graphite nodules in a ferrite matrix) consists of austenization at 950 °C for 1–3 h during which the matrix becomes fully austenitic, saturated with carbon as the nodules dissolve. The fully austenized casting is then quenched to around 350 °C and austempered at this temperature for 1–3 h. The austempering temperature is the most important parameter in determining the mechanical properties of ADI; high austempering temperatures (i.e. 350–400 °C) result in high ductility and toughness and lower yield and tensile strengths, whereas lower austempering temperatures (250–300 °C) result in high yield and tensile strengths, high wear resistance, and lower ductility and toughness. After austempering the casting is cooled to room temperature.

The desired microstructure of ADI is acicular ferrite plus stable, high-carbon austenite, where the presence of Si strongly retards the precipitation of carbides. When the casting is austempered for longer times than that to produce the desired structure, carbides are precipitated in the ferrite to produce bainite. Low austempering temperatures (~250 °C) lead to cementite precipitation, but at the higher austempering temperatures (300–400 °C) transition carbides are formed, ε-carbides at the lower temperatures and η-carbides at the higher. With long austempering times the high-carbon austenite precipitates χ-carbide at the ferrite–austenite boundaries. The formation of bainite does not result in any catastrophic change in properties but produces a gradual deterioration with increasing time of austempering. Typically, ADI will have a tensile strength of 1200–1500 MN m^{-2}, an elongation of 6–10% and $K_{1c} \approx 80$ MN $m^{-3/2}$. With longer austempering the elongation drops to a few percent and the K_{1c} reduces to 40 MN $m^{-3/2}$. The formation of χ-carbide at the ferrite–austenite boundaries must be avoided, since this leads to more brittle fracture. Generally, the strength is related to the volume fraction of austenite and the ferrite spacing. Figure 5.1.8 shows the microstructure of Si spheroidal graphite (SG) iron and the corresponding fracture mode.

5.1.4 Superalloys

5.1.4.1 Basic alloying features

These alloys have been developed for high-temperature service and include iron-, cobalt- and nickel-based materials, although nowadays they are principally nickel based. The production of these alloys over several decades (see Figure 5.1.9) illustrates the transition in the development of engineering materials from basic alloy composition achievements to a more process-dominated control.

In these alloys γ' (Ni_3Al) and γ^* (Ni_3Nb) are the principal strengthened by chemical and coherency strain hardening. The ordered γ'-Ni_3Al phase is an equilibrium second phase in both the binary Ni–Al and Ni–Cr–Al systems and a metastable phase in the Ni–Ti and Ni–Cr–Ti systems, with close matching of the γ' and the fcc matrix. The two phases have very similar lattice parameters ($\lesssim 0.25\%$, depending on composition) and the coherency (interfacial energy $\gamma_1 \approx$ 10–20 mJ m^2) confers a very low coarsening rate on the precipitate so that the alloy overages extremely slowly even at $0.7T_m$. In alloys containing Nb, a metastable Ni_3Nb phase occurs but, although ordered and coherent, it is less stable than γ' at high temperatures.

Another source of strengthening is due to solid-solution hardening; Cr is a major element, Co may be added up to 20% and Mo, W and Ta up to a total of 15%. These elements also dissolve in γ' so that the hardening effect may be twofold. Additions of Cr, like Co, also increase the γ' solvus and lower the stacking-fault energy.

In high-temperature service, the properties of the grain boundaries are as important as the strengthening by γ' within the grains. Grain boundary strengthening is produced mainly by precipitation of chromium and

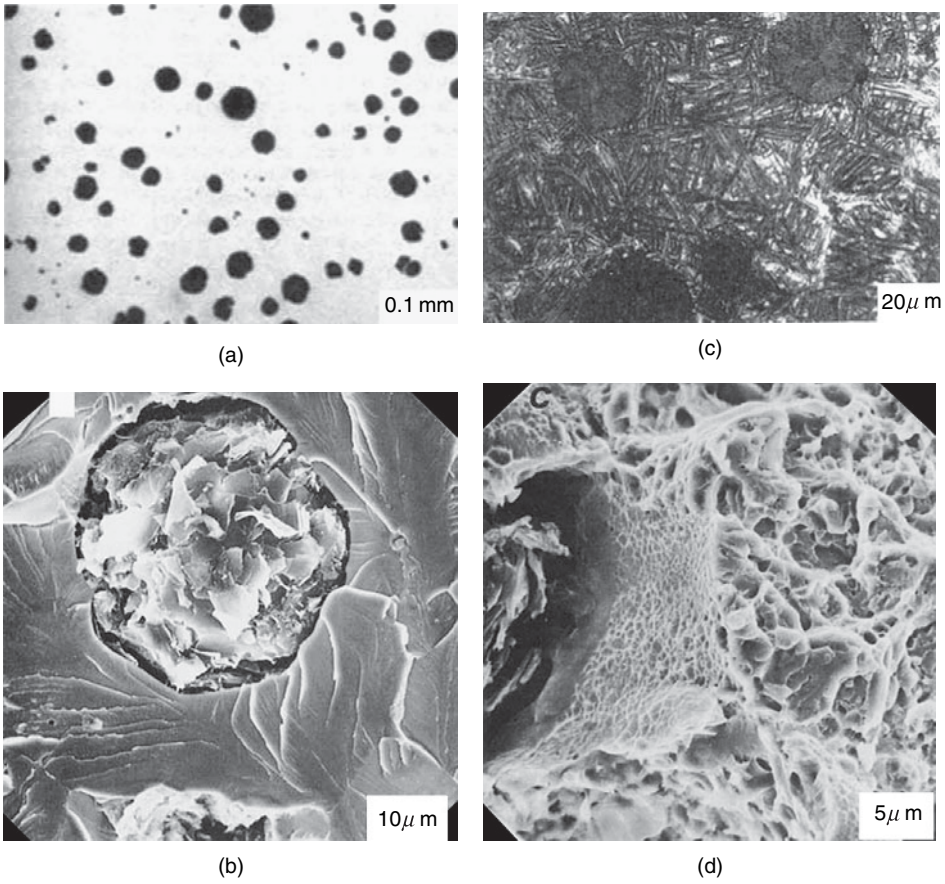


Fig. 5.1.8 Microstructure and fracture mode of silicon spheroidal graphite (SG) iron: (a, b) as-cast and (c, d) austempered at 350 °C for 1 h (L. Sidjanin and R. E. Smallman, 1992; courtesy of Institute of Materials, Minerals and Mining).

refractory metal carbides; small additions of Zr and B improve the morphology and stability of these carbides. Optimum properties are developed by multistage heat treatment; the intermediate stages produce the desired grain boundary microstructure of carbide particles enveloped in a film of γ' and the other stages produce two size ranges of γ' for the best combination of strength at both intermediate and high temperatures. Table 5.1.3 indicates the effect of the different alloying elements.

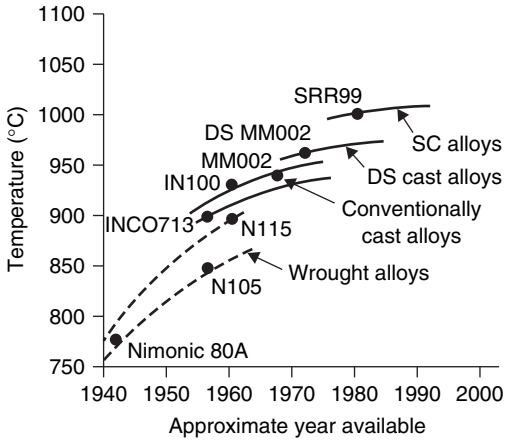


Fig. 5.1.9 Increases in temperature capability for turbine blade alloys, based on creep rupture in 1000 h at 150 MN m^{-2} (from Driver, 1985, by permission of Institute of Materials, Minerals and Mining).

Some of the nickel-based alloys have a tendency to form an embrittling σ-phase (based on the composition FeCr) after long-term in-service applications, when composition changes occur removing σ-resisting elements such as Ni and enhancing σ-promoting elements such as Cr, Mo or W. This tendency is predicted in alloy design by a technique known as Phacomp (phase computation), based on Pauling's model of hybridization of 3d electrons in transition metals. While a fraction of the 3d orbitals hybridize with P and s orbitals to create the metallic bond, the remainder forms non-bonding orbitals which partly fill the electron holes in the d-shell, increasing through the transition series to give electron hole numbers N_v for Cr (4.66), Mn (3.66), Fe (3.66), Co (1.71) and Ni (0.66). Computation shows that the γ/σ phase relation depends on the average hole number $\overline{N}_v$, given by

$$\overline{N}_v = \sum_{i=1}^{n} m_i (N_v)_i,$$

Table 5.1.3 Influence of various alloying additions in superalloys

Influence	Cr	Al	Ti	Co	Mo	W	B	Zr	C	Nb	Hf	Ta
Matrix strengthening	✓			✓	✓	✓						
γ' formers		✓	✓							✓		✓
Carbide formers	✓		✓		✓	✓				✓	✓	✓
Grain boundary strengthening							✓	✓	✓		✓	
Oxide scale formers	✓	✓										

Table 5.1.4 Composition of some Ni-based superalloys

		Composition (wt%)								
	Ni-alloy	Cr	Co	Ti	Al	Mo	C	Zr	B	Others
Waspaloy	Balance	19.5	13.5	3.0	1.3	–	0.08	0.06	0.006	
Astroloy	Balance	15.0	17.0	3.5	4.0	5.25	0.06	–	0.030	
Inco 718	Balance	19.0	3.0	0.9	0.6	3.0	0.04	–	–	Fe 20 Nb 5.2
Inco 901	42	13.0	–	3.0	0.3	5.7	0.04	–	–	Bal. Fe

where m_i, is the atomic fraction of the ith element of electron hole number N_v and n is the number of elements in the alloy. The limit of γ-phase stability is reached at $\overline{N}_v \approx 2.5$.

5.1.4.2 Nickel-based superalloy development

A major application of superalloys is in turbine materials, jet engines, both disk and blades. Initial disk alloys were *Inco 718* and *Inco 901* produced by conventional casting ingot, forged billet and forged disk route. These alloys were developed from austenitic steels, which are still used in industrial turbines, but were later replaced by *Waspaloy* and *Astroloy* as stress and temperature requirements increased. These alloys (Table 5.1.4) were turbine blade alloys with a suitably modified heat treatment for disks. However, blade material is designed for creep, whereas disk material requires tensile strength coupled with low cycle fatigue life to cope with the stress changes in the flight cycle. To meet these requirements *Waspaloy* (Table 5.1.4) was thermomechanically processed to give a fine grain size and a 40% increase in tensile strength over the corresponding blade material, but at the expense of creep life. Similar improvements for disks have been produced in *Inco 901* by TMP. More highly alloyed nickel-based disks suffer from excessive ingot segregation, which makes grain size control difficult. Further development led to alloys produced by powder processing by gas atomization of a molten stream of metal in an inert argon atmosphere and consolidating the resultant powder by HIPing to near-net shape. Such products are limited in stress application because of inclusions in the powder and, hence, to realize the maximum advantage of this process it is necessary to produce 'superclean' material by electron beam or plasma melting.

Improvements in turbine materials were initially developed by increasing the volume fraction of γ' in changing *Nimonic 80A* up to *Nimonic 115*. Unfortunately, increasing the (Ti + Al) content lowers the melting point, thereby narrowing the forging range, which makes processing more difficult. Improved high-temperature oxidation and hot corrosion performance has led to the introduction of aluminide and overlay coatings, and subsequently the development of *IN 738* and *IN 939* with much improved hot-corrosion resistance.

Further improvements in superalloys have depended on alternative manufacturing routes, particularly using modern casting technology. Vacuum casting was first used to retain high Ti and Al contents without oxidation loss. With 9–11% (Ti + Al), a 70% volume fraction of γ' has been produced in *IN 100 (Nimocast PK 24)* which does not require supplementary solid solution strengtheners and therefore gives a saving in density.[1]

[1] A Pratt and Whitney version of *IN 100 (B1900)* replaced Ti with Ta to improve the castability.

Additions of high melting point elements such as W extend the high-temperature capabilities at the expense of density. *M200* contains 12% W and 1% Nb but has limited ductility around 760 °C, which can be improved by additions of hafnium. The significant improvement in ductility and reduced porosity produced by Hf has led to its addition to other alloys (e.g. *Mar 001 (IN 100* + Hf) and *Mar 004 (IN 713* + Hf) and *M002*, which contains 10% W, 2.5% Ta and 1.5% Hf).

Creep failures are known to initiate at transverse grain boundaries and, hence, it is reasonable to aim to eliminate them in the turbine blade to gain further improvement in performance. Technologically this was achieved by directional solidified (DS) castings with columnar grains aligned along the growth direction with no grain boundaries normal to that direction. By incorporating a geometric constriction into the mold or by the use of a seed crystal it has been possible to eliminate grain boundaries entirely and grow the blade as one single crystal.

The elimination of grain boundaries immediately removes the necessity for adding grain-boundary strengthening elements, such as C, B, Zr and Hf, to the superalloy. The removal of such elements raises the melting point and allows a higher solution heat-treatment temperature with consequent improvement in chemical homogeneity and more uniform distribution of γ' precipitates. Particularly important, however, is the control of the growth direction along the [10 0] direction. The [1 0 0] alignment along the axis of the blade gives rise to an intrinsic high creep which enables thermal stresses caused by temperature gradients across the blade to be minimized.

Single-crystal blades have now been used successfully for both civil and military engines, *SSR99* replacing *Mar M002* but with improved tensile, creep and fatigue properties and a lower-density alloy *RR2000* replacing IN 100.

5.1.4.3 Dispersion-hardened superalloys

All γ'-hardened alloys experience a reduction in strength at elevated temperatures because of the solution of γ' precipitate. To produce improved high-temperature strength, alloys hardened by oxides (ODS or oxide dispersion strengthened), particularly thoria, have been developed. TD nickel (i.e. thoria dispersion-strengthened nickel or nickel with 2% ThO_2) and *TD-Nichrome* (i.e. nickel–20% Cr–2% ThO_2) are produced by mixing thoria sols with nickel-containing solutions to ensure a good dispersion. Drying the oxide powder mixture, followed by a hydrogen reduction process, produces a fine composite nickel–thoria powder. Compaction and controlled TMP during extrusion and rolling develops the structure and strength. These materials have excellent high-temperature creep resistance but poor low-temperature properties, which precludes aero-engine applications, although some other applications have been found. It is not possible to add the γ'-forming elements via the above process to produce lower-temperature strengthening because aluminum and titanium oxide are not reduced by hydrogen.

MA is a dry powder process and overcomes this problem. MA, a high-energy ball-milling process, produces a homogeneous mixture of the matrix material and dispersoid. The final product is then produced by TMP with the grain structure on recrystallization elongated in the direction of working.

These materials are generally anisotropic in their properties, but a range of MA materials is emerging with more acceptable dispersoids. *MA754* is a *Nichrome* matrix with Y_2O_3 dispersoids and *MA6000* is a more complex nickel superalloy-based yttria material (Ni–12Cr–2Mo–4W–2Ta–4.5Al–2.5Ti and $1.1Y_2O_3$) suitable for turbine blades, having the creep characteristics of the γ'-strengthened materials at low temperatures combined with the advantages of dispersion strengthening at high temperature (see Figure 5.1.10). TMP limits the amount of γ' strengthening that can be introduced and, while *MA6000* can be run at higher temperatures than conventional nickel-based superalloys, application is then limited to areas where the loadings can be kept low.

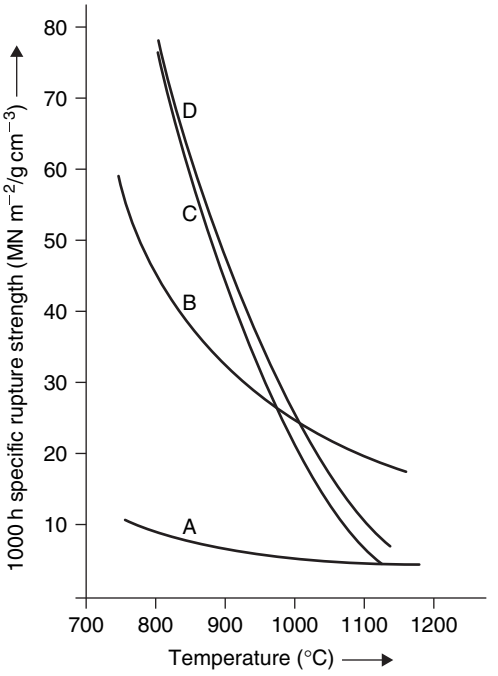


Fig. 5.1.10 Comparison of *MA6000* with other high-strength nickel alloys: (A) TD-nickel, (B) *MA6000*, (C) directionally solidified *Mar M200* and (D) single-crystal *PWA 454*.

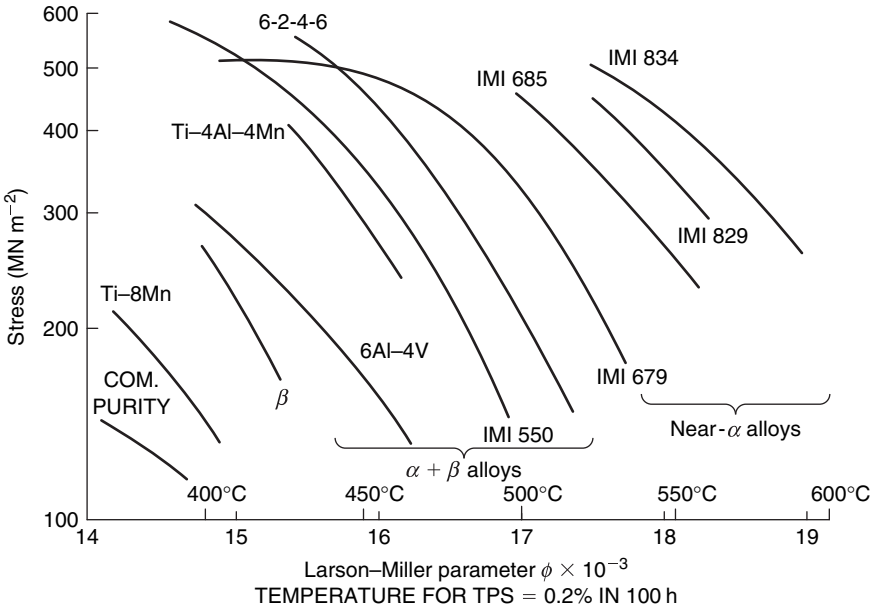


Fig. 5.1.11 Plot of stress versus Larson–Miller parameter ϕ for a range of titanium alloys.

5.1.5 Titanium alloys

5.1.5.1 Basic alloying and heat-treatment features

Since the emergence of titanium as a 'wonder metal' in the 1950s, the titanium industry has developed a wide range of alloys with different compositions (see Figure 5.1.11).[2] These alloys rely on the high strength/weight ratio, good resistance to corrosion, combined low thermal conductivity and thermal expansion of titanium, properties which make it attractive for aerospace applications in both engine and airframe components.

Titanium exists in the cph α form up to 882 °C and then as bcc β to its melting point. Alloying additions change the temperature at which the α to β transition takes place, solutes that raise the transus are termed α-stabilizers and those that lower the β-transus temperature are termed β-stabilizers (Figure 5.1.12). The predominant α-stabilizer is aluminum. It is also an effective α-strengthening element at ambient and elevated temperatures up to 550 °C, and thus a major constituent of most commercial alloys. The low density of aluminum is an important additional advantage. α-Phase strengthening is also achieved by additions of tin and zirconium. These metals exhibit extensive solubility in both α and β titanium but have little influence on the β-transus and are thus regarded as neutral additions. β-Stabilizers may be either β-isomorphous (i.e. have the bcc structure like β-Ti) or β-eutectoid elements. β-Isomorphous elements have a limited α-solubility and are completely soluble in β-titanium, typical additions being molybdenum, vanadium and niobium. In contrast, β-eutectoid elements have a restricted solubility in β-titanium and form intermetallic compounds by eutectoid decomposition of the β-phase. In some alloy systems containing β-eutectoid elements, such as silicon or copper, the compound formation (i.e. respectively Ti_5Si_3 and Ti_2Cu) leads to an improvement in mechanical properties. Titanium will also take interstitial solutes in solid solution, hydrogen being a β-stabilizer while carbon, nitrogen and oxygen are strong α-stabilizers. To minimize gas in Ti leads to a high cost of manufacture and heat treatment requires vacuum or inert gas conditions and freedom from refractory contact.

In describing titanium alloys it is conventional to classify them in terms of the microstructural phase (α alloys, β alloys, ($\alpha + \beta$) alloys or near-α alloys, i.e. predominantly α-phase but with a small volume of β-phase). Commercial alloys are usually heat treated to optimize the mechanical properties by controlling the transformation of the β to α phase, the extent of which is governed by the alloy composition and the cooling rate. The α alloys can transform completely from the β to α phase no matter what the cooling rate. Such treatments have a negligible effect on properties and α alloys tend to be used in the annealed state.

[2] The Larson–Miller parameter ϕ is given by $\phi = T(A + \log_{10}t)$ where T is the temperature in degrees kelvin, t the time in hours and A a constant, and defines the conditions to produce a given amount of plastic strain (e.g. 0.2%).

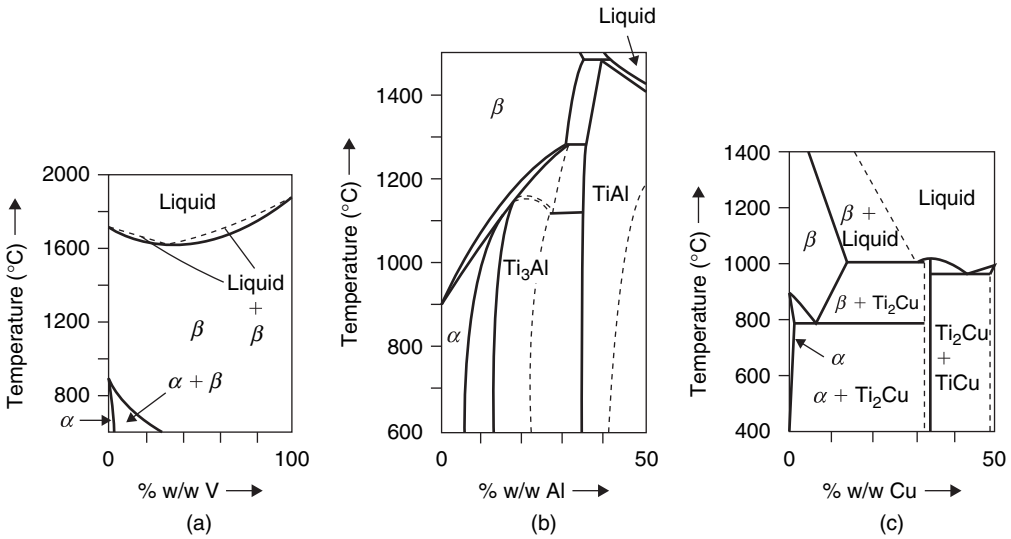


Fig. 5.1.12 Representative phase diagrams for Ti alloys, (a) Ti–V. (b) Ti–Al. (c) Ti–Cu (after E. A. Brandes and G. B. Brook, *Smithells Metals Reference Book*, 1998).

Rapidly cooled alloys containing β-stabilizers form martensitic α from the β-phase, whereas slower cooling rates favor α formation by a nucleation and growth process. Several morphologies of α can be produced by controlling the nucleation and growth mechanism; slow cooling, for example, tends to produce similarly aligned α platelets in colonies, combined with primary α at the grain boundaries. Faster cooling and higher α-stabilizer contents result in a basket-weave microstructure. Metastable-β, when aged, precipitates fine α, giving increased strength. The less stable the β, the more α can be precipitated and hence the higher the strength attained.

β-Stabilizing elements improve strength by strengthening the β-phase. The microstructure consists of primary α combined with the β-phase, which can be strengthened by an aging treatment to precipitate acicular-α. Further strengthening is achieved by the limited solubility of the β-stabilizing element. Generally, these alloys have poor ductility properties.

The most important alloys contain both α- and β-stabilizers which, after working and annealing, give good strength and fabrication properties. For good creep strength an α-titanium base, strengthened as much as possible by solute elements, is required. To meet this requirement the near-α alloys have been developed. These alloys combine the high α stability with sufficient β-stabilizer to give adequate strength. By β heat treating, the $(\alpha + \beta)$ microstructure changes to a totally transformed β structure containing basket-weave α. These alloys have good creep resistance and reasonable room-temperature properties. The basket-weave morphology is effective in inhibiting crack growth, and near-α alloys exhibit lower crack propagation rates than the $\alpha + \beta$ microstructures. Most of the commercial alloys which have been developed recently are of this type. The major factor influencing the post-forging microstructure is the cooling rate; an oil quench results in a basket-weave structure (see Figure 5.1.13a) and an air cool gives a typically aligned microstructure (see Figure 5.1.13b). The alloy is usually stress-relieved by annealing for 2 h at 625 °C or above.

5.1.5.2 Commercial titanium alloys

α Alloys transform entirely to α on cooling from the β-phase, regardless of cooling rate. Commercially pure titanium with a nominal oxygen content of 500, 1000, 1550 and 2700 ppm respectively gives tensile properties which range from 450 to 640 MN m^{-2} (see Figure 5.1.14). These are used mainly in sheet form. Solid-solution strengthening by aluminum, tin or zirconium increases tensile and creep strength and *IMI 317* (Ti–5Al–2.5Sn) is typical.

Increasing the α-stabilizing composition increases creep strength but makes fabrication more difficult and can lead to embrittlement during prolonged exposure at temperature in service due to the formation of the coherent ordered phase α_2 (Ti_3 Al). To avoid this it was empirically established that the aluminum equivalent Al* must be no greater than 9, where Al* (in wt%) is given by:

$$Al^* = Al + \frac{Sn}{3} + \frac{Zr}{6} + (10 \times O_2).$$

IMI 317 is difficult to fabricate and is often replaced by an *IMI 230*, which is an α-phase alloy containing the

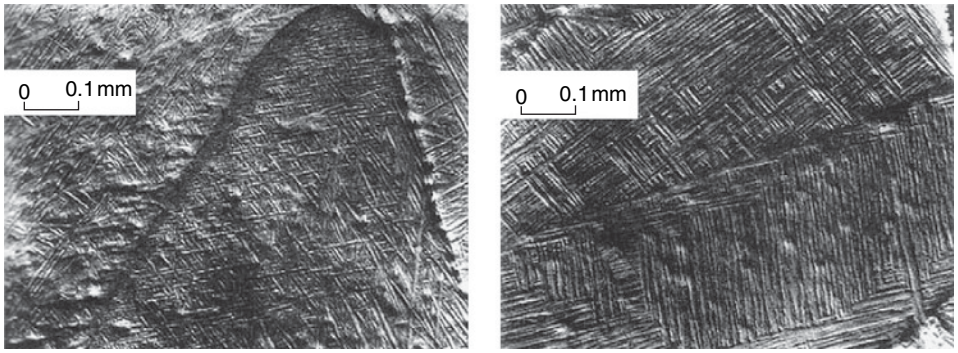


Fig. 5.1.13 Microstructure of near-α titanium alloy (*IMI 829*) initially ft heat-treated at 1050 °C for 1 h. (a) Oil quenched. (b) Air cooled (after Woodfield, Loretto and Smallman, 1988).

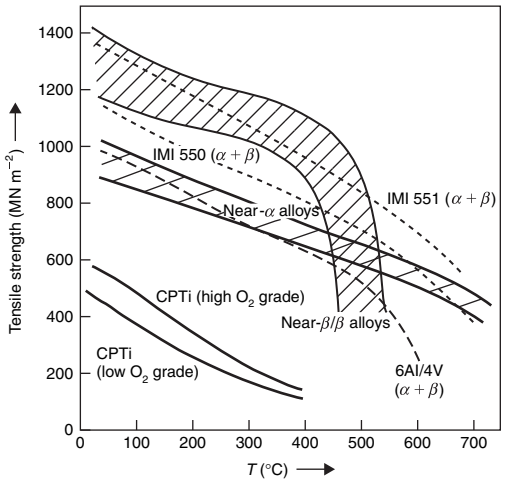


Fig. 5.1.14 Variation of tensile strength with temperature for a range of commercial titanium alloys.

precipitation-hardening phase Ti_2Cu; it can be fabricated and welded, and has good strength and ductility up to 400 °C.

β Alloys contain enough β-stabilizing elements to maintain the bcc β-phase to room temperature. Unfortunately, bcc β–Ti alloys are prone to embrittlement. The binary alloys, titanium with Fe, Cr, Mn, Nb, Mo, Cr or V, all precipitate the embrittling ω-phase. More complex β–Ti alloys containing Cr also suffer embrittlement from $TiCr_2$. More stable alloys have been developed (e.g. (Ti–11.5Mo–6Zr–45Sn)) but are little used. However, β-phase alloys, such as Ti–10V–2Fe–3Al, have potential as airframe construction materials offering high strength (1250 MN m^{-2}) and toughness (45 MN m$^{-3/2}$) in relatively thick cross-sections (90 mm) and superior hot-working characteristics which are attractive in expensive forging operations.

Highly stabilized β alloys have been developed for burn resistance. Ti alloys can burn when they rub together in gas turbine engines and steel stators are commonly used to separate the blades in compressors. Replacement of steel components by Ti alloys is a considerable weight saving. Pratt and Whitney developed a composition (wt%) Ti–35V–15Cr where the burn resistance is associated with the high V-content which produces a volatile V_2O_5 to shield the component from oxygen. A cheaper version with lower V-content has been developed by Rolls-Royce with composition Ti–25V–15Cr–2Al–0.2C. This is made from a low-cost V/Cr/Al master alloy readily available in the steel industry. Unfortunately, the introduction of Al results in a B2 structure which embrittles the alloy at the operating temperature of 550 °C when α is precipitated at the grain boundaries. Alpha is very much influenced by the presence of oxygen which is an α-stabilizer. The addition of carbon forms Ti(CO) carbides, which increases the oxygen concentration in the β-matrix and prevents oxygen segregation at the grain boundaries, with a corresponding reduction in grain boundary precipitation of α.

($\alpha + \beta$) alloys are probably the most widely used titanium alloys and contain alloying additions which strengthen both phases. These alloys are thermomechanically processed to control the size, shape and distribution of both α and β. The most versatile ($\alpha + \beta$) alloy is *IMI 318*, which contains 6% Al and 4% V; it can be used at temperatures up to 350 °C and has good forging and fabrication properties. It initially replaced steel as a disk material in jet engines, leading to 20% weight saving. Another important ($\alpha + \beta$) alloy is *IMI550* (Ti–6Al–2Sn–4Mo–0.5Si), which has higher strength and good creep resistance up to 400 °C. ($\alpha + \beta$) alloys remain the principal materials for fan disks and blades, and for low- and intermediate-pressure compressor disks and blades of current gas turbine engines, ($\alpha + \beta$) alloys with extra low interstitial (ELI) content are attractive as 'damage-tolerant' materials for critical airframe components. Ti–6Al–4V with low oxygen has a tensile strength

8% lower than the standard alloy but, more importantly, the minimum fracture toughness is 60 MN $m^{-3/2}$.

Near-α alloys have increased the strength and the volume of the more creep-resistant α-phase at the expense of the bcc β-phase, which imports good low-temperature strength and forgeability. *IMI 685* was the first titanium alloy to operate above 500 °C. It contains Ti–6Al–5Zr–0.5Mo–0.2Si with Al and Zr instead of Sn as α-stabilizers, reduced Mo, the β-stabilizer, to minimize β at the α-needles and Si to improve creep resistance. These alloys are worked and heat treated in the β-range, have a tensile strength of about 1000 MN m^{-2} and give less than 0.1% creep strain in 100 h under a stress of 310 MN m^{-2} at 520 °C. *IMI 829* (Ti–5.6Al–3.5Sn–3Zr–1Nb–0.25Mo–0.3Si) has been derived from *IMI 685* by replacing some of the Zr with the more potent strengthener Sn. It is β heat treated and has sufficient higher-temperature capability to be used in the hotter regions of engines. *IMI 834* (Ti–5.8Al–4Sn–3.5Zr–0.7Nb–0.5Mo–0.35Si–0.06C) has been developed for use up to 600 °C and combines the high fracture toughness and crack propagation resistance of a transformed β-structure with the typical equiaxed structure of the $\alpha + \beta$ alloys, providing good fatigue resistance and ductility. The small addition of carbon allows a controlled high α/β heat treatment. Hot working is carried out in the $\alpha + \beta$ field and heat treatment involves solution treatment for 2 h at 1025 °C, consistent with about 15% primary α, followed by oil quenching prior to aging for 2 h at 700 °C, then air cooling. With such good high-temperature properties the alloy is being specified for engine compressor applications.

5.1.5.3 Processing of titanium alloys

Some of the titanium alloys have excellent superplastic-forming characteristics and fabricating manufacturers have taken advantage of these properties in developing new processing technologies. Fine-grained thin sheets of *IMI 318* have been superplastically deformed at 900 °C under slow strain-rate conditions to produce a variety of complex parts. In conjunction with diffusion bonding, a weight saving of about one-third has been achieved in, for example, the wing access panels of the A320 Airbus. Near-α alloys have also been shown to exhibit superplastic behavior.

Casting technology has also been developed. Investment casting *of IMI 318* is widely used but there is now increasing interest in the use of high-temperature alloy castings as weight-saving alternatives to steel and nickel alloys above 500 °C. The introduction of HIPing for titanium castings has widened their potential application. For *IMI 829*, the best properties are obtained in the α/β HIP and solution-treated and aged condition, generally meeting the specification for the alloy in the wrought condition. Rapid solidification processing (RSP) may also offer possibilities by incorporating Er_2O_3 and other rare-earth oxide dispersions into the titanium matrix. At present, the control of these dispersoids remains a problem.

To develop higher-performance materials than the near-α alloys into the 700–800 °C range, attention is being given to strengthening high-temperature alloy materials, including titanium aluminides (see Section 5.1.6) with ceramic fibers such as SiC and B_4C.

5.1.6 Structural intermetallic compounds

5.1.6.1 General properties of intermetallic compounds

In terms of their properties, intermetallic compounds are generally regarded as a class of materials between metals and ceramics, which arises from the bonding being a mixture of metallic and covalent.

Intermetallics are intrinsically strong (and in the $L1_2$-ordered fcc compounds strength increases with temperature up to about 600 °C) with high elastic modulus. The strong bonding and ordered structure also gives rise to lower self-diffusion coefficients and hence greater stability of diffusion-controlled properties. Some of the compounds of current interest are shown in Table 5.1.5. Intermetallics containing aluminum or silicon exhibit a resistance to oxidation and corrosion because of their adherent surface oxides. Those based on light elements have attractive low density giving rise to high specific properties particularly important in weight-saving applications.

Like ceramics, however, the greatest disadvantage of intermetallics is their low ductility, particularly at low and intermediate temperatures. The reasons for the lack of ductility vary from compound to compound, but include (1) a limited number of easy deformation modes to satisfy the von Mises criterion, (2) operation of dislocations with large slip vectors, (3) restricted cross-slip, (4) difficulty of transmitting slip across grain boundaries, (5) intrinsic grain boundary weakness, (6) segregation of deleterious solutes to grain boundaries, (7) covalent bonding and high Peierls–Nabarro stress, and (8) environmental susceptibility. It has been demonstrated, however, that some intermetallics can be ductilized by small alloying additions: Ni_3Al with boron, TiAl with Mn, Ti_3Al with Nb. This observation has encouraged recent research and development of intermetallics, and the possibility of application of those materials.

Table 5.1.5 Comparison of physical properties of some intermetallic compounds

Compound	Crystal structure	Melting temp. (°C)	Density (kg m^{-3})	Young's modulus/density
Ni_3Al	$L1_2$ (ordered fcc)	1400	7500	45
Ni_3Si	$L1_2$ (ordered fcc)	1140	7300	
NiAl	B_2 (ordered bcc)	1640	5860	35
Ti_3Si	$D0_{19}$ (ordered cph)	1600	4200	50
TiAl	$L1_0$ (ordered tetragonal)	1460	3910	24
FeAl	B2 (ordered bcc)	1300	5560	47

5.1.6.2 Nickel aluminides

Ni_3Al (nickel aluminide) is the ordered fee γ'-phase and is a major strengthening component in superalloys. Ni_3Al single crystals are reasonably ductile but in polycrystalline form are quite brittle and fail by intergranular fracture at ambient temperatures. The basic slip system is $\{1\ 1\ 1\}\langle 1\ 1\ 0\rangle$ and has more than five independent slip modes, but still exhibits grain boundary brittleness. Remarkably, small additions of ~0.1 at.% boron produce elongations up to 50%. General explanations for this effect are that B segregates to grain boundaries, and (1) increases the cohesive strength of the boundary and (2) disorders the grain boundary region so that dislocation pile-up stresses can be relieved by slip across the boundary rather than by cracking. This general explanation is no doubt of significance but, additionally, there are distinct microstructural changes within the grains which must lead to a reduced friction stress and ease the operation of polyslip. For example, the addition of B reduces the occurrence of stacking-fault defects. Addition of solutes, such as B, are not expected to raise the stacking-fault energy and hence this effect possibly arises from the segregation of B to dislocations, preventing the superdislocation dissociation reactions.

Microhardness measurements inside grains and away from grain boundaries indeed show that boron softens the grains. The ductilization effect is limited to nickel-rich aluminides and cannot be produced by carbon or other elements, although some substitutional solutes such as Pd, which substitutes for Ni, and Cu produce a small improvement in elongation. Small additions of Fe, Mn and Hf have also been claimed to improve fabricability. Grain size has been shown to influence the yield stress according to the Hall–Petch equation and B appears to lower the slope k_y and facilitate slip across grain boundaries. These alloys are also known to be environmentally sensitive. Hf, for example, which does not segregate to grain boundaries but still improves ductility, has a large misfit (11%) and possibly traps H from environmental reactions, such as $Al + H_2O \rightarrow Al_2O_3 + H$. Ti, which has a small misfit, does not improve the ductility.

The most striking property of Ni_3Al is the increasing yield stress with increasing temperature up to the peak temperature of 600 °C (see Figure 5.1.15). This behavior is also observed in other $L1_2$ intermetallics, particularly Ni_3Si and Zr_3Al. This effect results from the thermally activated cross-slip of screw dislocations from the $\{1\ 1\ 1\}$ planes to the $\{1\ 0\ 0\}$ cube planes, where the APB energy is somewhat lower. The glide of superdislocations is made more difficult by the formation of Kear–Wilsdorf (K–W) locks and their frequency increases with temperature. Electron-microscopy measurements of APB energies given in Table 5.1.6 shows that the APB energy on $\{1\ 0\ 0\}$ decreases with aluminum content and this influences the composition dependence of the strength, shown in Figure 5.1.15. The cross-slip of screw

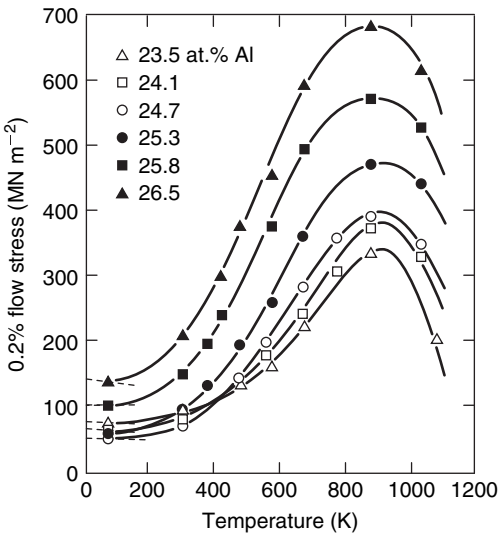


Fig. 5.1.15 Effect of aluminum content on the temperature dependence of the flow stress in Ni_3Al (after Noguchi, Oya and Suzuki, 1981).

Table 5.1.6 Anti-phase boundary energies in Ni_3Al

Alloy	γ_{111} (mJ m^{-2})	γ_{100} (mJ m^{-2})	$\gamma_{111}/\gamma_{100}$
Ni–23.5Al	183 ± 12	157 ± 8	1.17
Ni–24.5Al	179 ± 15	143 ± 7	1.25
Ni–25.5Al	175 ± 13	134 ± 8	1.31
Ni–26.5Al	178 ± 12	113 ± 10	1.51
Ni–23.5Al + 0.25B	170 ± 13	124 ± 8	1.37

dislocations from the {111} planes to cube planes also gives rise to a high work-hardening rate.

Although the study of creep in γ'-based materials is limited, it does appear to be inferior to that of superalloys. Above $0.6T_m$ creep displays the characteristic primary and secondary stages, with steady-state creep having a stress exponent of approximately 4 and an activation energy of around 400 kJ mol^{-1}, consistent with climb being the rate-controlling process. At intermediate temperatures (i.e. around the 600 °C peak in the yield stress curve) the creep behavior does not display the three typical stages. Instead, after primary creep, the rate continuously increases with creep strain, a feature known as inverse creep. In primary creep, planar dissociation leads to an initial high creep rate which slows as the screws dissociate on {1 0 0} planes to form K–W locks. However, it is the mobile edge dislocations which contribute most to the primary creep strain and their immobilization by climb dissociation which brings about the exhaustion of primary creep. The inverse creep regime is still not fully researched, but could well be caused by glide on the {1 0 0} planes of the cross-slipped screw components.

The fatigue life in high-cycle fatigue is related to the influence of temperature on the yield stress and is invariant with temperature up to about 800 °C, but falls off for higher temperatures with cracks propagating along slip planes. With boron doping the fatigue resistance is very sensitive to aluminum content and decreases substantially as Al increases from 24 to 26 at.%. Nevertheless, crack growth rates of Ni_3Al + B are lower than for commercial alloys.

Hyperstoichiometric Ni_3Al with boron can be prepared by either vacuum melting and casting or from powders by HIPing. Fabrication into sheets is possible with intermediate anneals at 1000 °C. At present, however, the application of Ni_3Al is not significant; Ni_3Al powders are used as bond coats to improve adherence of thermal spray coatings. Nevertheless, Ni_3Al alloys have been tested as heating elements, diesel engine components, glass-making molds and hot-forging dies, slurry-feed pumps in coal-fired boilers, hot-cutting wires and rubber extruders in the chemical industry. Ni_3Al-based alloys as matrix materials for composites are also being investigated.

Nickel aluminide (NiAl) has a cesium chloride or ordered β-brass structure and exists over a very wide range of composition either side of the stoichiometric 50/50 composition. It has a high melting point of 1600 °C and exhibits a good resistance to oxidation. Even with such favorable properties it has not been commercially exploited because of its unfavorable mechanical properties. Because it is strongly ordered, low-temperature deformation occurs by an $a\langle 1\ 0\ 0\rangle$ dislocation vector and not by $a/2\langle 1\ 1\ 1\rangle$ superdislocations. {1 1 0} ⟨1 0 0⟩ slip therefore leads to insufficient slip modes to satisfy the general plasticity criterion and in the polycrystalline condition β-NiAl is extremely brittle. The ductility does improve with increasing temperature but above 500 °C the strength drops off considerably as a result of extensive glide and climb. Improvements in properties are potentially possible by refinement of the grain size and by using alloying additions to promote ⟨111⟩ slip, as in Fe Al, which has the same structure. In this respect, additions of Fe, Cr or Mn appear to be of interest. For high-temperature applications, ternary additions of Nb and Ta have been shown to improve creep strength through the precipitation of second phases and mechanical alloying with yttria or alumina is also beneficial.

A further commercial problem of this material is that conventional production by casting and fabrication is difficult, but production through a powder route followed by either HIPing or hot extrusion is more promising.

5.1.6.3 Titanium aluminides

Because of the limited scope for improvements in the properties of conventional titanium alloys above 650 °C, either by alloy development or by TMP, increased attention is being given to the titanium intermetallics, Ti_3Al (α_2-phase) and TiAl (γ-phase). With low density, high modulus and good creep and oxidation resistance up to 900 °C, they have considerable potential if the poor ductility at ambient temperatures could be improved. A comparison of Ti_3Al- and TiAl-based materials with conventional Ti alloys is given in Table 5.1.7.

Electron microscopy studies of Ti_3Al or α_2 have shown that deformation by slip occurs at room temperature by coupled pairs of dislocations with $b = 1/6\langle 1\ 1\ \bar{2}\ 0\rangle$, which glide only on $\{1\ 0\ \bar{1}\ 0\}$ planes and by very limited glide on $\{1\ 1\ \bar{2}\ \bar{1}\}$ with pairs of dislocation with $b = 1/6\langle 1\ 1\ \bar{2}\ 6\rangle$. The ductility increases at higher temperatures due to climb of the $\langle 1\ 1\ \bar{2}\ 0\rangle$ dislocations and to the increased glide mobility of $1/6\langle 1\ 1\ \bar{2}\ 0\rangle$ and $1/6\langle 1\ 1\ \bar{2}\ 6\rangle$ dislocations through thermal activation.

Table 5.1.7 Comparison of α_2 and γ superalloys with conventional titanium alloys

Property	Titanium alloys	$(\alpha_2 + \beta)$	$(\gamma + \alpha_2)$
Density (g cm^{-3})	4.54	4.84	4.04
E, stiffness (GN m^{-2})	110	145	176
RT tensile strength (MN m^{-2})		1100	620
HT (760 °C) tensile strength (MN mm^{-2})		620	550
Max. creep temp. (°C)	540	730	900
RT ductility (%)	20	4–6	3
Service temp. ductility (%)	High	5–12	5–12

Only limited activity of the $\{0\,0\,0\,1\}\langle 1\,1\,\bar{2}\,0\rangle$ slip systems is observed, even at high temperatures.

The most successful improvements in the ductility of Ti_3Al have been produced by the addition of β-stabilizing elements, particularly niobium, to produce α_2 alloys. An addition of 4 at.% Nb produces significant slip on $\{1\,0\,\bar{1}\,0\}\langle 1\,1\,\bar{2}\,0\rangle$, $\{0\,0\,0\,1\}\langle 1\,1\,\bar{2}\,0\rangle$ and $\{1\,1\,\bar{2}\,1\}\langle 1\,1\,\bar{2}\,6\rangle$ as well as some slip on $\{1\,0\,\bar{1}\,1\}\langle 1\,1\,\bar{2}\,0\rangle$. This improvement is attributed to the decrease in covalency as Nb substitutes for Ti with a consequent reduction in the Peierls–Nabarro friction stress. Alloys based on α_2 are Ti–(23–25)Al–(8–18)Nb, of which Ti–24Al–11Nb has excellent spalling resistance. Most Ti_3Al + Nb alloys, such as super α_2, also contain other β-stabilizers including Mo and V, i.e. Ti–25Al–10Nb–3V–lMo, which exhibits about 7% room temperature elongation. Alloying Ti_3Al with β-stabilizing elements to produce two-phase alloys significantly increases the fracture strength. These Ti_3Al-based alloys can be plasma-melted and cast followed by TMP in the $(\alpha_2 + \beta)$ or β-range. The improved ductility of Ti_3Al alloys has led to aerospace applications in afterburners in jet engines, where it compares favorably in performance with superalloys and gives a 40% weight saving.

Developments are taking place in RSP to include a second phase (e.g. rare-earth precipitates) and to provide powders, which may be consolidated by HIPing, to produce fully dense components with properties comparable to wrought products. There are also developments in intermetallic matrix composites by the addition of SiC or Al_2O_3 fibers ($\sim$10 μm). These have some attractive properties, but the fiber/intermetallic interface is still a problem.

The γ-phase Ti–(50–56)Al has an ordered fc tetragonal ($L1_0$) structure up to the m.p. 1460 °C, with $c/a = 1.02$ (Figure 5.1.16). Deformation by slip occurs on {1 1 1} planes and, because of the tetragonality, there are two types of dislocations, namely ordinary dislocations $1/2\langle 1\,1\,0\rangle$ and superdislocations $\langle 0\,1\,1\rangle = 1/2\langle 0\,1\,1\rangle + 1/2\langle 0\,1\,1\rangle$. Another superdislocation $1/2\langle 1\,1\,2\rangle$ has also been reported.

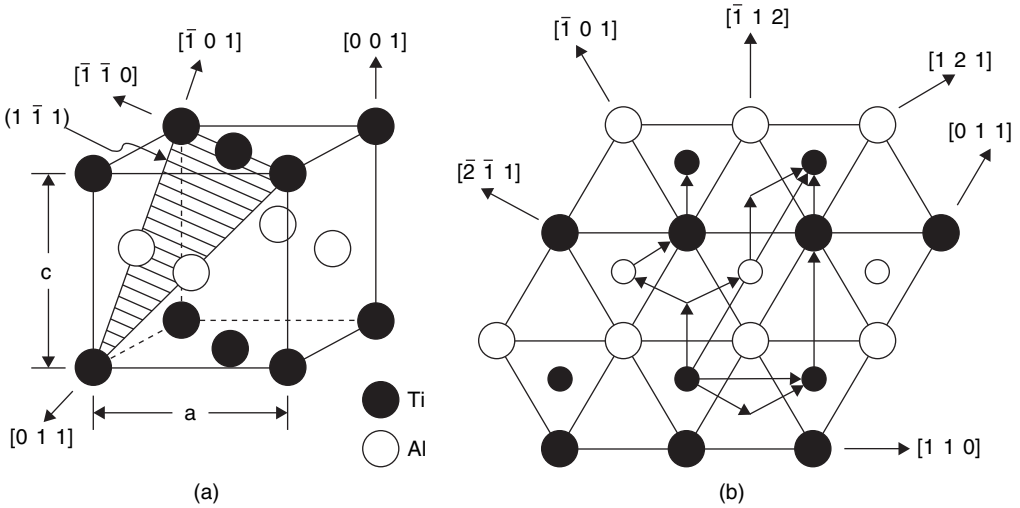


Fig. 5.1.16 Structure of TiAl ($L1_0$) (a) and (1 1 1) plane (b) showing slip vectors for possible dissociation reactions, e.g. ordinary dislocations 1/2[1 1 0], superdislocations [0 1 1] and 1/2[1 1 2], and twin dislocations 1/6[1 1 2] (after Kim and Froes, 1990).

At room temperature, deformation occurs by both ordinary and superdislocations. However, [0 1 1] and [1 0 1] superdislocations are largely immobile because segments of the trailing superpartials 1/6⟨1 1 2⟩ type form faulted dipoles. The dissociated 1/2⟨1 1 0⟩ dislocations bounding complex stacking faults are largely sessile because of the Peierls–Nabarro stress. Some limited twinning also occurs. The flow stress increases with increasing temperature up to 600 °C as the superpartials become mobile and cross-slip from {1 1 1} to {1 0 0} to form K–W-type locks, the 1/2⟨1 1 0⟩ slip activity increases and twinning is promoted.

The two-phase ($\gamma + \alpha_2$) Ti–Al alloys have better ductility than single-phase γ with a maximum at 48 at.% Al. This improvement has been attributed to the reduced c/a with decreased Al, further promotion of twinning and the scavenging of O_2 and N_2 interstitials by α_2. The combination of high stiffness ($E = 175$ GPa at 20 °C to 150 GPa at 700 °C), density-normalized strength similar to cast Ni-based alloys, high temperature strength and reasonable oxidation resistance to 750 °C, low thermal expansion coefficient and high thermal conductivity have led to a high level of interest in TiAl-based alloys. The major limitations to their application are the intrinsic low room temperature ductility (no better than 2–3%), the low fracture toughness (between 10 and 20 MPa $m^{1/2}$ at 20 °C) and the high growth rate of fatigue cracks.

Alloy and process development have resulted in some successful applications of these alloys over the last 10 or so years. Cast turbochargers are now manufactured in Japan for cars and wrought exhaust valves were used in Formula 1 cars for some years. Major applications for these alloys are still awaited despite the success of these two applications; the high cost of processing is holding commercial developments back. In the case of TMP the costs are high because the alloys are strong at normal hot-working temperatures and because some sort of protection (such as canning with steel) from oxidation must be used during working. In the case of castings the efficiency of material usage is very low, both because casting technology is not efficient and because melting and casting are difficult because of the reactivity of molten TiAl-based alloys.

The compositions of TiAl-based alloys which are of commercial interest lie within the range of about Ti/45–8 Al (at.%) but all alloys contain other elements in attempts to improve the properties of the binary alloy. Additions of Nb between about 5 and 8 at.% are important in improving oxidation resistance and also imparting some solid solution strengthening.

An understanding of the microstructures which can be obtained in cast or in wrought products of TiAl-based alloys requires knowledge of the phase changes that occur over the temperature range from the melting point to room temperature. The relevant part of the binary phase diagram between Ti and Al is shown in Figure 5.1.17, which also indicates schematically the influence of some alloying additions on phase boundaries.

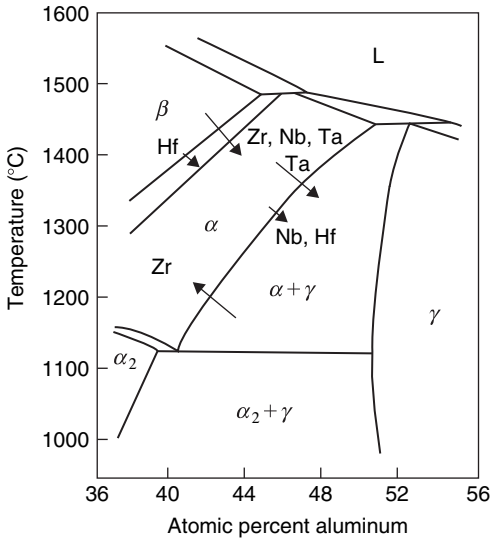


Fig. 5.1.17 Partial Ti–Al phase diagram showing the influence of ternary additions on the position of the various boundaries (courtesy M. H. Loretto).

The various phase transformations in the Ti–Al system offer the possibility of microstructural control both for the wrought route and for the casting route. Thus, cooling samples containing less than about 44 at.% Al the solidification will take place through the formation of β, which may or may not be removed via the peritectic reaction. Subsequent cooling of the α-phase results in precipitation of γ which, under typical cooling rates encountered with castings, results in the formation of a lamellar structure. This 'fully lamellar' structure consists of parallel lamellae of γ and α and of twinned γ. These lamellae are formed on the (0 0 0 1) plane of the α-phase and thus their length is defined by the pre-existing α grain size. Somewhat slower cooling results in some of the γ lamellae coarsening at colony boundaries to form γ grains through local growth of the lamellae, to form a 'near fully lamellar' structure. Hot working in the two-phase region results in the formation of equiaxed γ and α grains; the ratio of the amounts being defined by the average alloy composition and the hot-working temperature. Subsequent cooling results in the equiaxed α grains forming lamellae to yield a duplex microstructure, or if extensive hot working is used, a structure consisting of γ and α grains is formed, termed 'near γ'.

If the cooling rate is increased, as in oil or water quenching, the α-phase transforms massively to γ if the Al content is above about 44 at.% (below this Al-content, α is retained) and this transformation offers a further opportunity for microstructural control in cast samples by heat treating in the two-phase field so that α can precipitate on all four {1 1 1} planes, throughout the γ grains. The tendency to transform massively is strongly dependent upon

the composition, which has important consequences upon the choice of alloy composition in cast samples.

5.1.6.4 Other intermetallic compounds

A number of intermetallic compounds are already used in areas which do not rely on stringent mechanical properties. Fe_3Al, for example, is used in fossil-fuel plants, where resistance to both sulfur attack and oxidation is important. Ni_3Si is used where resistance to hot sulfuric acid is required. There are several compounds with rare-earth elements used in magnet technology. PdIn is gold colored and a possible dental material. Zr_3Al has a low neutron capture cross-section and is a possible reactor material.

The β-compound NiTi (*Nitinol*) is an important shape-memory alloy. The shape-memory effect (SME) manifests itself when the alloy is deformed into a shape while in a low-temperature martensitic condition but regains its original shape when the stress is removed and it is heated above the martensitic regime. Strains of the order of 8% can be completely recovered by the reverse transformation of the deformed martensitic phase to the higher-temperature parent phase. The martensite transformation in these alloys is a thermoelastic martensitic transformation in which the martensite plates form and grow continuously as the temperature is lowered and are removed reversibly as the temperature is raised. NiTi was one of the original SME alloys, but there are many copper-based alloys which undergo a martensitic transformation, e.g. Cu–17Zr–7Al. Application of SME alloys relies on the characteristic that they can change shape repeatedly as a result of heating and cooling and exert a force as the shape changes. By composition control (increasing the Ni content or substitution of Cu lowers the M_s temperature of TiNi), the shape memory can be triggered by normal body temperature or any other convenient temperature to operate a device. Several biomedical applications have been developed in orthopedic devices (e.g. pulling fractures together), in orthodontics, in intrauterine contraceptives and in artificial hearts. Industrial applications include pipe couplings for ships which shrink during heating, electrical connectors, servo-mechanisms for driving recording pens, switches, actuators and thermostats.

5.1.7 Aluminum alloys

5.1.7.1 Designation of aluminum alloys

Aluminum has an attractive combination of properties (i.e. low density, strong, easy to fabricate) which can be developed and modified by alloying and processing. In this section some of the alloys developed for particular industries such as the transportation, construction, electrical and packaging industries will be discussed.

Table 5.1.8 Aluminum alloy designation systems

Wrought alloys	Designation	Casting alloys	Designation
99.00% (min.) aluminum	1XXX	99.00% (min.) aluminum	1XX.X
Copper	2XXX	Copper	2XX.X
Manganese	3XXX	Silicon with added copper and/or magnesium	3XX.X
Silicon	4XXX	Silicon	4XX.X
Magnesium	5XXX	Magnesium	5XX.X
Magnesium and silicon	6XXX	Zinc	6XX.X
Zinc	7XXX	Tin	7XX.X
Others	8XXX	Others	8XX.X

Aluminum alloys are identified by a four-digit system[3] based on the main alloying element. This is summarized in Table 5.1.8. For wrought alloys the first digit identifies the alloy group and the second digit any modification to the original alloy, which is identified by the last two digits. The system is slightly different for casting alloys. The first digit again identifies the group, the second two digits identify the alloy and the last digit, preceded by a decimal point, indicates the product form (i.e. 0 for casting and 1 for ingot).

5.1.7.2 Applications of aluminum alloys

With the need for fuel economy and weight saving, aluminum alloys are increasingly used in cars, and its two most important properties are density and thermal conductivity. Over the past 15 years the aluminum content of cars has increased from around 5% to 13% by both volume and weight. In engines they are used for pistons, cylinder heads and sumps. Al–Si casting alloys of the 3XX.X series are being used for engine blocks and Al–Si pistons with cast iron cylinder liners for wear resistance. The superior thermal conductivity reduces the volume of coolant in the system. Aluminum wheels, vacuum cast or forged, are replacing conventional steel

[3] The International Alloy Designation System (IADS) was first introduced by the Aluminum Association of the USA and is now standard for wrought alloys.

wheels in sports cars. Heat-treatable[4] 2XXX and 6XXX as well as 5XXX series can be used for body sheet. However, the modulus of aluminum is only one-third that of steel and hence significant design changes are necessary to maintain rigidity and stiffness. A straightforward increase in gauge thickness would lead to a doubling cost, which limits the replacement to 'quality' cars. For bulk market cars, gauge-for-gauge substitution for steel is a future objective with structural reinforcement to enhance body torsional characteristics. This is possible with the use of adhesives in a weld-bonding approach which can reduce the weight by half, and fuel consumption and CO_2 emission by almost 15%. Aluminum-structured vehicle technology is likely to be essential when emission control to reduce global warming effects is tightened.

The Honda NSX all-aluminum car[5] is manufactured with conventional design and assembly, and has three different 6000 series alloys for external panels and a structural subframe of 5182 alloy.

In aircraft construction use is made of the high strength-density ratio of the Al–Cu (2000 series) and Al–Zn–Mg (7000 series) alloys in extruded form for wing spars, fuselage and landing gear, and for the skin in plate or sheet form; typically 7075 (5.6Zn–2.5Mg–l.6Cu) is used in the T6 condition and 2024 (4.4Cu–1.5Mg–0.6Mn) in the T3 or T8 conditions. The alloys of the 7000 series have higher strength than the 2000 alloys but lower resistance to fracture. However, higher purity levels (e.g. 2124 alloy) give enhanced toughness. Alloys of both series lose strength above 100 °C and are thus not suitable for supersonic aircraft. The *RR58* alloy used for the Concorde at temperatures up to 175 °C was originally an early engine material, for compressor blades and impellers.

Aluminum is used extensively in the construction industry because of its light weight, resistance to atmospheric attack and surface finish. For decorative applications, dyed anodic films produce a permanent durable finish. Generally, the Al–Mg–Si 6000 series is used, i.e. 6063 medium-strength and 6082 higher-strength alloys in the T6 condition for extrusions or T4 where forming is required during fabrication. The Al–Cu (2014A) alloy is also used for heavily loaded primary structures.

The packaging industry also provides a large market for aluminum alloys. The main requirement is for low-cost, simple alloys capable of being formed, and the Al–Mg (5000) series is often used. Impurity control is essential and liquid metal filtering is necessary in the production of thin sheet.

In the electrical industry, electrical conductivity grade aluminum and higher-strength 5000 and 6000 series alloys are used for transmission lines, replacing the more expensive copper. Dispersion-strengthened alloys containing a fine dislocation substructure stabilized by small precipitates are used for electrical wiring.

5.1.7.3 Aluminum–lithium alloys

The advantages of aluminum–lithium alloys have been known for a long time, but lower density and increased elastic stiffness were offset by poor ductility and fracture performance. Basic Al–Li alloys precipitate the $(Al_3Li)\delta'$, a spherical ordered precipitate. Precipitation hardening leads, however, to localized deformation with limited cross-slip and poor fracture behavior. Additions of copper to the alloy so that the Li/Cu ratio is high leads to the formation of both δ' and a T_1 -phase (Al_2CuLi). This gives some improvement in fracture toughness by independent control of the two precipitates. In the quaternary system Al–Li–Cu–Mg the S-phase precipitates in addition to the δ' and T_1. The S-phase is better at dispersing slip than T_1 and with adjustment of composition can be made to dominate the structure. Both S and T_1 are nucleated heterogeneously on dislocations and the best results are obtained by cold working the alloy after solution treatment.

Commercial alloys based on this background are *Lital A*, *B* and C, which have been developed to match the (1) conventional medium-strength 2014-T6, (2) high-strength 7075-T6 and (3) damage-tolerant 2024-T3 alloys, with a 10% reduction in density and 10% improvement in stiffness (see Table 5.1.9).

Lital A in T6 sheet form typically has 365 MN m^{-2} 0.2% proof stress, 465 MN m^{-2} TS, 6% elongation, 66 MN $m^{-3/2}$ fracture toughness, an elastic modulus of 80 GN m^{-2} and density of 2550 kg m^{-3}. *Lital B* has roughly 10% improvement in strength. *Lital* C is a variant of the 8090 alloy and is heat treated to increase toughness ($\sim$76 MN $m^{-3/2}$) at the expense of strength (TS $\sim$440 MN m^{-2}).

Lithium additions are also being made to conventional aluminum alloys. The addition of lithium has a major influence since Li possesses a significant vacancy binding energy of about 0.25 eV Lithium atoms therefore trap vacancies and form Li–V aggregates. This decreases the concentration of mobile vacancies available for the

[4] Alloy treatment is usually described by a suffix letter and digit system (e.g. F – as fabricated, O – annealed, H – work-hardening and T – heat treated). Digits following H specify the work-hardened condition, and that following T the type of aging treatment (e.g. T6 is solution heat treated and artificially aged, T4 solution heat treated and naturally aged, T3 solution heat treated and cold worked).
[5] Launched in the USA in 1992.

Table 5.1.9 Composition of commercial aluminum–lithium alloys

Alloy	Li	Cu	Mg	Fe	Si	Zr
(8090) Lital A	2.5	1.3	0.7	≤0.2	≤0.1	0.12
(9091) Lital B	2.6	1.9	0.9	≤0.2	≤0.1	0.12

transport of zone-forming atoms, and therefore inhibits the diffusion of Zn and Mg in 7075, and Si and Mg in 6061, into zones. Second, the Li–V aggregates, very probably present during quenching and immediately after aging, act as heterogeneous sites for subsequent clustering of zone-forming atoms during aging.

Additions of Li into either Al–2Mg–0.6Si–0.3Cu–0.3Cr (6061) or Al–5.9Zn–2.4Mg–l.7Cu (7075) modify the precipitation scheme and age-hardening behavior of the original alloys. The precipitates which form in the base alloys are inhibited or even suppressed. For the 6061 the addition of 0.7% Li retards the precipitation of needle-shaped GP zones and produces a ternary compound AlLiSi, whereas the addition of 2.0% Li results in the dominant precipitation of δ' and extremely delayed and limited formation of needle-shaped GP zones and AlLiSi. For 7075 the addition of 0.7% Li alters the conventional precipitation scheme from solute-rich GP zone $\rightarrow \eta' \rightarrow \eta MgZn_2$ into vacancy-rich GP zone $\rightarrow$ T′ $\rightarrow T(AlZn)_{49}Mg_{32}$, whereas the addition of 2.0% Li produces the dominant δ' precipitate and limited and delayed formation of T-phase. As a result, the age-hardening response relating to these major hardening phases in both base alloys is delayed or decreased. Such additions can produce narrower PFZs and give improved fracture properties.

A further commercial alloy is *UL*40, which is essentially a binary alloy containing 4% Li. The alloy is cast using a spray-deposition process resulting in a fine-grained microstructure, with uniform distribution of second phase, free from oxide. The high Li content alloy has a very low density (2400 kg m^{-3}) and is almost a third lighter than conventional aluminum and magnesium alloys. It extrudes well and can be welded with Al–Mg–Zr filler, producing components for aircraft and helicopters, such as pump housings and valves, and for yachting with good corrosion resistance.

5.1.7.4 Processing developments

5.1.7.4.1 Superplastic aluminum alloys

Superplastic forming is a cost-effective manufacturing process for producing both simple and complex shapes from aluminum alloy sheet, because of its low-cost tooling and short lead times for production. A range of alloys is available, including 2004 (Al–6Cu–0.4Zr) or *Supral*, 5083 SPF, 7475 SPF and *Lital* 8090 SPF (Al–Li–Mg–Cu). *Supral* and *Lital* dynamically recrystallize to a fine grain size during the early stages of deformation (~500 °C), which is stabilized with $ZrAl_3$ particles. The grain size in 7475 is stabilized by TMP and submicron Cr particles and, in 5083, by an Mn-dispersoid. Components are formed by clamping the alloy sheet in a pressure chamber and then applying gas or air pressure to force the sheet slowly into contact with a tool surface; both the tool and sheet are maintained at the forming temperature throughout the process. During normal superplastic forming, the alloys tend to develop voiding. This void formation is minimized by forming in hydrostatic conditions by introducing a gas pressure on both the front and back surfaces of the sheet being formed. The sheet is forced against the tool surface with a small pressure differential.

5.1.7.4.2 RSP of aluminum alloys

RSP has been applied to aluminum alloys to produce a fine grain size and extend solid solubility, particularly for the transition metals iron, molybdenum, chromium and zirconium, which usually have low solid solubility and low diffusion rates in aluminum. Interestingly, RSP alloys containing Fe and Cr, on annealing, precipitate metastable spherical quasi-crystals of icosahedral phase with fivefold symmetry. These are extremely stable and hardly coarsen after extensive heat treatment, which indicates a potential for alloy development.

A series of commercially available high-temperature Al–Fe–V–Si alloys has been developed and consist of very fine, spherical $Al_{13}(FeV)_3Si$ silicides uniformly dispersed throughout the matrix, which display much slower coarsening rates than other dispersoids. A typical alloy with 27 vol.% silicides is 8009 (Al–8.5Fe–1.3V–l.7Si) and, without any needle or platelet precipitates in the microstructure, has a $K_{1c} \sim 29$ MN $m^{-3/2}$. The tensile properties as a function of temperature are shown in Figure 5.1.18 in comparison with a conventional 7075-T6 alloy. At all temperatures up to 480 °C, the 8009 alloy has a higher specific stiffness than a Ti–6Al–V alloy. The fatigue and creep rupture properties are better than conventional aluminum alloys with excellent corrosion resistance. These RSP silicide alloys can be readily fabricated into sheet, extruded or forged, and the combination of attractive properties makes them serious candidates for aerospace applications. Other alloys developed include Al–Cr–Zn–Mn, Al–8Fe–2Mo and Al–Li.

5.1.7.4.3 Mechanical alloys of aluminum

Mechanical alloys of aluminum contain dispersions of carbides or oxides, which not only produce dispersion strengthening but also stabilize a fine-grained structure. An advantage of these alloys arises because the strength is derived from the dispersoids and thus the composition of

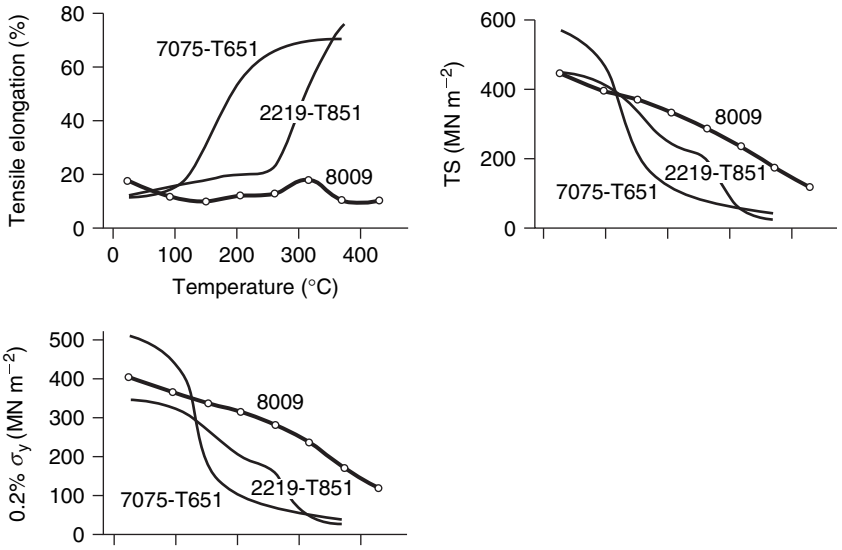


Fig. 5.1.18 Tensile properties of RSP Al alloy 8009 as a function of temperature compared with conventional aluminium alloys (from Gilman, 1990; courtesy of Institute of Materials, Minerals and Mining).

the alloy matrix can be designed principally for corrosion resistance and toughness rather than strength. Thus, the alloying elements which are usually added to conventional aluminum alloys for precipitation strengthening and grain-size control may be unnecessary.

Mechanical alloying is carried out with elemental powders and an organic process control agent, such as stearic acid, to balance the cold-welding and powder-fracture processes. No dispersoid is added because the oxide on the surface of the powders and process control agent are consolidated during mechanical alloying as hydrated oxides and carbonates. The process produces a fine dispersion of ~20 nm particles in a dynamically recrystallized structure with grains as fine as 0.05 μm. Subsequent vacuum degassing at elevated temperature removes the H_2 and N_2 liberated, improves the homogenization of the matrix and reduces carbonates to Al_4C_3, which forms most of the dispersoid. The final grain size is around 0.1 μm. The powder is then compacted by HIPing or vacuum hot-pressing and conventionally extruded to produce a material with a stable grain size of 0.3 μm, with grain boundaries pinned by the dispersoid.

Mechanical alloys have been developed corresponding to the 2000, 5000 and 7000 aluminum series alloys. *IN 9021* is heat treatable by solution treatment and natural or elevated temperature aging to give 500–560 MN m^{-2} proof stress, 570–600 MN m^{-2} TS, 12% elongation and 40 MN m$^{-3/2}$ K_{1c}. *IN 9052* is the equivalent of a 5000 series alloy, requiring no heat treatment and offering good strength in thick sections, 390 MN m^{-2} proof stress, 470 MN m^{-2} TS, 13% elongation and 46 MN m$^{-3/2}$ K_{1c}. Mechanically alloyed Al–Mg–Li offers inherent high strength in thick section, 430 MN m^{-2} proof stress, 500 MN m^{-2} TS, 10% elongation and 30 MN m$^{-3/2}$ K_{1c}.

References

Baker, C. (ed.) (1986). *Proc. 3rd International Aluminium – Lithium Conference*. Institute of Metals, London.

Bhadeshia, H. K. D. H. (1992). *Bainite in Steels*. Institute of Materials, London.

Honeycombe, R. W. K. (1981). *Steels, Micro-structure and Properties*. Edward Arnold, London.

Janowak, J. R. *et al.* (1984). A review of austempered ductile iron metallurgy. *First International Conference on ADI*.

Kim, Y.-W. (1989). Physical metallurgy of titanium aluminides. *TMS/ASM Symposium on High Temperature Aluminides and Intermetallics*.

Materials in defence (1988). *Metals and Materials*, 4(7). Institute of Materials, London.

Meetham, G. W. (1981). *The Development of Gas Turbine Materials*. Applied Science, London.

Peters, H. (ed.) (1991). *Proc. 6th International Aluminium–Lithium Conference*. Deutsche Gesellschaft für Material-kunde.

Polmear, I. J. (1989). *Light Alloys*. Edward Arnold, London.

Sims, C. T. and Hagel, W. C. (eds) (1972). *Superalloys*. John Wiley, Chichester.

Sims, C. T., Hagel, W. C. and Stoloff, N. S. (1987). *Superalloys II*. John Wiley, Chichester.

Stoloff, N. S. (ed.) (1984). Ordered alloys. *International Metal Reviews*, 29(3).

Stoloff, N. S. (1989). Physical and mechanical metallurgy of Ni_3Al and its alloys. *International Metal Reviews*, 34(4).

Yoo, M. H. *et al.* (1993). Deformation and fracture of intermetallics. Overview No. 15. *Acta Metall. Mater.*, No. 4.

Section **Six**

Design and ergonomics

This page is intentionally left blank

Chapter 6.1

6.1

Personal factors

Dennis Attwood, Joseph Deeb and Mary Danz-Reece

6.1.1 Introduction

In the human/system interface (HSI) model, the operator continuously receives information from the environment and processes it. There may or may not be follow-up action. If a response is required, then the appropriate information is transmitted back to the system in the form of manual manipulation of controls (e.g., valves) and buttons, lifting bales of rubber, climbing ladders and stairs, and so forth. This, in turn, changes the status of the system. The operator receives new information from the displays about the status of the system and, again, may or may not take action to bring the system to the desired level or maintain its status. Furthermore, the information the operator receives does not come only from displays of machines. As an example, when using a hand tool such as a saw, the external status of the line to be cut is perceived by the operator and continuous corrections are made to end up with a straight cut. Thus, the word *display* has a more general meaning than just a temperature or pressure gauge.

This HSI model functions as a closed-loop system. This means that, as long as we interact with something (i.e., machine, tools, other humans, or even animals), we continuously receive information, process information, look for status changes, receive new information, and so on. The first set of limitations that can affect the efficient working of this HSI closed-loop system is the sensory, perceptual, and cognitive capabilities of the operator. The second set of limitations is the physical capabilities that allow the body to move in the desired way. This chapter discusses both sets of limitations.

At the end of this chapter is a case study related to examining human force capabilities when interacting with valves at different positions. In addition, a set of review questions are included to help check your understanding of the material covered in this chapter.

6.1.2 Sensory and cognitive capabilities

Among the human sensory mechanisms are the five senses (vision, hearing, touch, smell, and taste). These senses deal with stimuli external to the body and are referred to as the *exteroceptors*. In addition, we have senses that deal with the position of our body in space and assist us in jumping, climbing stairs and ladders, reaching, and the like. These are referred to as *proprioceptors* and are embedded in the muscles, tendons, joints, and inner ear. This section covers the exteroceptors, specifically vision and hearing, for two reasons:

1. The majority of information we need to perform tasks is received through the visual and hearing senses, such as written procedures, warning labels and signs, control position, dials, warning lights, and alarms.
2. The majority of human factors research has been conducted in the visual and auditory senses.

6.1.2.1 Visual sense

Vision has been the most studied and researched sense. In addition, given current technologies, vision is probably the most overloaded sense at work. Our visual system receives energy from the environment around us and converts the energy into something meaningful for

Ergonomic Solutions for the Process Industries; ISBN: 9780750677042
Copyright © 2003 Elsevier Inc. All rights of reproduction, in any form, reserved.

us to understand. Our visual system works on two levels: physiological, includes the eyes and brain, and psychological, involves both the immediate visual sensation (energy received) and our interpretation of it based on our knowledge and experience (the perception process).

In general, we receive visual information either directly (direct observation of a coworker similar to the observation process of a behavior-based safety program) or indirectly (seeing things through a computer, television, radar, or the like).

The most important part of the eye is the retina, which covers the back of the eye. It converts energy from the environment (i.e., light) into electrical energy and passes it to the brain through optic nerves. The retina is made up mostly of cones and rods (so-called because of their shape). Cones function at high illumination levels and are responsible for daylight vision. Cones are also important in visual acuity (ability of the eye to see details) and differentiate extremely well among colors. Rods, on the other hand, respond to low illumination levels, such as at night. Rods also differentiate between shades of black and white.

Because of the characteristics of the cones and rods, the eyes function well over a wide range of illumination levels. However, several factors determine visual capabilities and effectiveness, and have important implications in designing visual displays. These factors are discussed next.

6.1.2.1.1 Accommodation of the eye

Accommodation refers to changing the shape of the lens in the eye to properly focus the image on the retina. In other words, accommodation is the ability of the eye to focus on objects at different distances. For example, the ability to read signs on the side of the road is an example of far vision. The ability of a control room operator to monitor displays is an example of near vision.

In normal accommodation, the lens of the eye flattens to focus a far object on the retina and bulges to focus a near object. The eyes may not accommodate an object for two reasons:

- *Nearsightedness*. Individuals cannot focus on far objects. The lens tends to remain in a bulged condition and the light from the far object is focused in front of the retina. Nearsightedness is also referred to as myopia.
- *Farsightedness*. Individuals cannot focus on near objects. The lens tends to remain in a flattened condition and the light from the near object is focused in the back of the retina.

With aging, there is a reduction in the ability of the eyes to accommodate. This is related to the gradual loss of the elasticity of the lens. This reduction is measured by the closest distance (near point) at which the eyes can focus. The table below lists average near point distances against age.

Age	Near point in cm (in.)
10	7 (2.8)
15	8 (3.2)
30	11 (4.4)
45	25 (10)
50	50 (20)
60	85 (34)
70	100 (40)

6.1.2.1.2 Visual field

Visual field is defined as "what can be seen when head and eyes are kept fixed (motionless)." We divide the visual field roughly into three areas (Grandjean, 1988):

1. *Optimal field* of view with an angle of view of 1° where objects are sharp and seen clearly with details.
2. *Middle field* with a visual angle ranging between 1° and 40°, where objects are not sharp; however, movements and strong contrasts are noticed.
3. *Outer field* with a visual angle ranging between 40° and 70° where objects are not sharp or clear. Objects must move to be noticeable.

It is important that the objects we need to see clearly are located in the optimal field area.

6.1.2.1.3 Adaptation process

Adaptation is the process or ability of the eyes to adjust to different light levels. We need time to adapt to darkness or brightness before we can discriminate details. For example, when we first move from bright sunlight to a dark movie theater, we have difficulty finding our way around. After a while, we begin to see distinct objects and faces. This process is called *dark adaptation*. Dark adaptation is the process of transitioning from cone to rod vision. The process can take from 30 minutes to an hour depending on the initial illumination levels. Conversely, moving from darkness to bright sunlight, we experience high glare and find it difficult to see details. The process of transitioning from dark to light is called light adaptation. Light adaptation is the process of transitioning from rod to cone vision. Light adaptation is completed relatively fast, a minute or two. The process of adaptation has practical work implications:

1. Allows sufficient time for the eyes to adapt when performing tasks at low levels of lighting.

2. Avoids excessive changes in illumination levels, since the constant need for adaptation results in visual fatigue.
3. Avoids excessive differences in brightness in all important surfaces within the visual filed.

6.1.2.1.4 Color vision

Color, in the workplace, is used to help the operator to distinguish objects in the visual field. For example, the green-colored areas on a temperature display indicate acceptable levels, yellow means caution, and red means unacceptably high levels. It is important to point out here that the eyes do not have the same sensitivity to color under different lighting conditions. When it is bright, the eyes are more sensitive to red and less sensitive to blue. The reverse occurs when it is dark. Therefore, we need to consider the lighting conditions under which the color-coded temperature display is viewed. More details on lighting are provided in Chapter 6.3.

People with normal color vision are capable of distinguishing among hundreds of different colors. Complete lack of color vision (color blindness) is extremely rare.

Color deficiency, such as the inability to discriminate between red and green is more common. Usually, these colors are seen as poorly saturated yellows and browns. Overall, about 8% of the male population and less than 1% of the female population have this form of color deficiency. Therefore, color should always be used as a redundant (additional) method of coding, and not the primary method.

6.1.2.1.5 Visual acuity

Acuity refers to sharpness of vision. Visual acuity is the ability to see fine details; for example, detecting differences in the position of two controls, recognizing the presence of an object in the visual field, inspecting a tank for cracks, or even reading the small details on a bottle of a prescription drug. A small distinction between visibility and visual acuity is warranted at this point. The ability to see the pressure gauge on a field display panel refers to visibility, whereas the ability to identify the numbers on the gauge refers to visual acuity.

Visual acuity depends largely on the individual's visual capabilities, such as accommodation of the eyes, visual field, the process of adaptation, color vision, and age (discussed in Section 6.1.2.1.6) and situational factors. Situational factors include:

- *Size of the object*. As the size of the object increases, so does our ability to see the object (Figure 6.1.1). However, this depends on the viewing distance. The further away we are from the object, the less is our visual acuity to see details. For example, we can see many more ground details when sitting in an airplane at an altitude of 2000 feet than at 35,000 feet. Similar logic can be applied in the work environment. If we need to design a sign in the plant or a visual display in the control room, we need to consider what will be the greatest viewing distance (Figure 6.1.2). For example, a process control room operator needs to read information from monitors and displays located at different distances from his or her eyes. Yet, the sizes of characters on most screens are designed for a viewing distance of about 71 cm (28 in.). This may have effects on increased eye strain and visual fatigue as well as the possibility of misreading the information, leading to human error. Formulas are available to calculate the recommended height of letters or figures (Grandjean, 1988; Sanders and McCormick, 1993; Konz, 1990). For favorable viewing conditions (good lighting, good contrast, noncritical display),

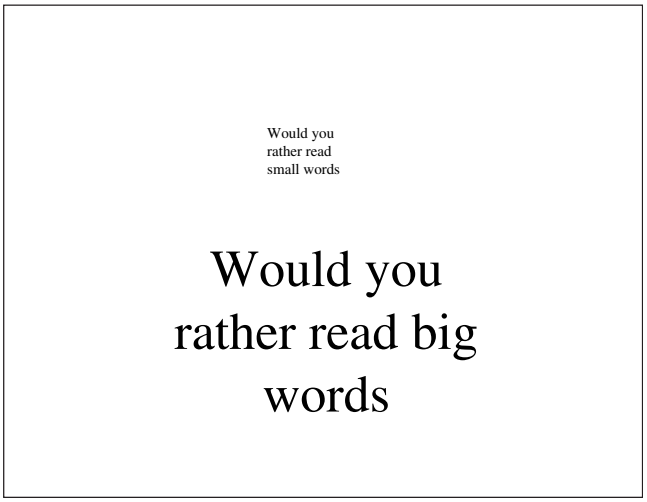


Fig. 6.1.1 Size of object and visual acuity.

$$H = D/200$$

where H is the character height in meters and D is the viewing distance in meters. For unfavorable viewing conditions (low light levels, reduced contrast, critical displays),

$$H = D/120$$

For example, the size of letters to be used for a sign that summarizes evacuation instructions and must be clearly visible from 10 meters (30 feet) away under dim lighting conditions is

$$H = D/120 = 10/120$$
$$= 0.08 \text{ m or } 8.0 \text{ cm } (3.15 \text{ in.}) \text{ high}$$

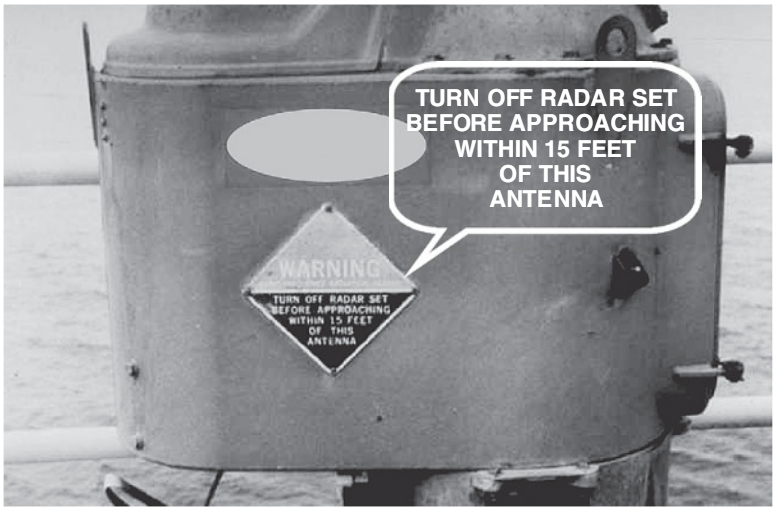


Fig. 6.1.2 Poor design of a warning sign; the text is too small to be read from 15 feet away.

- *Contrast*. This refers to the difference in brightness between a visual target and its background. Generally, as the contrast increases, the ability to see details (visual acuity) increases. For example, detecting dark (black) characters on white background (90% contrast) is much easier than detecting dark characters on relatively darker background (Figure 6.1.3).
- *Illumination*. Up to a point, visual acuity and contrast sensitivity (the ability of the eye to detect the smallest difference, contrast, between a lighter and a darker spatial area) increase, or improve, with increasing illumination. However, high, intense levels of illumination are not always warranted, as they may contribute high levels of glare, weaken information clarity, and lead to temporary blindness. Therefore, the amount of illumination necessary to perform visual tasks is task dependent. Illumination is discussed in detail in Chapter 6.3.
- *Exposure time*. Visual acuity increases with increased exposure time. The longer we view and focus on an object, the more details we see.
- *Observer or target movement*. Our visual acuity decreases as the observer or target move. This is so because movement affects exposure time. For example, consider the difficulty in reading the details of an object on the side of the road while driving a car at 100 km per hour (65 miles per hour).

Dark letters on a dark background are hard to read

Dark letters on a light background are easy to read

Fig. 6.1.3 Example of poor and good contrast.

6.1.2.1.6 Age

The fact that our visual abilities and visual performance decline with age should not be a point of dispute or debate. For example, our visual acuity is reduced by about 30% between the ages of 20 and 60 years. Faster depreciation of vision generally takes place after age 40, and we wonder why we cannot read the fine print on the bottle of a prescription drug.

This decrease in visual acuity is related to the following changes in the structure of the eye:

- Yellowing of the lens, where the amount of light reaching the retina is significantly reduced.
- Decrease in the accommodation power of the eye around the age of 40, which affects depth perception and distance vision. This is caused by deterioration in the muscles that contract and change the lens shape. Depth perception is our ability to judge our distance from objects and the distance between objects in three-dimensional space.
- Retina and nervous system changes around the age of 60, affecting the size of the visual field and sensitivity to low quantities of light.

If the workforce is older, special considerations must be made to accommodate workers who are more likely to experience visual difficulties. Some of these considerations are:

- Higher levels of lighting and larger targets such as characters.
- More viewing time to absorb the information.

- Higher level of contrast between the target and its background, such as black characters on white background.
- Targets set at the same distance from the eyes, since speed of accommodation decreases with age. For example, the displays viewed by a control room process operator should be at the same distance from the eyes.

6.1.2.2 Auditory sense

Hearing is second only to vision in providing information through which we can understand and adjust to our surroundings. For example, a field operator and a control room operator communicate through a radio to change a state of a system such as opening or closing specific valves. Other areas where hearing is key is in the design of warning signals on machines, a buzzer on a fire alarm, speech communication, and the like.

The two defining characteristics of any auditory stimulus are frequency and intensity. The corresponding subjective attribute (what people hear) for frequency is tone and, for intensity, is loudness. This area is discussed in detail in Chapter 6.3 in the section on noise. In many situations, the designer has to choose between using visual and auditory displays or even a combination of the two. The decision on when to use an auditory or visual presentation of information depends on:

- The nature of the message. An alarm that calls for immediate action should be auditory. The message here is short and simple and deals with events in time. Moreover, a person need not look at an alarm to hear it.
- Characteristics of people involved, such as young versus older people (refer to the noise section in Chapter 6.3).
- Conditions under which information must be received. For instance, if the visual system is overburdened, then a buzzer makes sense to alert operators of equipment failure without having to look at a visual display.

In deciding on the type and role of auditory displays, a number of factors must be considered:

- Signal characteristics, which include sound frequency, intensity, duration, and rate of sound occurrence like the on/off emergency signal. Signal characteristics are discussed in Chapter 6.3.
- Individual characteristics, which include the hearing ability and experience of the person, age, and length of exposure to signal.
- Environmental characteristics, which include the type and amount of background (masking) noise.

6.1.2.3 Cognitive capabilities

In the previous section, we discussed how people receive information from the environment through their visual and auditory senses. This section reviews how we process and interpret this information and the associated limiting factors.

The senses (eyes, ears, etc.) receive signals (data) from the environment and pass them to the brain through nerves. Such data are referred to as *sensory information*. The brain acts on the sensory information through complex processing mechanisms to give it meaning. This is referred to as *information processing*. Information processing influences people's behavior, and people's behavior at work is a consequence of information processing.

Information processing involves several different processes: attention, perception, memory, and decision making. These processes are interdependent and serve different functions in the flow and interpretation of information. As we will see and repeat later, receiving and interpreting information requires knowledge, experience, mental alertness, skills, and the ability to formulate new ideas.

6.1.2.3.1 Attention

Consider the following two examples to illustrate the process of attention:

1. When driving a car, we want to ensure that we are watching the road, steering the car, maintaining velocity by looking at the speedometer and adjusting the pressure on the gas pedal, listening to the radio, and perhaps talking to a passenger and drinking coffee.
2. A control room process operator scans and monitors a number of displays (each display presenting different information, temperature and pressure of different units), talks to coworkers, answers the phone, and communicates through the radio with an operator in the field.

It is clear that we are continuously bombarded with information from the environment around us. However, because we have limited capacity to process the information we receive, we must select out the important messages and ignore all other information. Attention is the process of information reduction. It allows us to select important information, and it controls which information is processed further. Occasionally, we do not "pay attention" and miss important information that may be the cause of task errors and accidents. For example, while driving the car we decided to change the radio station (most important activity at that moment). Now the attention is diverted from the road and, suddenly,

a bumper-to-bumper incident takes place. The limiting factors here are:

1. We behave like a single-channel information processing device. We make active decisions about one thing at a time, even though we think we are attending to all things going around us, at all times.
2. We process information at a constant rate because of our limited channel capacity in terms of speed and quantity.

For example, we behave roughly like the automatic feeder of a photocopy machine. If we put in a stack of 100 pages to copy, a single feeder (single channel) takes in a fixed number of pages (process at constant rate) per minute or hour. Four types of situations define the direction of attention (Goldstein and Dorfman, 1978; Broadbent, 1982; Brouwer et al., 1991; Parasuraman and Nestor, 1991; Wickens, 1992):

- Selective attention is a conscious decision to attend voluntarily to important (relative to the individual) information. For example, the control room process operator scans the displays to detect unexpected but important events, such as excessive temperature or pressure.
- Focused attention allows us to filter out unwanted information and attend to one source of important information. For example, the driver changing the radio station is focused on the radio as the one source of important information at that moment.
- Divided attention occurs where we need to attend to several things at once. For example, the control room operator monitors displays and talks to coworkers while eating.
- Sustained attention, also known as *vigilance*, is where we must monitor a situation for a long period of time. For example, operators scan the radar in an airport control tower.

6.1.2.3.2 Perception

After the signals from the environment enter the brain through the eyes or ears, for example, they are interpreted to produce a recognizable pattern of sight or sound. This process, called *perception*, is entirely subjective in nature because it is greatly influenced by the individual's knowledge, experience, expectations, feelings, and wishes (Anderson, 1995). A few facts about perception follow:

- Perception is a complex, active process that can occur with or without external input. Internal input could be a stomach ache, eye pain, or the like.
- It overrides what we know to be true.
- It takes place without awareness (subconsciously).
- It need not correspond with reality.
- Our behavior is determined by perceived rather than actual information.

The main interest for an ergonomist is to design the actual outcome of a job to fit in with the operator's perception of the expected outcome. Two issues can be briefly discussed here: adaptation and direction of movement.

Adaptation is the process of adapting to our environment, which affects our expectations in making a decision and response. For example, if a control room process operator, over a period of time, experiences a high number of false alarms, the expectation is that the next alarm is also false. Therefore, over time, the alarms no longer arouse the operator, and the next true alarm will not be responded to accordingly. Therefore, we should keep the occurrence of alarms to a minimum and ensure that attention is drawn to priority conditions so that arousal level stays high.

The direction of movement or expected relationship between controls and displays needs to meet population stereotypes. Population stereotypes are what most of the population will do on most occasions. For example, in North America, we push a light switch up to turn on the lights. In other parts of the world, the switch is pushed down.

6.1.2.3.3 Memory

Memory is the mechanism of storing and retrieving information. The information stored in the memory is based on our learning and past experiences. For example, driving a car, operating machines, using tools, monitoring a process in the control room—all depend on our capacity to learn and store information. Three stages of memory can be identified. Each stage has different levels of information processing, storage capacity, representation of information, and time span for retention. A simple model of memory is in Figure 6.1.4.

This simple model shows that the stimuli are captured and perceived by the senses (eye, ears, etc.) and held briefly in the sensory memory (SM) storage. The filtering process, through attention, passes on the information considered relatively important to the individual to the short-term memory (STM) storage. Further rehearsal (i.e., repeating, rereading, reviewing material) of the stimuli allows information to pass on and be stored in long-term memory (LTM) (Bailey, 1989; Wickens, 1992; Anderson, 1995; Best, 1995). Each of the stages is briefly presented.

SM is relatively automatic in nature, meaning we need not attend to the information for it to be maintained at this stage. This memory is specific to the sensory system (visual, auditory, etc.) being stimulated. For example, visual information is held in the visual

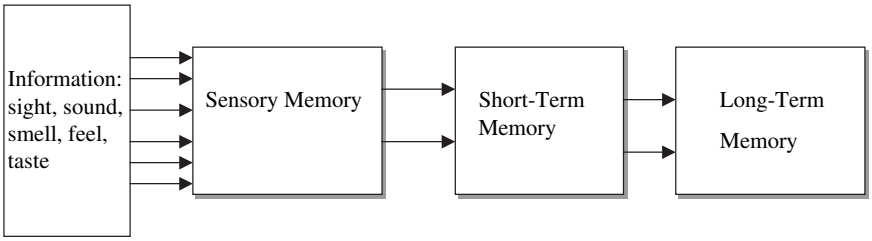


Fig. 6.1.4 Stages of memory storage.

sensory storage (iconic memory), and auditory information is held in the auditory sensory storage (echoic memory). Iconic memory storage lasts for about 1 second, while echoic memory storage can last for 3–5 seconds before the information is lost or decayed. For example, a phone number for an advertised product is flashed briefly on the television screen. As we start writing down the first digits we forget the last ones simply because it was lost from the visual sensory storage. Or, someone is talking to us and we do not hear a single word the person has said. The reasons some of the information is lost in the SM are: (1) decay, where information fades very rapidly; and (2) because material considered important and attended to (paying attention) is passed on to STM, the rest is lost.

In addition, the SM stage serves as a limiting factor in a task sequence. For example, a control room process operator has the task of scanning many pages of alarm messages on a display throughout the shift. While the operator thinks he may have scanned all the important messages, one or more messages did not register and may not have been passed on into the LTM for proper action. This is because the messages were not attended to properly and the memory trace was lost.

STM, also referred to as *working memory*, is the stage where we maintain information we consider important and pay direct attention to. It is also the stage where we try to organize and encode information into the LTM through the process of rehearsal. An example of encoding is Morse code, where either of two codes consisting of dots and dashes or long and short sounds is used for transmitting messages by audible or visual signals. The characteristics of STM are:

- *Short duration and decay*. STM storage lasts for a few seconds. If we pay attention and use the process of rehearsal, then we strengthen the information in the STM and avoid decay; for example, looking up a phone number and then holding it in the STM through rehearsal (repeating the number over and over again) until we have dialed it. Sometimes, we think of a few items we want to do as we drive to the office. If we do not constantly rehearse the items or write them down, we get to the office and cannot remember a thing. The information in this case is lost or decayed simply because of other interfering information and competing demands on the STM. This is why some people use a small audiotape recorder to record things, write phone numbers on the palm of their hand, or write a sequence of valves, to open or close, on their hands or gloves.
- *Distraction, disruption, and interference*. STM faces interference, distraction, and disruption all the time, depending on the competing demands; for example, being interrupted in the middle of a conversation by a phone call and not knowing where we were in the conversation.
- *Limited capacity*. The storage capacity in the STM is limited to seven plus or minus two (five to nine) independent chunks or single pieces of information (Miller, 1956). These chunks could be five to nine digits, words, steps in a procedure, or the like. For example, if you were given two 10-digit telephone numbers to memorize (which exceeds the five to nine rule), you may make mistakes when attempting to remember both of them. However, if you group or chunk the information using your knowledge and association already learned, you can do a much better job remembering. For example, would you rather have to remember these two numbers:
  8008474869 and 2255252752
  or 800-847-4869 and 225-525-2752
  or 800-VISIT-NY and CALL-ALASKA
  Some of the guidelines to improve STM limits are:
  - Minimize the load in terms of time and number; for example, a touch tone phone is faster to enter numbers and, therefore, the memory time is less than when operating a rotary phone.
  - Use chunking as just discussed: Use three to four letters or numbers per chunk (Bailey, 1989). For example, 800-847-4869.
  - Use letters instead of numbers where possible and meaningful; for example, CALL-ALASKA.
  - Keep letters and numbers separate (Preczewski and Fisher, 1990); for example, car license plate B68RN2 is harder to remember than 628 BRN or BRN 628.

- Create meanings for characters; for example, the number 628 BRN can have the following meaning: 6 + 2 = 8, and BRN can be an abbreviation for the color word brown.

The transfer of information from STM to LTM relies on the attention and rehearsal processes. The rehearsal process must involve *not* just rote repetition but the use of associations (information analyzed, compared, and related to other information) between the new information and information already in LTM. This makes retrieving information later on much easier; for example, storing the number 225-252-2752 versus CALL-ALASKA. LTM contains all the learning, knowledge, and experience we obtain during our lifetime. Retrieving information from LTM presents challenges for some of us. We often have difficulty recalling a 25-step procedure or meeting times, names, or even more serious events such as birthday and anniversary dates. However, we do better in recognition (i.e., faces, checklists) than recall. Some of the characteristics of recall and recognition are:

Recall	Recognition
Limited	Almost unlimited
About 2000 words	About 100,000 words
Identify about 8 colors	Discriminate between 500 colors

The factors that affect our ability to remember are:

- Initial learning: how well we analyze, compare, and relate new information to that already learned determines how much and how well we learn. Factors that affect this process are motivation and the desire to learn, prior learning, anxiety or stress—too much anxiety and stress can affect learning, and the method of learning.
- Importance and meaningfulness of the material we are trying to learn.
- Interference, which is best illustrated with an example. An individual who learned to drive a car on the right-hand side of the road is in a different country, learning to drive on the left-hand side. After a while, she is back driving on the right-hand side. Learning the new left-hand side driving interferes with her previous learning and affects her ability to drive safely on the right-hand side for a while, until relearning take place.
- Age: memory ability declines with age and can affect day-to-day activities, as most of us Baby Boomers experience.

6.1.2.3.4 Decision making

Decision making is the process of evaluating alternatives and selecting an action. We tend to search for information related to the problem at hand, estimate the probabilities of different alternatives, and attach meanings and values to anticipated outcomes. Therefore, decisions are a choice among courses of actions. People who must make too many decisions too quickly have to trade off speed of decision making against accuracy of decision outcome. In addition, the time to make an accurate decision is related to the amount of uncertainty in the decision. Naturally, the more uncertainty we have, the longer it takes us to search for the information, estimate probabilities of different alternatives, and attach values to outcomes. Possibilities to improve human decision making include:

- *Design or redesign support systems*. Designing the support system appropriately improves decision making and therefore performance. For example, if we eliminate the need to make manual mathematical calculations by the control room process operator and build that into the process control system, we enhance the speed and accuracy of the decision-making process and eliminate any possibility for human error due to miscalculation.
- *Decision tools*. Decision tools or aids can eliminate a number of the steps the operator has to take to reach a decision; for example, using an expert system to calculate and predict future outcomes based on a large database of previous events. For instance, if there is a continuous rise in temperature or pressure of a boiler, the expert system can predict what may happen, based on the data from previous events, and present the operator with specific decisions to make. The decisions may be to keep the process going or shut the system down. To arrive to such a decision without the aid of an expert system requires the operator to gather information from different displays and sources; consult procedures, supervisors, and coworkers; and so forth. This increases the mental workload and creates an environment for human error.
- *Training*. Training proves especially useful when a system is in an abnormal state. Abnormal situations do not happen frequently, but when they happen, operators may not be ready simply because they have not been regularly trained on such events. Therefore, placing a simulator of the process in the control room center provides the process operators with training, especially in abnormal states. This helps develop the process of decision making and action taking during both routine and abnormal situations.

6.1.2.4 Summary of Information processing

The following is a summary of points to remember on information processing.

- Inattention:
  - Minimize conflicting demands.
  - Ensure redundant checks in critical situations.
- Forgetting when distracted: minimize distraction.
- Inability to retain enough information:
  - Chunk information to ensure no more than 7 ± 2 concepts are addressed at once.
  - Develop logical associations between the information and the action.
- Inability to retain information for prolonged periods: design systems to retain information until it is no longer needed by the operator.
- System demands exceed information processing ability: design calculations into the process control system.
- Inability to recall information from LTM:
  - Replace "recall" with "recognition" through the use of cues and reminders, such as checklists, pull-down menus, help screens, set points.
  - Conduct refresher training (using procedures).
  - Design systems to eliminate the need for procedures or the use of LTM.
- Performing activities out of sequence or too slow: design a logical relationship between one activity and the next, such as building sequence into the task.
- Confusion due to information presentation:
  - Ensure that plant equipment operates in ways that the operator expects and understands.
  - Make the design compatible with operator's expectations.

6.1.3 Physical capabilities

This section of the HSI model, discussed in Section 6.1.1, deals with converting a decision to an action. This is where the operator communicates with the system using bodily actions such as pulling a lever, pressing a button, or speaking over the radio to communicate with others in the plant. For the body, or parts of it, to move, reach, and push a button, or turn a valve requires the involvement of the musculoskeletal (muscles, bones, and connective tissues) system, which provides the primary components for muscular activity. Musculoskeletal movement is also referred to as *motor performance*.

When an operator is called on to perform a physical activity like turning a valve, the amount of work required must be within his or her physical capabilities. The aspects of anthropometry (operator's relevant body dimensions) and biomechanics (mobility of bones, joints, and muscles) define these capabilities. Muscles, in particular, play a key role as limiting factors in many physical tasks and activities. This section considers the factors that limit motor performance to ensure that the operator can reach and apply the necessary force to activate a control. It focuses mainly on two areas of interest: muscular strength and endurance and anthropometry.

6.1.3.1 Muscular strength and endurance

Muscles have limits to their strength and their ability to maintain that strength and resist fatigue (also referred to as muscular endurance). Strength is defined as the maximum force one can exert voluntarily. It is measured in kilograms (kg) or pounds (lb). The work that muscles may have to do is classified into two types:

- Static work, also referred to as *static effort*, is characterized by long-term muscular contraction without body motion occurring; for example, when a posture must be held for long periods like sitting, standing in one place, bending, and twisting to perform a maintenance job.
- Dynamic work, also referred to as *dynamic effort*, is characterized by body motion accompanying muscular tension; for example, repeated muscular contractions with rest periods in between, such as walking.

Most physical activities have both static and dynamic work components. Take, for example, carrying jars filled with products from a sampling point in the plant back to the laboratory. The upper body, especially the arms, shoulders, and back muscles, is under static effort to support the weight and maintain (stabilize) the body in an upright position. However, the lower body, especially the legs, is under dynamic effort to allow the body to move.

Strength is strictly defined as "the maximal force muscles can exert isometrically in a single, voluntary effort" (Kroemer, 1970; Chaffin, 1975). However, strength under dynamic conditions can also be measured (Sanders and McCormick, 1993). Information about isometric muscular strength capability is needed to specify, for example, the minimum valve resistances and the forces required to safely carry bags of chemicals. Most tasks or activities that operators perform require the integrated exertion of many muscle groups. For example, lifting a bag of chemicals off the floor requires squatting, extending arms, grabbing the bag with both hands, lifting the bag with the arms and shoulders, lifting the body using the

legs, and the like. In this complex situation, the maximum force that can be exerted is determined by the weakest link in the different muscle groups involved.

6.1.3.1.1 Factors affecting strength

There is a wide range of muscular strength capabilities among people, from heavy weight lifters to sedentary TV watchers. In addition, many variables play the role of limiting factors and can influence muscular strength:

- *Age*. Strength increases rapidly in the teens, reaches its peak in the middle to late twenties, levels off, and starts decreasing, slightly in the forties and faster in the fifties and sixties. Most tasks and jobs are designed around the young worker. But, the strength required may exceed the capacity of older groups. This is truly an issue today, since the current workforce contains many older workers.
- *Gender*. In general, women have less-developed muscles and their overall mean strength is about 67% that of males (Hettinger, 1961; Roebuck, Kroemer, and Thomson, 1975; Lauback, 1976). However, we must not generalize this figure (Wilmore, 1975; Redgrove, 1979; Pheasant, 1983; O'Brien, 1985; Chaffin and Andersson, 1991; Sanders and McCormick, 1993). The reasons are:
  1. This 67% is an average value from various groups of muscles. The range is 35–85% of male mean strength.
  2. For the lower extremities, the average strength of females is comparable with males. This may reflect the use of lower extremities by both males and females for daily regular activities. But, in general, men use their upper extremities more than women.
  3. Trained female athletes are stronger than many untrained men.

In terms of designing manual tasks, such as lifting, the maximum weight that can be lifted is determined by the weakest member of the two genders involved.

- *Occupation*. Relates to the different levels of muscle use (exertion) by people in different occupations; for example, a process operator versus a desk clerk given that neither is a fitness fanatic.
- *Physical training*. Physical training, especially weight lifting, tends to increase muscular strength and endurance by as much as 30–50%.
- *Others*. Other factors that have been associated with influencing muscular strength are:
  1. Gloves. Wearing gloves, on average, results in a loss of about 20% in grip strength. Of course, this depends on the thickness of the gloves: the thicker the gloves, the less the strength.
  2. Body dimension, such as height and weight. Body weight is an important determinant of arm strength. Likewise, height correlates highly with pulling strength.
  3. Body position. For example, the strength of elbow flexion is at a maximum with the elbow at about 90° but drops off to about half this value at extreme angles.
  4. Diet. Inadequate and improper nutrition decrease body strength.
  5. Other personal factors: genetics, health, fatigue, drugs, motivation.

6.1.3.1.2 Endurance and fatigue

As mentioned earlier in this chapter, *muscular endurance* refers to the ability of the muscle to continue to work. In the case of static work, the length of time (endurance) a force can be maintained depends on the proportion of the available strength being exerted. For example, maximum force can be maintained for a very short time (few seconds). However, the smaller the force exerted, the longer it can be maintained (Caldwell, 1963, 1964; Monod, 1985; Deeb, 1988). The applications of these data are important. If we expect an operator to apply and maintain static force to hold an object (i.e., a drill weighing 15 kg or 33 lb) for a period of time, the designer must consider endurance. In the case of dynamic work, *endurance* is defined in terms of force and frequency of repetition. However, the endurance time is significantly longer. For example, 25% maximum voluntary contraction can be maintained for about 10 minutes under static contraction compared with over 4 hours under dynamic work.

Fatigue is a product of the work performed, whether dynamic or static. The degree and severity of fatigue constitutes the level of discomfort, pain, distraction, and reduced performance that may lead to accidents. It is, therefore, very important to design tasks by eliminating factors that induce fatigue.

6.1.3.2 Anthropometry: body size

Anthropometrists, measurers of the human body, have collected body size data for many years. For example, the measuring units of foot and hand have been derived from the dimensions of body parts. The term *anthropometry* is derived from two Greek words: antropo(s), or human, and metricos, or measurement. Anthropometry is used extensively by ergonomists to design tools, equipment, plants, manufacturing lines, clothes, shoes, and the like to ensure the proper fit to the person. Therefore, to achieve proper fit, it is important to have details on the dimensions of the appropriate body part. For example, the size of the hand is used to design the dimensions of

controls such as switches and pushbuttons, while details of arm reach are necessary to position controls at appropriate distances.

Two primary categories of anthropometric data interest ergonomists:

- *Structural anthropometry*, also referred to as *static anthropometry* or static dimensions. These are measurements with the body in a still or fixed position; for example, stature or height, weight, head circumference.
- *Functional anthropometry*, also referred to as *dynamic anthropometry* or *dynamic dimensions*. These are measured with the body engaged in various work postures, indicating the ranges of motion of individual body segments; for example, arm reach.

It is important to point out that static anthropometry data are often measured on unclothed (nude) individuals, mainly to ensure consistent results. Therefore, corrections must be made to account for increases in body size due to clothing, such as a process operator working outside during an Alaskan winter and another during a Texas summer. In addition, allowances must be made for wearing safety shoes and hard hats, which could add about 10–12 cm (4–5 in.) to the stature that must be considered in the design.

6.1.3.2.1 Sources of body size variability

The various genetic, biological, and physiological differences between humans influence the way they vary with respect to body dimensions in terms of height, weight, shape, and the like. This can be noticed if you observe people in a shopping mall. Therefore, we need to be very careful using anthropometric data if they are to be of value (Pheasant, 1982a, 1982b).

The common practice in ergonomics is to specify anthropometric data in terms of percentiles. A percentile refers to a percentage of the population with a body dimension up to a certain size or smaller. For example, if a 95th percentile height (stature) is 170 cm (66.9 in.), it indicates that 95% of the population have heights up to 170 cm. Or, 95% of the population have heights of 170 cm or less and 5% are taller than 170 cm. If, on the other hand, the 5th percentile stature is 150 cm (59 in.), it indicates that 5% of the population are shorter than 150 cm and 95% are taller.

When a particular design (i.e., placing a valve hand wheel at a given height) is expected to be used by many people, we need to consider the following variables in influencing body size and adjust the design accordingly:

- *Gender*. Men are generally larger than women at most percentiles and body dimensions. The extent of the difference varies from one dimension to another. For example, hand dimensions of men are larger than women sizes—hand finger thickness of men is about 20% larger, while fingers are about 10% longer for men than for women (Garrett, 1971). Women, in general, exceed men in five dimensions: chest depth, hip breadth and circumference, thigh circumference, and skin-fold thickness. O'Brien (1985) proposes a complete list of anthropometric differences between men and women.
- *Age*. Body dimensions generally increase from birth to early twenties, remain constant to around age 40, and decline afterward into old age as part of the normal aging process. For example, stature or height reaches full growth at around age 20 for males and 17 for females (Trotter and Gleser, 1951; Damon, Stoudt, and McFarland, 1971; Roche and Davila, 1972; Stoudt, 1981). Decline in stature is more pronounced in women than in men. Therefore, it is important to define the user population early in the design cycle.
- *Nationality and culture*. Nationalities and cultures differ in body sizes. For instance, Asians tend to be somewhat shorter on average than Northern Americans, while certain cultures from southern Sudan (Africa) tend to be taller. Therefore, it is important for the designer to define and use the anthropometry data related to the user population nationality and culture. Table 6.1.1 (and Figure 6.1.5) presents an example of anthropometric values for three different nationalities: North American, Japanese, and Hong Kong.
- *Occupation*. Differences in body size dimensions among occupational groups is common and well documented. For example, manual workers, on the average, have larger body sizes than sedentary workers. Sanders (1977) found truck drivers to be taller and heavier than the general civilian population. This difference among occupations may be the result of

  1. diet;
  2. exercise;
  3. physical activities imposed by the job (i.e., manual handling);
  4. imposed selection, such as individuals need to be a certain height to be accepted in a particular job;
  5. self-selection, such as individuals with a given height choose a particular job for practical or sociological reasons;

  We must take great care of not using anthropometric data obtained from groups of one occupation, such as armed forces, to design the environment of another, such as office workers.
- *Historical trends*. The average size of people has been increasing over the years. For instance, the average adult height in Western Europe and the United States

Table 6.1.1 Selected anthropometry data for different nationalities in centimeters (cm) and inches (in.)

Anthropometric Dimensions	North Americans				Japanese				Hong Kong			
	95th% man		5th% woman		95th% man		5th% woman		95th% man		5th% woman	
	(cm)	(in.)	(cm)	(in.)	(cm)	(in.)	(cm)	(in.)	(cm)	(in.)	(cm)	(in.)
A. Vertical grip/reach	217.5	85.7	177.8	70.0	207.5	81.7	168.0	66.1	210.5	82.9	168.5	66.3
B. Stature/head height	184.4	72.6	149.5	58.9	175.0	68.9	145.0	57.1	177.5	69.9	145.5	57.3
C. Shoulder height	152.4	60.0	121.1	47.7	143.0	56.3	107.5	42.3	146.0	57.5	118.0	46.5
D. Elbow height	119.0	46.9	93.7	36.9	110.5	43.5	89.5	35.2	108.0	42.5	87.0	34.3
E. Eye height	172.7	68.0	138.2	54.4	163.5	64.4	135.0	53.1	164.0	64.6	133.0	52.4
F. Forward grip/reach	88.3	34.8	64.0	25.2	75.0	29.5	57.0	22.4	77.0	30.3	58.0	22.8
G. Knuckle height	80.5	31.7	64.3	25.3	80.5	31.7	65.0	25.6	81.5	32.1	65.0	25.6
H. Knee height	59.2	23.3	45.2	17.8	42.0	20.9	42.0	16.5	54.0	21.3	41.0	16.1
I. Waist height	110.5	43.5	86.4	34.0	89.5	35.2	70.0	27.6	92.0	36.2	71.5	28.1

Note: Add about 4 cm (1.5 in.) for shoes and 7.5–10 cm (3–4 in.) for hard hat.

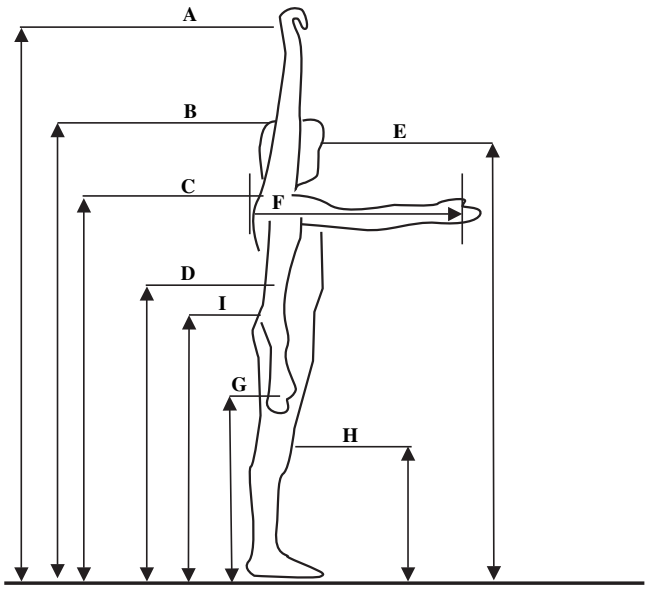


Fig. 6.1.5 Anthropometry figure to guide Table 6.1.1.

increased about 1 cm (0.4 in.) per decade (Sanders and McCormick, 1993). This is so, perhaps, because of better diet, medical care, hygiene, and living conditions. As designers, we need to consider present-day users as well as future generations to ensure proper systems design few decades down the road.

- *Body position*. Posture affects body size. For example, restraints such as seat belts, affect data applicability of forward reach.
- Clothing. As mentioned earlier, almost all anthropometric data are obtained from nude individuals. Therefore, the type (material) and amount of clothing add to body size and can also create restriction of movement such as affecting overhead and forward reach. Another example is the use of gloves where allowance must be made to accommodate different thicknesses of gloves.

6.1.3.2.2 Principles of body size application

When determining the proper anthropometric data to be used in a design, the following must be observed: we need to carefully define the population or group we are designing for and ensure that the data are reasonably representative. For most purposes, a range of dimensions from the 5th to the 95th percentile is generally acceptable. The range can also increase, if possible, from the 2nd to the 98th percentile or even larger. The choice of design percentile is largely a matter of cost. Three general principles of body size application to specific design problems are accepted:

1. *Design for the average*. The average value is taken as the 50th percentile, meaning 50% of the population is above and 50% below this value.
2. *Design for the extreme*. In designing for the extreme, the ergono-mists constantly apply the following two principles:
   - Design fit or clearance dimensions for the largest individual.
   - Design reach dimensions for the smallest individual.

It is frequently the practice to use the 95th percentile male for the clearance or fit dimensions and the 5th percentile female for reach dimensions. Therefore, it is safe to say that a design that would accommodate individuals at one extreme would also accommodate virtually the entire population.

3. Design for the adjustable range. Designing for the adjustable range is generally the preferred method to accommodate individuals of varying sizes. This use of adjustability can be seen in car seats, office chairs, desk heights, bench heights in a maintenance shop, adjustable tables for manual materials handling jobs, and the like. However, this may not always be possible:
   - If we strictly use the range, then we must accommodate people from 3 feet tall to 9 feet tall and weighing from 23 kg (50 lb) to 227 kg (500 lb). This is why the 5th and 95th percentile values are traditionally advocated.
   - Adjustability may not have any practical value and the cost-out weighs the benefit; for example, an eye wash station in a plant or a bathroom toilet height.

6.1.4 Case study

The force an operator is able to apply to a valve hand wheel depends on the location of the wheel relative to the operator, the orientation of the valve stem, the diameter of the wheel, and the quality of the interface between the wheel and the hands. The objective of this study was to create a database of human force capabilities, to apply a hand wheel to open or close a valve, that can be used by plant designers in the selection, placement, and orientation of valves. A summarized version of this study is presented here. For more details, consult Attwood et al. (2002).

6.1.4.1 Method

6.1.4.1.1 Participants

Sixty-six volunteer process operators participated in the study. Data from nine volunteers were rejected: seven subjects were removed because of irregularities in their data; two women were removed to provide uniform gender data. Analyses were performed on 57 men.

Subjects were screened for health problems before participation. Subjects signed a consent form.

Warm-up and practice sessions were conducted. Each subject was provided with a small token of appreciation.

6.1.4.1.2 Equipment

Force data were collected on the "primus" device manufactured by the Baltimore Therapeutic Equipment Company, Hanover, Maryland (Figure 6.1.6).

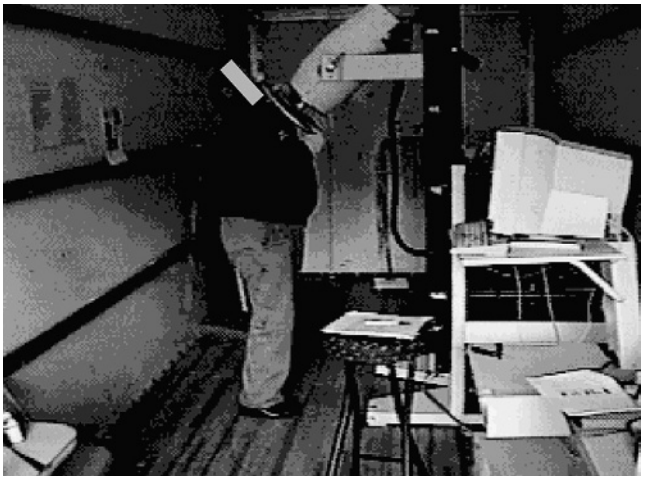


Fig. 6.1.6 Subject participating in study.

6.1.4.1.3 Procedure

- Nine valve positions were identified to represent typical heights and stem angles.
- Valve positions were tested randomly within low, medium, and high heights.
- Operators completed three exertions at each valve position.
- Operators were instructed to assume standard position and exert maximum force on valve wheel in a counterclockwise direction, increasing force smoothly to maximum and holding for 6 seconds.
- Operators were allowed 30 seconds rest between each repetition and 2 minutes rest between different valve positions.

6.1.4.2 Data collected

- Age, gender, height, weight, and functional reach were compared with population norms.
- Average force, 3–5 seconds into repetition.
- Peak force per repetition.

6.1.4.2.1 Data analyses

The maximum average force was identified from three repetitions per participant for each valve position. Also determined were the mean and standard deviation of force for each of the nine valve positions, as well as the 5th percentile of force for each valve position and the relationship between force and the stature of the operator.

6.1.4.3 Conclusion

The data suggested a relationship between stature and force capability. The valve locations that produced the

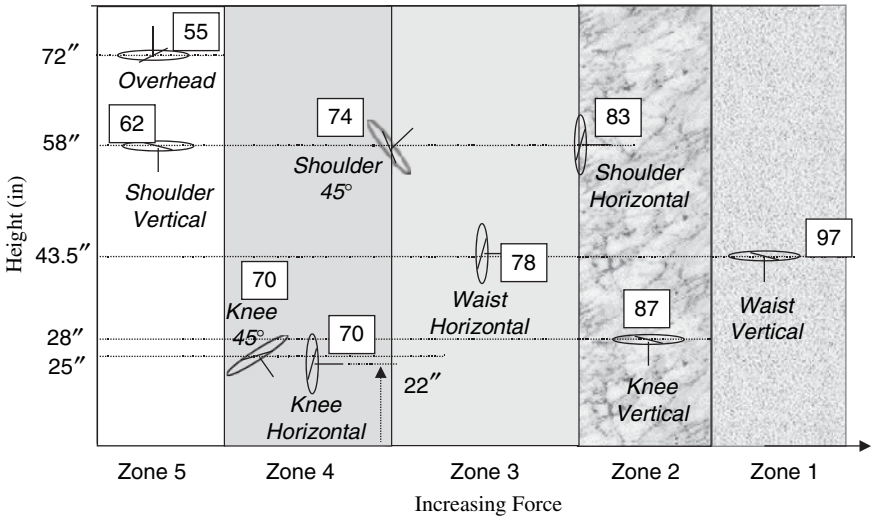


Fig. 6.1.7 The 5th percentile force (in lb) as a function of valve height and stem orientation. Note: Means within the same column are not statistically different from each other but are different from those in different columns. Each column is defined as a "Zone."

highest turning forces differ from those typically used by designers. The 5th percentile force capability is much lower than recommended historically. Figure 6.1.7 presents the 5th percentile force (in lb) as a function of valve height and stem orientation.

The recommendations for existing operations are

- Identify critical manually operated valves that are operated frequently, in an emergency, and have time constraint for operation.
- Measure the valve force required to crack critical valves.
- Modify those that exceed operator force capability (for the existing location and orientation).

6.1.5 Review questions

1. What is the function of the retina of the eye?
2. Define accommodation of the eye.
3. What is nearsightedness and farsightedness?
4. What are the areas of the visual field?
5. Describe the process of adaptation of the eye.
6. Why should we not use color as the primary method of coding?
7. What situational factors affect visual acuity?
8. Describe how attention is essential in processing information.
9. Why is perception a subjective and not an objective process?
10. Describe the three stages of memory: sensory, short-term, and long-term.
11. Describe what to do to improve the limited capacity of STM.
12. What factors affect strength?
13. What is anthropometry?
14. Discuss the concept of adjustable design.
15. Why there is no such thing as an average human?
16. Give examples of the two categories of anthropometry.
17. List the sources of body size variability.
18. What does percentile mean?
19. Discuss the three principles of body size application.
20. Why design for the extremes? Give examples.

References

Anderson, J. R. (1995) *Cognitive Psychology*. Fourth Ed. New York: W. H. Freeman.

Attwood, D. A., Nicolich, M. J., Doney, K. P., Smolar, T. J., and Swensen, E. E. (2002) "Valve Wheel Rim Force Capabilities of Process Operators." *Journal of Loss Prevention in the Process*, 15, pp. 233–239.

Bailey, R. W. (1989) *Human Performance Engineering*. Second Ed. Englewood Cliffs, NJ: Prentice-Hall.

Best, J. B. (1995) *Cognitive Psychology*. Fourth Ed. St. Paul, MN: West Publishing.

Broadbent, D. (1982) "Task Combination and Selective Intake of Information." *Acta Psychologica* 50, pp. 253–290.

Brouwer, W. H., Waterink, W., Van Wolffelaar, P. C., and Rothengatter, T. (1991) "Divided Attention in Experienced Young and Older Drivers: Lane Tracking and Visual Analysis in

a Dynamic Driving Simulator." *Human Factors* 33, pp. 573–582.

Caldwell, L. S. (1963) "Relative Muscle Loading and Endurance." *Journal of Engineering Psychology* 2, pp. 155–161.

Caldwell, L. S. (1964) "Measurement of Static Muscle Endurance." *Journal of Engineering Psychology* 3, pp. 16–22.

Chaffin, D. (1975) "Ergonomics Guide for the Assessment of Human Strength." *American Industrial Hygiene Association Journal* 35, pp. 505–510.

Chaffin, D., and Andersson, G. (1991) *Occupational Biomechanics*. Second Ed. New York: Wiley.

Damon, A., Stoudt, H. W., and McFarland, R. A. (1971) *The Human Body in Equipment Design*. Cambridge, MA: Harvard University Press.

Deeb, J. M. (1988) "An Exponential Model of Isometric Muscular Fatigue as a Function of Age and Muscle Group." Dissertation, Department of Industrial Engineering, State University of New York at Buffalo.

Garrett, J. W. (1971) "The Adult Human Hand: Some Anthropometric and Biomechanical Considerations." *Human Factors* 13, pp. 117–131.

Goldstein, I., and Dorfman, P. (1978) "Speed Stress and Load Stress as Determinants of Performance in a Time-Sharing Task." *Human Factors* 20, pp. 603–610.

Grandjean, E. (1988) *Fitting the Task to the Man*. London: Taylor and Francis.

Hettinger, T. (1961) *Physiology of Strength*. Springfield, IL: Charles C Thomas.

Konz, S. (1990) *Work Design: Industrial Ergonomics*. Third Ed. Worthington, OH: Publishing Horizons.

Kroemer, K. H. E. (1970) "Human Strength: Terminology, Measurement, and Interpretation of Data." *Human Factors* 12, pp. 297–313.

Lauback, L. L. (1976) "Comparative Muscular Strength of Men and Women: A Review of the Literature." *Aviation, Space, and Environmental Medicine* 47, pp. 534–542.

Miller, G. (1956) "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review* 63, pp. 81–97.

Monod, H. (1985) "Contractility of Muscle during Prolonged Static and Repetitive Dynamic Activity." *Ergonomics* 28, pp. 81–89.

O'Brien, M. (1985) "Women in Sport." *Applied Ergonomics* 16, pp. 25–39.

Parasuraman, R., and Nestor, P. G. (1991) "Attention and Driving Skills in Aging and Alzheimer's Disease." *Human Factors* 33, pp. 539–557.

Pheasant, S. T. (1982a) "A Technique for Estimating Anthropometric Data from the Parameters of the Distributions of Stature." *Ergonomics* 25, pp. 981–992.

Pheasant, S. T. (1982b) "Anthropometric Estimates for British Civilian Adults." *Ergonomics* 25, pp. 993–1001.

Pheasant, S. T. (1983) "Sex Differences in Strength—Some Observations on Their Variability." *Applied Ergonomics* 14, pp. 205–211.

Preczewski, S. C., and Fisher, D. L. (1990) "The Selection of Alphanumeric Code Sequences." Proceedings of the Human Factors Society 34th Annual Meeting, Santa Monica, CA, pp. 224–228.

Redgrove, J. (1979) "Fitting the Job to the Women: A Critical Review." *Applied Ergonomics* 10, pp. 215–223.

Roche, A. F., and Davila, G. H. (1972) "Late Adolescent Growth in Stature." *Pediatrics* 50, pp. 874–880.

Roebuck, J., Kroemer, K., and Thomson, W. (1975) *Engineering Anthropometry Methods*. New York: Wiley-Interscience.

Sanders, M. S. (1977) "Anthropometric Survey of Truck and Bus Drivers: Anthropometry, Control Reach, and Control Force." Westlake Village, CA: Canyon Research Group. Cited in M. S. Sanders and E. J. McCormick, *Human Factors in Engineering and Design*. New York, McGraw-Hill, 1993.

Sanders, M. S., and McCormick, E. J. (1993) *Human Factors in Engineering and Design*. New York: McGraw-Hill.

Stoudt, H. W. (1981) "The Anthropometry of the Elderly." Human Factors 23, no.1, pp. 29–37.

Trotter, M., and Gleser, G. (1951) "The Effect of Aging upon Stature." *American Journal of Physical Anthropometry* 9, pp. 311–324.

Wickens, C. D. (1992) *Engineering Psychology and Human Performance*. Second Ed. New York: HarperCollins.

Wilmore, J. H. (1975) "Inferiority of Female Athletes: Myth or Reality?" *Journal of Sports Medicine* 3, pp. 1–6.

This page is intentionally left blank

Chapter 6.2

6.2

Physical factors

Dennis Attwood, Joseph Deeb and Mary Danz-Reece

The objective of this chapter is to provide an overview and methods on how to reduce work-related risk factors that contribute to musculoskeletal disorders (MSDs). A musculoskeletal disorder is an injury that affects the bones, muscles, or other related tissues of the body joints.

In petrochemical operations, the two primary types of physical work tasks on which to focus are manual handling tasks and repetitive tasks. Since numerous physical work tasks are done at a plant, it is important to identify the tasks that hold high priority; that is, those that should have reduced risk factors to reduce the potential for MSDs. A qualitative screening approach assists in selecting the physical work tasks that need further attention and possible risk reduction. The screening approach categorizes work tasks into high-, medium-, and low-priority groups based on relative risk.

In some cases, further assessment of a task is needed to focus in on the specific risks that need reduction and guide interventions to reduce the risks. This chapter includes an overview of the main tools for assessing risk factors in manual handling and repetitive tasks. The focus is on assessment tools for the practitioner to use to evaluate the tasks and less reliance is placed on self-reporting by the individuals doing the task, since people tend to be inaccurate at estimating weights and forces, especially in unfamiliar situations (Wiktorin et al., 1996).

This chapter also offers a framework for an ergonomics program so that risk identification and reduction can be sustained on an ongoing basis. The ergonomics program recommended in this chapter is a specific application of the general program model.

6.2.1 Musculoskeletal disorders

Musculoskeletal disorders affect the bones and muscles of the body and the tissues that form the body joints. There are two categories of disorders, based on the type of event that caused it: (1) conditions caused by an acute trauma, such as a slip or fall resulting in, for example, a strained back, bruised leg, or sprained ankle; and (2) conditions due to exposure to a repeated, or chronic, type of physical activity, resulting in, for example, soreness from inflamed tendons or ligaments. Conditions associated with repeated exposure to physical activity are called *cumulative trauma disorders* (CTDs). *Repetitive strain injury* (RSI) is another term for a cumulative trauma disorder specifically related to repetitive tasks. Examples of CTDs include

- Tendon disorders: tendinitis, tenosynovitis, DeQuervain's disease.
- Nerve disorders: carpal tunnel syndrome, ulnar or radial nerve damage.
- Neurovascular disorders: thoracic outlet syndrome, white finger (Raynaud's syndrome).

Some people also consider low back pain a type of CTD, if a single acute cause cannot be identified. CTDs can result from, be precipitated by, or be aggravated by intense, repeated, sustained, or insufficient recovery from exertion, motions of the body, vibration, or cold. CTDs generally develop over periods of weeks, months, and years. Physical activity risk factors related to CTDs include (Occupation Safety and Health Administration, 1995)

- *Application of force*. Higher forces translate into higher loads on the muscles, tendons, and joints, which can quickly lead to muscular fatigue.

Ergonomic Solutions for the Process Industries; ISBN: 9780750677042
Copyright © 2003 Elsevier Inc. All rights of reproduction, in any form, reserved.

- *Repetitive motion*. This is defined as performing the same motion every 30 seconds or less or where 50% of the work cycle involves similar upper extremity motion patterns.
- *Awkward posture*. An awkward posture requires more muscular force because muscles cannot work as effectively.
- *Contact stress*. Tools, objects, or equipment that create pressure against the body (usually the hands and arms) can inhibit nerve function and blood flow.
- *Overall muscular fatigue*. Insufficient recovery time between muscular contractions may lead to overall muscular fatigue.

The magnitude of risk associated with a specific quantity of exposure to these factors is not well defined, but there is consensus that exposure to high levels or combinations of these risk factors increases the risk of the CTDs. For example, force and repetition are recognized multiplicative factors; that is, if they are simultaneous risk factors, the interaction between them likely results in an increased risk of CTD compared to a single factor. Also, tasks with high repetition and high forces are 14 times as likely to be associated with a type of CTD in the wrist called *carpal tunnel syndrome* (CTS) than low-repetition, low-force tasks (Silverstein, Fine, and Armstrong, 1987).

Because there can be multiple confounded causes of an MSD, and particularly a CTD, it is often difficult to determine whether a condition is work or nonwork related. The ergonomist's goal is to identify the most critical risk factors that may be present in the work setting that could aggravate or contribute to the cause of an injury. The critical risk factors are those that the ergonomist believes are most important to address for the best possibility of reducing the risk of an injury.

To identify the critical risk factors, an ergonomist conceptually categorizes the physical work activities into two types, those that consist mainly of single exertions involving the entire body and those that are more repetitive and most likely involve more intensive use of the upper body or arms and hands. The physical work activities that involve whole-body exertions typically involve carrying or moving an object, so they are called *manual handling tasks*. The ergonomist uses different approaches for assessing the risk factors depending on whether the task is a manual handling task or a more repetitive task.

6.2.2 Manual handling tasks

Manual handling tasks are activities during which workers move objects from place to place by lifting, lowering, carrying, pushing, or pulling. In petrochemical plants, examples of manual handling activities include

- Lifting a sack of additive and holding the sack while pouring the contents into a mixer.
- Picking up boxes of packaged motor oil from a pallet.
- Pushing and pulling a loaded pallet jack.
- Tilting and rolling drums of chemicals in staging areas.
- Grasping and turning a handwheel to manually open and close a valve.
- Holding and pouring a large bottle of liquid additive into a blender.
- Picking up a hose, dragging it, and holding it in position to connect at a manifold station.
- Picking up and carrying plant spare parts or toolboxes.

6.2.2.1 Manual handling risk factors

There are general guidelines that can be used to screen out the tasks where the force required could be a concern. A task that involves any of the following features may pose risk that could be reduced:

- Repetitive or prolonged kneeling or squatting.
- Significant sideways twisting of the body.
- Repetitive or prolonged reaching with hands above shoulder height.
- Performance in a confined area that restricts natural posture.
- More than 16–27 kg (35–60 lb) handled and cannot be held close to the chest or torso.
- More than 27 kg (60 lb) handled.
- More than 8 kg (17 lb) lifted from or below knee height or above shoulder height.
- Unbalanced carrying or lifting (e.g., the bulk of the load supported by one side of the body).
- More than 5 kg (10 lb) handled by only one hand (e.g., a heavy tool).
- Unstable, unbalanced, difficult, or awkward to handle loads being moved.
- Object of more than 75 cm (30 inches) in two dimensions carried or lifted.
- Object weighing more than 8 kg (17 lb) carried more than 9 m (30 ft).
- Combinations of physical tasks performed for more than 1 hour at a time.
- Physical tasks requiring heavy exertion performed more than once per minute for continuous periods of more than 15 minutes.

For manual handling tasks in petrochemical plants, the ergonomist's concern is usually about the maximum force required in a single or just a few exertions. The

MSD risk of a manual materials handling task is determined by comparing the force actually required to execute the task to an estimate of the maximum force that would be safe for this person to exert in similar circumstances. For a lifting task, the actual weight that must be lifted is compared to a recommended weight limit for that situation. For pulling, the force required to pull an object can be measured with a force gauge and compared to the recommended maximum pulling forces that can be safely exerted under various circumstances (see Figure 6.2.1).

Fewer highly repetitive lifting tasks are encountered in petrochemical operations than in other industrial operations, such as parcel package mailing and delivering. But, there may be occasions when the recommended weight limits or forces would be determined for multiple exertions (e.g., several lifts per hour, as in a warehousing operation unloading a pallet and stacking the boxes on shelves).

For lifting, aspects of the object, task, individual's technique, and the environment influence the amount of load that can be safely handled in various situations (as listed in Figure 6.2.2).

The size, weight, stability, and contents of the load influence the amount of weight that can be handled safely. So does the location of the load when it is picked up: the height of the load from the floor and horizontal distance of the load from the body are critical factors in assessing risk of a lifting situation. The number of times the person must lift a load during a time period is also considered. The technique the individual uses to move the load also influences the risk, as does individual factors such as the person's strength, size, and endurance. In addition, environmental factors such as quality of lighting in the work area and path and housekeeping factors such as obstructions and condition of the floor surface also influence how safely a load can be handled.

Fig. 6.2.1 Using a force gauge to measure a pushing/pulling task.

A safe lifting task, of course, is one in which the weight of the load does not exceed the weight that could

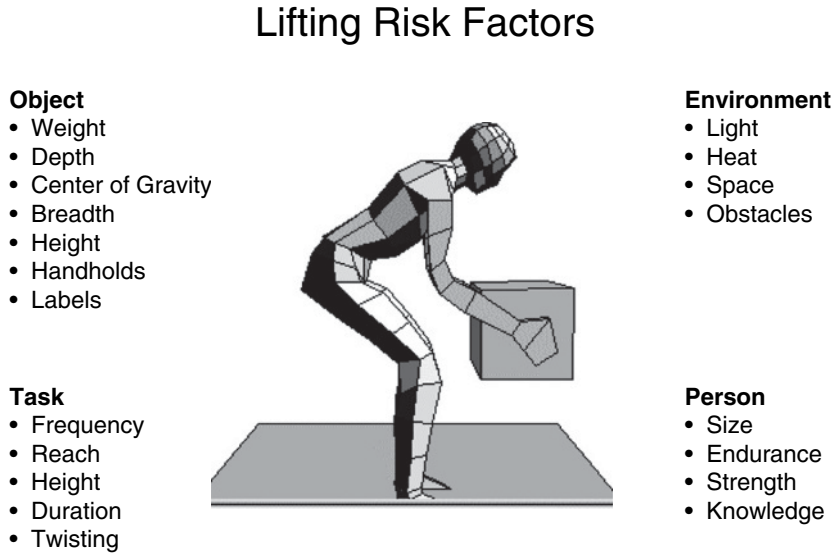


Fig. 6.2.2 Risk factors that influence the weight of the load.

be safely lifted. An optimal lifting task is one in which the object is compact, easy to grasp and handle, and requires no awkward movements to maneuver the load. The object is easily accessible from between waist and elbow height and oriented so that it can be approached from straight on without bending or stooping toward it. This configuration is the least stressful on the body for lifting and most favorable for safely lifting a load.

Situations that deviate from the optimum scenario contain risk factors. Examples of these situations include bulky and unstable objects, objects on the floor or overhead, and objects that are located so that twisting the torso or deep stooping is required for access.

General guidelines for manual handling tasks are

- Avoid static work as much as possible.
- Analyze and redesign jobs requiring the manual handling (lifting, pulling, and pushing) of heavy material and equipment, as necessary.
- Provide mechanical assistance for jobs in which manual forces exceed the limits of the operator.
- Provide mechanical assistance for jobs requiring frequent handling of heavy loads.
- Design manual handling jobs to avoid twisting as much as possible.

6.2.2.2 Methods for evaluating manual handling tasks

Three main tools are available to evaluate MSD risk of manual handling tasks: (1) checklists and postural observation, (2) National Institute of Occupational Safety and Health (NIOSH) lifting equation, and (3) mathematical models based on biomechanical principles. By design these tools tend to yield conservative results. Evaluations from any single tool should not be solely relied on, but should be considered as input into the prioritization and decision-making process. The training and skill required to use these tools depends on the type of tool. The checklists and postural observation tools described in this chapter can be effectively used by personnel who have minimal ergonomics training. Use of other tools, such as the NIOSH lifting equation, is more complex and requires more training for proficiency. For example, Dempsey et al. (1991) show that trained individuals can effectively apply the NIOSH lifting equation but untrained individuals make critical errors when using the equation. Use of mathematical and biomechanical models is not appropriate for the layperson and best left to the trained specialists. Training to use the models is available, however, and those that may have to use the model frequently get more out of the training than those who use it infrequently.

6.2.2.2.1 Postural observation checklists for manual handling tasks

The objective of postural observation methods is to identify the tasks or portions of tasks most physically stressful. To use most postural observation methods, all that is needed is the worksheet and a pencil, but for most methods it is recommended that the task be videotaped so that it can be reviewed while scoring the worksheet. Two available methods are rapid entire body assessment (REBA), developed by Hignett and McAtamney (2000), and Ovako working posture analyzing system (OWAS), developed and validated by Karhu, Kansi, and Kuorinka (1977). Subsequently, OWAS was applied in industrial settings and the analyses published (Karhu, Harkonen, Sorvali, and Vepsalainen, 1981; Kivi and Mattila, 1991; Mattila, Karwowski, and Vilkki, 1993; Scott and Lambe, 1996). The OWAS method is intended to be applied to whole body analyses, such as for manual handling tasks, and therefore, it provides very little precision for upper limb assessment. The method was developed specifically so that it could be used successfully by untrained personnel and provide unambiguous answers. The procedure and forms for using the OWAS can be found in Karhu et al. (1981).

Attachment 1 is a worksheet that can be used to evaluate and set priorities for materials handling tasks to identify those of highest priority for reducing MSD risk. The worksheet provides a rating system for each aspect that affects the safety of a manual handling task, such as nature of the object, weight lifting, posture used, repetition, and other conditions of the task. The ergonomist fills out one worksheet per task and tallies the score for the entire task. Then, tasks are compared and tasks with the higher scores receive the higher priority for follow-up.

6.2.2.2.2 Calculation of weight limit for two-handed lifting tasks

NIOSH used a conservative approach to develop an equation to calculate the recommended weight limit (RWL) for simple two-handed lifting tasks (Waters et al., 1993; Waters, Putz-Anderson, and Garg, 1994). The equation is applicable to specific types of lifting tasks and requires training for effective application (Dempsey et al., 1991). Table 6.2.1 is a checklist that allows the user to determine when the NIOSH equation should *not* be used. If the task meets any of these conditions, then the equation cannot be used and other means of determining a recommended weight limit should be applied.

The NIOSH lifting equation is

$$\text{RWL (Recommended Weight Limit)} = \text{LC} \times \text{HM} \times \text{VM} \times \text{DM} \times \text{AM} \times \text{FM} \times \text{CM}$$

Table 6.2.1 Criteria for determining when the NIOSH lifting equation is *not* applicable
Lifting or lowering with one hand.
Lifting or lowering for over 8 h.
Lifting or lowering while seated or kneeling.
Lifting or lowering in a restricted work space.
Lifting or lowering unstable objects.
Lifting or lowering while carrying, pushing, or pulling.
Lifting or lowering with wheelbarrows or shovels.
Lifting or lowering with high-speed motion (faster than about 30 inches. per second).
Lifting or lowering with unreasonable foot or floor coupling (<0.4 coefficient of friction between the sole of the shoe and the floor).
Lifting or lowering in an unfavorable environment (i.e., the temperature significantly outside the 19–26 °C [66–79 °F] range, relative humidity outside the 35–50% range).

The variables are defined in Table 6.2.2, and Figure 6.2.3.

The multipliers and the task variables (which determine the values of the multipliers) listed in the equation are summarized in Table 6.2.2. Finally, Tables 6.2.3 and 6.2.4 outline the values for the frequency multiplier (FM) and the coupling multiplier (CM).

Figure 6.2.4 illustrates use of an electronic spreadsheet programmed with the NIOSH equation to determine the recommended weight limit (RWL) for a lifting task that involves picking up a box located on a shelf 25 inches high and moving it to a shelf 10 inches higher (35 inches from the floor). The calculations indicate that the RWL is about 40 lb, for this situation. Figure 6.2.5 illustrates the use of the NIOSH equation to determine the RWL for a lifting task identical to the one in Figure 6.2.4, except the new task involves a 45° torso twist. The RWL for the task with twisting is reduced by about 12.5% to 35 lb.

6.2.2.2.3 Biomechanical models

A biomechanical model is a series of calculations that mimic the mechanics of the human body. The models can be used to predict the limits on force or strength exerted by a person while positioned in a specified posture. One widely used computerized model that was developed by the University of Michigan (1995; see Figure 6.2.6) is applicable to lifts that are infrequent and performed slowly.

This model estimates the limit based on two criteria:

1. Estimated percentage of the population that would be capable of exerting strength required to do the task.
2. Estimated compression forces on the spinal disk between the fifth lumbar and first sacral vertebrae in the region of the low back.

The percentage capable and the estimated compression forces can be used as guidelines to assist the practitioner in assessing the MSD risk of a particular task.

Table 6.2.2 Summary of NIOSH multipliers and task variables

Multiplier name	Multiplier abbreviation	Metric equations	U.S. Customary equations	Task variable
Load constant	LC	23 kg	51 lb	None
Horizontal multiplier	HM	$(25/H)$	$(10/H)$	H
Vertical multiplier	VM	$1 - (0.0031V - 751)$	$1 - (0.00751V - 301)$	V
Distance multiplier	DM	$0.82 + (4.5/D)$	$0.82 + (1.8/D)$	D
Asymmetric multiplier	AM	$1 - (0.0032 \times A)$	$1 - (0.0032 \times A)$	A
Frequency multiplier	FM	See Table 6.2.3	See Table 6.2.3	None
Coupling multiplier	CM	See Table 6.2.4	See Table 6.2.4	None

H = horizontal distance of hands from midpoint between the ankles. Measure at the origin and the destination of the lift (cm or inches).
V = vertical distance of the hands from the floor. Measure at the origin and destination of the lift (cm or inches).
D = vertical travel distance between the origin and the destination of the lift (cm or inches).
A = angle of asymmetry, angular displacement of the load from the sagittal plane. Measure at the origin and destination of the lift (degrees).
F = average frequency rate of lifting measure in lifts/min. Duration is specified <1 h, <2 h, or <8 h, assuming appropriate recovery allowances.
C = coupling factor from Table 6.2.4.

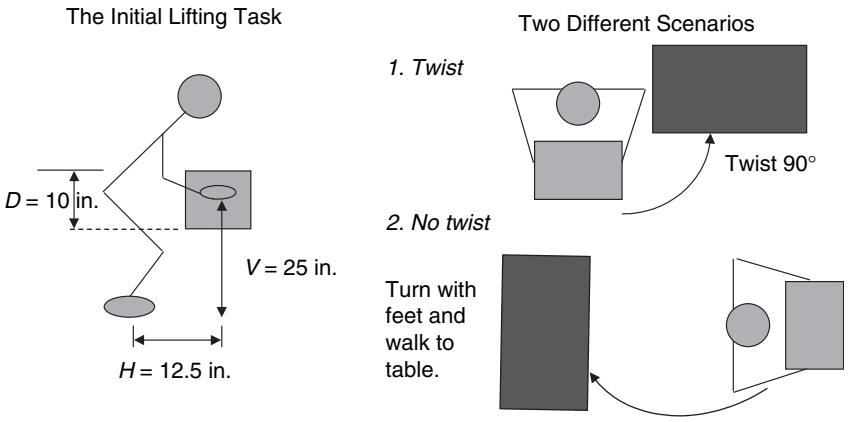


Fig. 6.2.3 Demonstration of the use of the NIOSH (1991) lifting guidelines.

Table 6.2.3 Frequency multipliers

Frequency (lifts/min)	Work duration					
	<1 h		<2 h		<8 h	
	V < 75	*V* ≥ 75	*V* < 75	*V* ≥ 75	*V* < 75	*V* ≥ 75
0.2	1.00	1.00	0.95	0.95	0.85	0.85
0.5	0.97	0.97	0.92	0.92	0.81	0.81
1	0.94	0.94	0.88	0.88	0.75	0.75
2	0.91	0.91	0.84	0.84	0.65	0.65
3	0.88	0.88	0.79	0.79	0.55	0.55
4	0.84	0.84	0.72	0.72	0.45	0.45
5	0.80	0.80	0.60	0.60	0.36	0.35
6	0.75	0.75	0.50	0.50	0.27	0.27
7	0.70	0.70	0.42	0.42	0.22	0.22
8	0.60	0.60	0.35	0.35	0.18	0.18
9	0.52	0.52	0.30	0.30	0.00	0.15
10	0.45	0.45	0.26	0.26	0.00	0.13
11	0.41	0.41	0.00	0.23	0.00	0.00
12	0.37	0.37	0.00	0.21	0.00	0.00
13	0.00	0.34	0.00	0.00	0.00	0.00
14	0.00	0.31	0.00	0.00	0.00	0.00
15	0.00	0.28	0.00	0.00	0.00	0.00
>15	0.00	0.00	0.00	0.00	0.00	0.00

Table 6.2.4 Coupling multipliers

Couplings	*V* < 75 cm (30 inches)	*V* > 75 cm (30 inches)
Good	1.00	1.00
Fail	0.95	1.00
Poor	0.90	0.90

For specific situations, the model estimates the percentage of the population capable and the compression force on the spine, based on the following inputs:

- Size of the person doing the task.
- Angle position of each body joint that determines the posture used to do the task.
- Weight of the object or the force being applied by the hands to the object being handled.
- Description of how the hands apply force to the object (e.g., lifting upward, pushing downward, pushing or pulling horizontally).

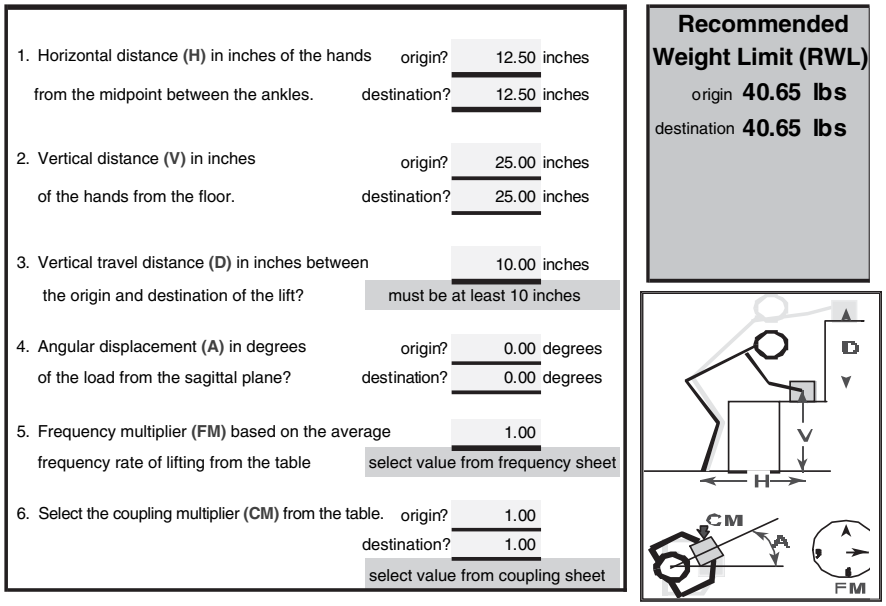


Fig. 6.2.4 NIOSH equation and figure, no twisting.

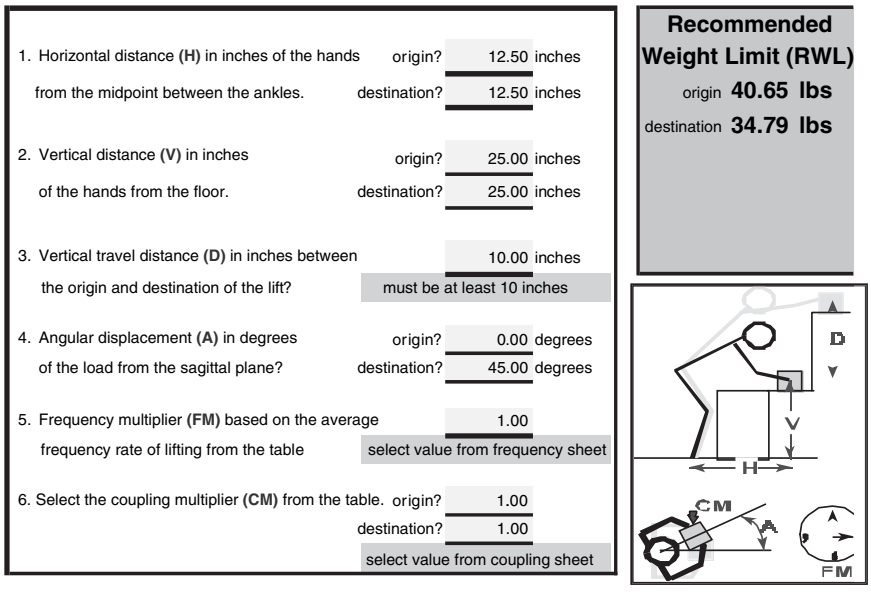


Fig. 6.2.5 NIOSH equation and figure, twisting at the waist.

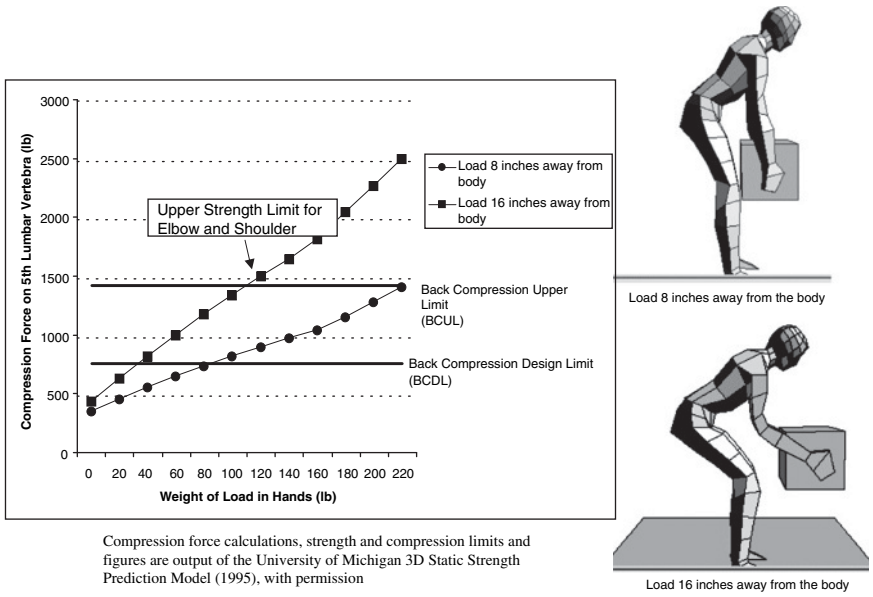


Fig. 6.2.6 University of Michigan model: plot of load (lb) versus compression force (lb) for 8″ horizontal and 18″ horizontal for best lifting posture.

When the percentage capable is less than 99% for men and 75% for women, the force required by the task is categorized as the strength design limit (SDL), which is similar to an "action limit" (NIOSH, 1981). When less than 25% of men and less than 1% of women are capable, the task requires force at the strength upper limit (SUL), which is, in effect, a "maximum permissible limit." If the maximum permissible limit were exceeded, there may be an increased risk of MSD injury.

If the compression force exceeds 770 lb (3400 newton), then the back compression design limit (BCDL), or "action limit," is exceeded. Compressive forces over 1430 lb (6400 Newton) exceed the back compression upper limit (BCUL), or "maximum permissible limit." The SDL and BCDL guidelines were developed to be "consistent with the biomechanical and psychophysical criteria used to develop the NIOSH lifting equation." Whenever the upper limits are exceeded, engineering controls or alterations in work practices are highly recommended. If the design limits are exceeded, decisions regarding workplace changes must be made with regard to

1. The percentage capable of safely performing the task considered.
2. The costs and benefits associated with the introduction of engineering controls.
3. Work practice alterations to accommodate the percentage of the population not capable of performing that task.

The NIOSH equation and the biomechanical models can be used for doing sensitivity (i.e., what if?) analyses on various situations to explore how dimensions of the work area can affect the ergonomic risk factors. Various task parameters, such as lifting height or weight of the load, can be increased or decreased and the equations used to recalculate the recommended weight limit, the percentage capable, and the spinal forces. The recalculated values can then be compared to the original case to determine how much risk reduction can be achieved with various modifications to the work setup.

Other tools also can be used to determine if the force requirement of a manual handling task exceeds recommended limits. Numerous tables of acceptable forces are listed in Ayoub, Mital, and Nicholson (1992) and published papers can be consulted for specific situations (e.g., Woldstad, McMulkin, and Bussi, 1995, for hand wheels). Suggested limits for pushing and pulling a cart, such as a flatbed, can be determined by referring to tabled values of maximum acceptable forces published by Snook and Ciriello (1991). The tables were developed by measuring the maximum acceptable forces that industrial workers chose to exert during pushing and pulling tasks. The maximum acceptable forces are indexed in tables according to height of the hands while pulling or pushing, number of times per minute the pushing or pulling is required, and gender. To determine if a pulling or pushing task requires excessive force, the actual force can be compared to the corresponding maximum acceptable force. To do this, the actual force required to execute the push or pull is measured with a force gauge. More extensive tables for pushing and pulling, as well as lifting, tasks can be found in Ayoub et al. (1992).

6.2.3 Hand-intensive repetitive tasks

For hand-intensive repetitive tasks, the emphasis of the assessment is on identifying which of the five CTD risk factors (force, repetition, awkward posture, contact stress, and muscular fatigue) are significant enough to warrant reduction. In petrochemical operations, some jobs consist of many different tasks, each possibly containing various combinations of the risk factors. Highly repetitive tasks that are repeated hundreds of times per hour are not as likely, compared to other industries such as electronics assembly and word processing.

6.2.3.1 Risk factors

If the task involves any of the following features, then it may pose a risk of MSD. More commonly, more than one of the risk factors is present:

- Reaching frequently reaching behind the body or more than 30 cm (12 inches) away from the body.
- Repetitive or prolonged reaching with hands above shoulder height.
- Repetitive or prolonged work with the hands above elbow height.
- Repetitive or prolonged finger-pinch gripping or hand grasping.
- Repetitive or prolonged rotation of the forearm (e.g., motion while using a screwdriver).
- Repetitive or prolonged awkward wrist posture (see Figure 6.2.7).
- The same or similar hand, arm, or upper body motions repeated frequently.
- Force is exerted by the hands over prolonged periods of time.
- Tool handle or grasped object digs into palm or fingers.
- High force is applied with a finger-pinch grip or a hand grip.
- The force required to manipulate controls, levers, or hand tools is unacceptably high.
- Repetitive pushing of buttons or manipulation of toggle switches on equipment.

Examples of risk factors in an electrician's job:

- Direct contact of tool handles to the palm and fingers while applying high amounts of force to the handle (e.g., utility blade handle, ridges on screw driver handle, short handles of pliers).
- Force from hitting, pushing, and pulling with the butt of the hand (e.g., hitting end of the wrench to turn it).
- Static grip and squeezing motion required to operate hand tool and support the weight of the hand tool with its power cord (e.g., cutting cable, power tools).
- Repetitive twisting motion of the hand and forearm to prep the work area while manually removing protective covers or threaded fasteners as well as replacing them after the electrical work is completed.

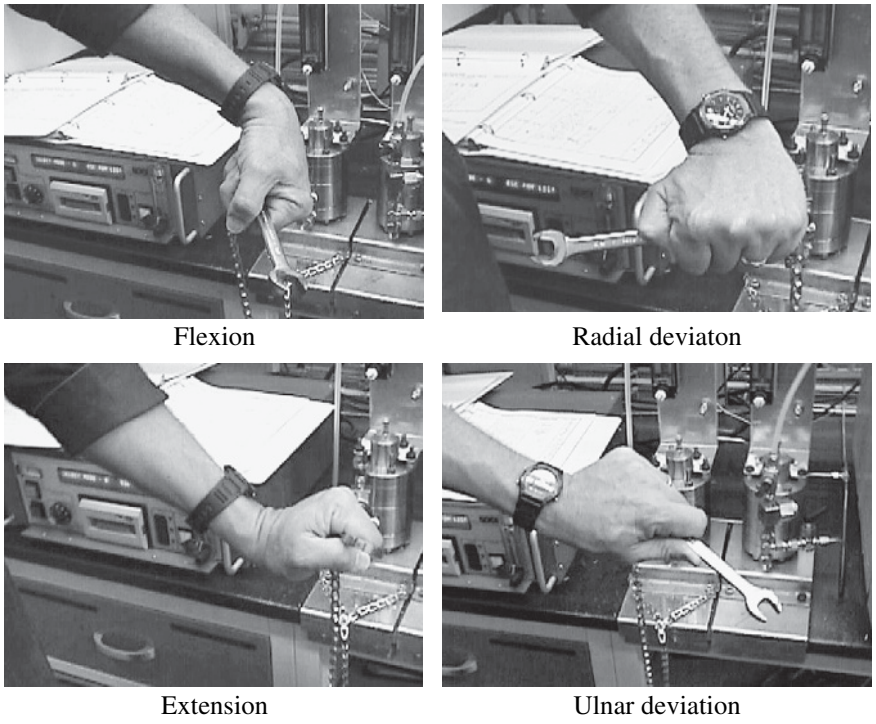


Fig. 6.2.7 Examples of improper wrist posture.

6.2.3.2 Survey and observation tools

More detailed analyses can be conducted by using one of the observation methods for tasks that involve use of the hands and arms. These methods are approaches for ranking tasks according to relative risk and are not measures of risk. In other words, a task scoring 4 according to this method is judged to contain more risk than a task scoring 2, but the task scoring a 4 need not have twice the risk of the task scoring 2.

Many methods are available for scoring tasks, but the best option is to choose a tool shown to be both valid and reliable for detecting risks related to CTDs. A valid tool reasonably identifies the risks related to the five CTD risk factors, the risk factors listed in the previous section. A reliable tool produces relatively similar results when tasks are scored by different raters.

One of these methods is the rapid upper limb assessment (RULA) method, developed by McAtamney and Corlett (1993) for analyzing upper-limb-intensive tasks (see Figure 6.2.8). RULA is easy for an untrained individual to apply to rate priority for tasks involving the hand, arm, and shoulder.

6.2.3.3 Hand tools

Use of hand tools potentially involves the five CTD risk factors as well as potential risk for acute injury, such as a finger pinch. Work areas and hand tools used by personnel with jobs where hand tools are used extensively, such as pipefitting, electrical, mechanical, and shop areas, should be evaluated for features that may contribute either to cumulative or acute injuries. The features of the hand tool that should be evaluated are

- Weight and balance.
- Grip characteristics.
- Trigger or other controls.
- Vibration (described in Chapter 6.3).
- Posture while using the tool.

In general, a tool should be easy to grasp and operate. The grip area in contact with the hand should minimize hard edges. Vibration should be minimized where possible. Heavy tools should be equipped for two-handed use. The following is a summary of tool guidelines based on detailed specifications found in Eastman Kodak Company (1983).

The recommended upper weight of a hand tool is about 4.5 lb (2 kg) with a maximum of 6 lb (2.7 kg) supported with one hand. If the tool is heavier, then provide an external device to counter or support the tool weight. A tool designed to be operated with two hands distributes forces, provides a more stable grip, overcomes torque, and provides the user better control over the tool. The hand or arm muscles should not have to be used to counteract a tool that does not have the weight balanced correctly.

A grip shape that is circular or oval with a diameter of 1.25 – 1.75 inches (3.1 – 4.4 cm) is recommended. If the grip requires a span across a lever (such as the handles of pliers), then the span should be in a range 2–3 inches (5–7.5 cm). All the hand and fingers should comfortably fit onto the grip. Operation of the tool should not require a restricted portion of the finger (such as the tip of the finger) or hand to provide prolonged static pressure on the tool. The thumb and index finger should slightly overlap around a closed grip for maximum strength. The grip is where the hands and fingers apply force to the tool, so it should be covered with a soft material that reduces hard edges and pressure points. The covering should protect the hands from the temperature of the tool (e.g., hot or cold surface) and reduce the slipperiness of the tool in the hand. Some coverings also claim to attenuate vibration exposure to the hand.

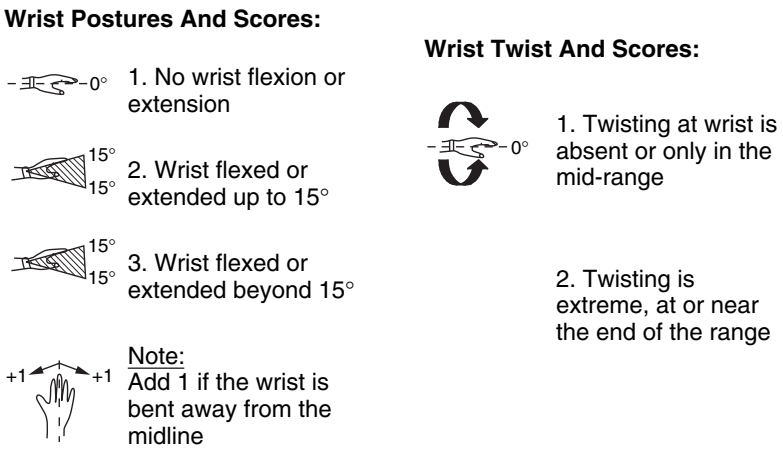


Fig. 6.2.8 Excerpt of RULA observation worksheet for wrist postures (McAtamney and Corlett, 1993).

The manner in which the tool is used also influences the potential for injury. Hand and wrist postures are important indicators. Tools should be able to be used without extreme bending of the wrist or awkward arm posture. The work should be arranged at about elbow height to avoid awkward arm and shoulder posture. The work should also be designed to minimize the need for static posture, such as prolonged gripping with the hands or holding out an outstretched arm. A correct glove fit is also important. Tight-fitting gloves can put pressure on the hands, while loose-fitting gloves reduce grip strength and dexterity.

6.2.4 Behavior

Individual work tasks can, of course, be assessed as needed with one or a few of the tools described in the previous sections of the chapter. The actual challenge of assessing work tasks, however, is to assess multiple work tasks and relative risk so that intervention efforts and resources can be focused at high-priority conditions—the aspects of work tasks identified as the highest potential contributors to a MSD injury.

The procedures described in this chapter are used to assess the risk associated with job and task characteristics. Risks associated with the work environment, human behavior, and management systems can be identified and addressed with other tools.

Conditions of the work environment that could contribute to a MSD injury, such as a back injury resulting from a slip, trip, or fall, include good housekeeping and adequate illumination of work areas. Examples of environmental issues that could contribute to an injury are

- Poor housekeeping (e.g., obstacles, such as hoses, left on the floor or in areas traversed by personnel carrying loads).
- Slippery walking surfaces (e.g., ice, rain, or slick substances).
- Dark areas, where the walking surface is not adequately illuminated.

These types of issues should be identified during safety inspections, job safety analyses, and task risk assessments.

Individuals can contribute to controlling MSD risk by proper behavior while doing physical tasks. Human behaviors that can help prevent MSD injuries are, but not limited to,

- Proper lifting technique.
- Proper pushing and pulling technique.
- Use of lift assist devices when indicated.

Proper lifting technique involves.

- Planning the lift.
- Positioning the feet.
- Using good posture.
- Getting a firm grip.
- Keeping close to the load.
- Lifting the load smoothly.
- Moving the feet.
- Putting down, then adjusting the load.

Proper pushing and pulling technique involves.

- Where possible, pushing instead of pulling.
- Where possible, locating the hands at about elbow height when standing erect.
- Staggering the feet, one foot forward.
- Not using sudden movements.
- Leaning slightly forward.

The results of studies on the effectiveness of back belts in preventing back injuries are equivocal, including the most recently published ones (Krause et al., 1996; NIOSH, 1994; Magnusson and Pope, 1996). The use of back belts as injury prevention devices should not be considered a substitute for a comprehensive ergonomics program. To prevent back injuries, it is necessary to train individuals to lift properly, provide adequate engineering controls, and ensure that injuries are reported early, so they can be treated and the situation ameliorated before any further, more serious, injury might occur.

To address behavioral risk factors, the site behavior-based safety program can incorporate critical behaviors that help prevent MSDs. In this type of program, critical MSD injury prevention behaviors can be identified and evaluated as safe or at risk during routine observations conducted as part of the program and are addressed as needed. As an example, for one program, the observers watch for the following critical lifting, bending, and twisting behaviors during a safety observation of warehouse employees:

- While lifting, is the employee using primarily the legs, keeping a curve in the back, the breastbone facing forward, and the load close to the body?
- Are the shoulders in line with the hips?
- Is the load held isometrically with the arms, as much as possible?
- Is the lift balanced and controlled?
- Does the employee avoid lifting or moving heavy or bulky objects without assistance?
- Is the load held close to the body?
- Does the employee keep frequently lifted material at a level between the waist and shoulders?
- While bending, is the employee keeping a curve in the back, and shoulders in line with hips?
- Does the employee bend from the hips or knees rather than from the waist, using a wide base of support as much as possible?
- Does the employee avoid bending where possible?

- Does the employee avoid twisting the spine at all times, keeping the shoulders in line with the hips?
- Does the employee turn by moving the feet and taking steps?

6.2.5 Ergonomics program

An ergonomics program provides a process to assess, evaluate, and control the risks associated with work-related MSD injuries. The following basic principles are recommended as a basis for any ergonomics program:

- Ergonomic processes are supported at every level of the organization.
- Ergonomic processes are proactive, designed with input from all levels of the organization and aligned with the culture, complexity, and level of risk of the site.
- Ergonomics applications are integrated into corporate safety practices, standards and management systems.
- Management is actively involved through participation in goal setting, training, and stewardship.
- Ergonomic processes support risk identification, assessment and interventions to reduce risks.
- The program complies with local ergonomics regulation.

The ergonomics program recommended in this chapter is a specific application of the general program model. There are other sources of recommendations for ergonomics program structures, such as those by NIOSH (1997), which grew from the proposed ergonomics regulations in the United States.

6.2.5.1 Planning an ergonomics program

The stages of ergonomics program implementation are planning, training, assessment, interventions, control, and continuous improvement.

Planning the program involves the following steps:

1. Develop the site plan and documentation (e.g., reference materials, manual).
2. Develop the training package.
3. Identify key site personnel and contacts.
4. Define the goals of the ergonomics program.

Training, awareness, and expertise in ergonomics should be developed and maintained. This includes the initial orientation and awareness training and periodic ongoing and refresher training for managers, supervisors, professionals, and other employees as appropriate. Training can be integrated into existing or ongoing safety and health training courses.

The level of training depends on the employee's role in the program. (Table 6.2.5 is an example of assigned roles and responsibilities for an ergonomics program.) General awareness training for all employees should include

- Common MSDs, signs and symptoms.
- The importance of reporting MSDs early.
- How to report MSDs, signs, and symptoms.
- Risk factors, jobs, and work activities associated with MSD hazards.
- Proper lifting and material handling techniques.
- A description of relevant local regulations and standards.

Safety meetings may be used to provide supplemental job specific ergonomic training, such as use of checklists, worksheets, and calculations. Management, supervisors, medical, and safety directors and the site skilled contacts should receive training on risk identification, setting risk priorities, and assessment. The site skilled contacts or subject matter experts should receive training on how to use detailed risk assessment tools for ergonomics agents (e.g., how to assess lifting, pushing, pulling, and upper extremity repetitive tasks). The project engineers and design specialists should receive training on how to consider minimizing design features that affect ergonomic hazards or risks.

6.2.5.2 Risk assessment process

The risk assessment process should cover the following:

- Reactive (e.g., response to incidents) and proactive (preventative) risk assessment.
- Screening using the following data to set priorities on tasks that need assessment: safety incident data, health illness data, absenteeism data, behavior-based safety observations on ergonomics (e.g., following lifting procedures, using proper lifting technique).
- Detailed assessments (e.g., use of NIOSH lifting equation, biome-chanical model) on specific high-risk tasks.
- Review of projects for new facilities or modified plant areas.

There are two levels of risk assessment for proactively assessing existing operations: screening and detailed. The first level is a screening process that may be used to proactively assess a specific job or task for potential risk factors to prevent injuries. The screening process is used to qualitatively assess a job or task, identify obvious concerns or conditions that can be quickly fixed, and assess priorities on interventions to reduce risk or initiate further detailed assessment.

When using these tools, a higher score indicates that a task contains a relatively high number of key MSD risk

Table 6.2.5 Example roles and responsibilities for an ergonomics site program

Management sponsor	Ensure implementation, execution, and maintenance of program Participate in an annual review of the program, facilitating improvements to the program and manual as necessary Assure that the management and local program manual administrators are properly trained and qualified Identify the responsible and accountable resources (administrator and skilled contacts) and ensure they are trained
Program administrator	Oversee the program Conduct periodic evaluations of the effectiveness of the ergonomics control program for the management sponsor Periodically review the program manual for necessary changes Assure that applicable personnel are properly trained Assure appropriate records are maintained, including key performance indicators Notify the employee and his or her supervisors of the results of the workplace evaluations
Line supervisor	Assure the program is implemented for their work areas Promote participation in the program Ensure proposed changes that affect the program are addressed through the management of change process Consider ergonomic principles when designing or reviewing job tasks Ensure employees receive ergonomics training Present job-specific ergonomics topics at safety meetings Promptly refer employees to medical personnel if they report symptoms and initiate incident investigation procedures as necessary Request workplace ergonomic evaluations as needed Support implementation of interventions and controls
Site skilled contact	Implement the program at each field location Ensure site-specific procedures are written Ensure worker training is conducted Maintain records and training Arrange general awareness training for site personnel Evaluate work environment or job tasks as indicated by risk assessment or referral (e.g., after an incident) Identify, prioritize, and assess risks Work with other personnel to develop and implement controls Report key performance indicators Serve as the subject matter expert (SME) for field-related questions Review injury, illness, and incident reports for ergonomic trends
Health clinic or medical personnel	Evaluate persons reporting discomforts or concerns with MSD signs and symptoms Report ergonomics injuries and illnesses to safety and management personnel for workplace evaluation follow-up Evaluate employee's fitness for work as needed
Trainer	Maintain documentation of ergonomics training
Ergonomist	Assist with development of training packages Train site contacts and other key personnel to assess risks Assure quality of data and analyses Provide technical expertise as needed Assist with intervention and controls as needed
Site engineer	Assist development of controls Include ergonomic principles in new design and modification projects Ensure project employees receive ergonomics training Include ergonomics considerations in design execution, including design reviews
Employees	Attend and participate in scheduled ergonomics training Use tools and equipment as instructed Use proper work practices as instructed in ergonomics training Report symptoms of injury or illness at their first appearance Report workplace hazards or conditions

factors. Management should consider these jobs or tasks at the highest priority for risk reduction. High- or moderate-risk tasks that have no simple or quick solution are quantitatively assessed with the best exposure assessment methods available. Tools such as simple checklists, postural observation, and task analysis worksheets are appropriate for screening and have already been described in this chapter. Data such as injury reports, medical, management, and employee concerns should also be considered in the screening process to assign priority to tasks for follow-up.

The second level is a detailed assessment process, which is used when quick fixes to high or moderate risks are not immediately evident or the identified solutions are considered costly or resource intensive. Detailed assessments take more effort and skill to conduct, but they provide quantitative insights that can help target the intervention for effective use of resources. The detailed assessment process is quantitative where possible and based on the best-available exposure assessment methods. Quantitative assessment provides insight on which portions of a job or tasks contribute the most risk, therefore guiding the cost-effective solution that most effectively reduces risk. In addition, quantitative assessments can be used to conduct sensitivity analyses, in which the tasks or jobs can be modeled with the proposed changes and the risk reduction estimated before any design changes actually take place. This process is illustrated in Figure 6.2.9.

More comprehensive analysis approaches encompass risk identification of factors beyond the physical risk factors. One approach, described by the Canadian Association of Petroleum Producers (2000), involves the use of a series of forms that guide the risk identification and assessment process. Several potential risk factors are reviewed on the forms, including a signs and symptoms questionnaire and posture, force, repetition, organizational, and environmental factors. The scores on each form are summarized and the tasks are categorized into low-, medium-, and high-risk categories based on the aggregate task score.

Another comprehensive approach is by NIOSH (1997) as part of proposed ergonomics regulations. This approach is based on criteria, called *triggers*, that initiated aspects of an ergonomics program based on certain conditions.

6.2.5.3 Solutions

Timely, successful solutions, regardless of size, build credibility, which is the foundation on which to support a longer-term process. In the solution development phase of the project, the emphasis is to "design in" ergonomic improvements in selected workplaces and implement appropriate engineering, work-practice, and administrative controls. These solutions take the form of modifying equipment, tools, or materials; changing work procedures; or modifying work schedules, where appropriate, by distributing difficult tasks among employees and shifts or job rotation.

In some cases, simple solutions are available through modifications in purchasing specifications, procedural changes, and the like. In other cases, solutions are not as obvious and require further development. Where solutions are costly or complex, it may be useful to request input from other similar operations, consultants, or equipment designers.

Engineering controls are physical changes to a job that reduce MSD hazards. Examples of engineering

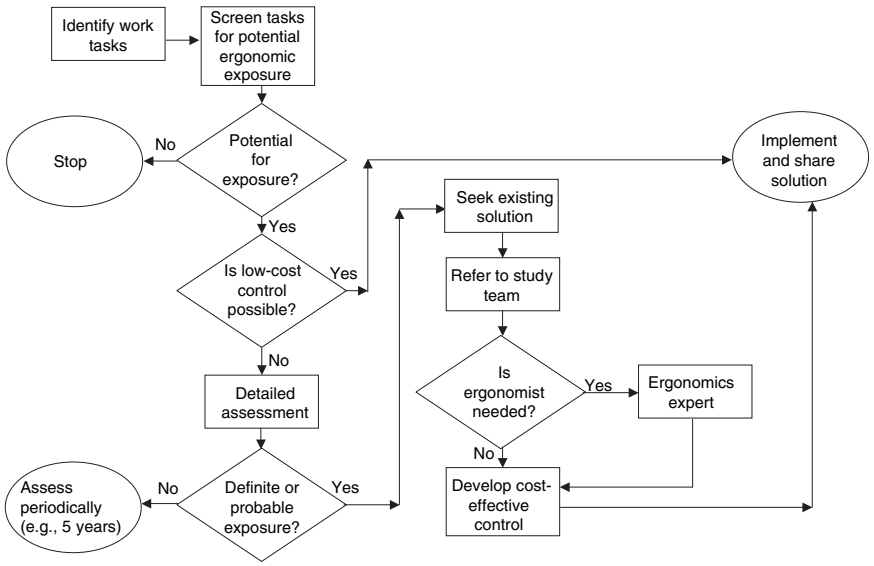


Fig. 6.2.9 Flowchart on risk assessment process.

controls include changing or redesigning workstations, tools, facilities, materials, or work processes. Examples of tools that reduce hazards include

- Box tables to raise loads.
- Tools to move loads (pallets, drums, boxes, or furniture).
- Hoists to assist lifting.
- Scissor lifts to keep loads at an appropriate height.
- Tilt tables to improve access.
- Portable stairs to prevent stretching.

Administrative controls are changes in the way work in a job is assigned or changes to schedules that reduce the magnitude, frequency, or duration of exposure to ergonomic risk factors. Examples of administrative controls include

- Rotating employees in positions to vary physical activity.
- Recombining jobs or tasks to adjust the physical workload.
- Providing alternative tasks.
- Changing the work pace (if the work pace is not under the control of the employee, such as a conveyor belt with product moving past).

6.2.5.4 Evaluating the ergonomics program

The ergonomics program should be assessed periodically to evaluate its effectiveness and identify improvement areas. Some indicators of effectiveness may include

- Percentage of incidents related to MSD injury.
- Evaluation of local work area.
- Percent training completed.
- Relative percentage safe and at-risk behavior observed.

The program evaluation checklist in Appendix 3 can used to conduct a gap assessment on existing programs.

6.2.6 Case study

This is a study on how to apply the musculoskeletal risk identification tools available from the Canadian Association of Petroleum Producers (CAPP, 2000). The risk identification process that it recommends consists of two levels: (1) screening and (2) priority setting and detailed assessment.

The CAPP provides a series of worksheets that support the screening and detailed assessment processes. The CAPP worksheets are used to screen and assess a series of work tasks to change out the catalyst in a reactor. The work tasks are

- Remove the 50-lb nuts from each bolt that secures the cover at the top of the reactor, then replace them and close the cover when finished changing the catalyst. This task requires hoisting an impact gun to about waist height and reaching upward to support it while it rests vertically over the nut to loosen and tighten it. The handles of the impact gun are above shoulder height.
- While sitting at the bottom of the reactor, push a long pole vertically up through the individual tubes within the reactor to loosen the catalyst so that it falls out of the reactor. Each push with the rod requires about 30 lb of upward force with one hand. The rod handle is a contact stress to the hand.
- While sitting near the top of the reactor, use a pneumatic drill to remove the caked catalyst from each tube within the reactor, so that the drilled pieces fall through the tubes to the bottom floor. Pushing the drill through the caked material requires about 20–35 lb of force depending on the psi used for the drill.
- Move the drums of catalyst to the top of the reactor—400-lb drums of catalyst are moved from an elevator to the reactor area. Drums are staged at a height of about 41 inches from the floor.
- Pour the new catalyst in the drums into the top of the reactor by pushing with 95 lb of force on the drum to tilt it so the contents pour into the reactor.

The level 1 screening process worksheet is shown in Figure 6.2.10.

The CAPP screening worksheet indicates a lower potential of issues associated drilling out the used catalyst and pouring in the new catalyst. The screening worksheet indicates that detailed assessments may be useful on the other tasks to help further assess risk. Level 2 detailed assessments consist of biomechanical analyses of the removal of the reactor cover, prodding the catalyst, and moving the drums.

Biomechanical assessment indicates that the recommended weight limit for the nuts on the reactor is about 44 lb for the current height of about chest height. A platform is installed to bring the operator up in height so that the work is at about 31 inches from the floor, resulting in a recommended limit of 50 lb. Loosening the catalyst by prodding the reactor with a vertical rod requires moderate shoulder and elbow strength and has no excessive compression force on the spine. For handling the drums, biomechanical assessment indicates that most people lack the shoulder strength to safely perform this task, although compression forces on the spine are not excessive. Drum lifting and handling equipment is acquired to minimize the amount of manual drum movement required to maneuver the drums over to the reactor.

	Task name	Perceived risk						Frequency and Duration Task Is Performed
	In distinct work activity comprised of several steps or actions	Moderate or Severe Body Part Discomfort: (Y/N)	Awkward Work Postures? (Y/N)	High Effort or Force? (Y/N)	High Repetition or Work Rate? (Y/N)	Contact Stress or Skin? (Y/N)	High Mental Stress? (Y/N)	1/month: 1/week: <1/week 1/day, and 1/hour
1.	Remove /replace reactor head	N	Y	Y	N	N	N	1/month
2.	Loosen catalyst with rod	N	Y	Y	N	Y	N	1/month
3.	Drill caked catalyst	N	N	N	N	N	N	1/month
4.	Move drums	N	N	Y	N	N	N	1/month
5.	Tilt drums to pour contents	N	N	N	N	N	N	1/month
6.								

Fig. 6.2.10 Level 1 summary form from CAPP risk identification process.

Appendix

Appendix 6.2.1 Manual handling screening checklist: Risk identification and priorities (Deeb, 1994)

Work Location _____ Date _____

Task Description

_____

Task Code _____

Assessed by _____

Please answer the questions in all three sections. You can have multiple answers to some questions. An asterisk next to a question means that multiple answers are possible. Please answer *all* questions to the best of your ability, leave no question unanswered. If you are unsure about an answer to a question, go to highest answer category. For example, in Question 2.3, if the object weighs 30lb (14kg), choose answer number 4.

Name: _____ Age: _____

Height: _____

Weight: _____ Male/Female

Left/Right Handed

For Authorized Use Only
SCORE

Section 1

1. Task Characteristics

1.1. The operator performs the task:

1. Sitting 2. Standing 3. Squatting or kneeling

Other comments:

1.2. When performing the task, the upper body is: (multiple answers are possible)*

1. Straight/erect 2. Bent forward 3. Bent backward 4. Twisted

Other comments:

1.3. Performing the task involves: (multiple answers are possible)*

1. Normal arm reaches
2. Awkward postures, such as deviation from stable body position
3. Reaching above shoulders
4. Uneven distribution of body weight on the feet, such as one foot is lower than the other

Other comments:

1.4. The task is performed at the pace of the operator

1. Yes
2. No

Other comments:

1.5. Performing the task involves: (multiple answers are possible)*

L 1. Lifting/lowering (answer 1.6, 1.7, 1.8, and go to Section 2)
P 2. Pushing/pulling (answer 1.9, 1.10, and go to Section 2)
C 3. Carrying (answer 1.11 and go to Section 2)

Other comments:

For Authorized Use Only
SCORE

L 1.6. This question has two parts, please answer Part A and Part B individually:

A. The initial lifting/lowering height is at
   1. Waist or elbow height
   2. Near the floor
   3. Above the shoulders

B. The location of object in relation to the worker's body is:
   1. Close to the front of the body
   2. Far from the body

Other comments:

L 1.7. The object is moved (lifted/lowered): (multiple answers are possible)*

1. At the same level
2. From waist to shoulder level/shoulder to waist level
3. From floor to waist level/waist to floor level
4. From floor to shoulder level/shoulder to floor level

Other comments:

L 1.8. This question has two parts, please answer Part A and Part B individually:

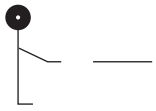
A. The task of lifting/lowering objects is performed
   1. For less than half an hour at a time
   2. Between half an hour and 1 h at a time
   3. Between 1 h and 2 h at a time
   4. More than 2 h

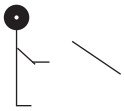
B. The frequency of lifting/lowering of objects is performed
   1. Less than once per minute
   2. Between one and six times per minute
   3. More than six times per minute

Other comments:

P 1.9. In what direction is the pushing/pulling applied?

1. In fornt of the body 2. Across the front of the body





Other comments:

P 1.10. The force applied to push/pull the object is sustained for

1. Less than 10 s
2. Between 10 and 20 s
3. Between 20 and 30 s
4. More than 30 s

Other comments:

For Authorized Use Only
SCORE

C 1.11. Carrying the object requires a holding grip or position for
1. Less than 5 s
2. Between 5 and 10 s
3. More than 10 s

Other comments:

1.12. The task (please answer parts A, B, C, and D individually)
A. Requires unusual capability to be performed (i.e., beyond expectations of the normal worker)
1. No
2. Yes
B. Aggravates those with a health problem (i.e., back and muscular problems)
1. No
2. Yes
C. Aggravates those who are pregnant
1. Not applicable (for men)
2. No
3. Yes
D. Requires special training/individual received training
1. No/not applicable
2. Yes/yes
3. Yes/no

Other comments:

Total of Section 1

Section 2

2. Material/object characteristics:
2.1. The type of object can best be described as a
1. Bucket
2. Box
3. Bag
4. Other (e.g. hose, drum, pump)_____

Other comments:

2.2. The geometric shape of the object can best be described as
1. Small
2. Medium
3. Large, bulky, or unwieldy

Other comments:

2.3. The weight of the object handled is
1. Below 10 lb (4.5 kg)
2. Between 10 and 20 lb (4.5 and 9 kg)
3. Between 20 and 30 lb (9 and 14 kg)
4. Between 30 and 40 lb (14 and 18 kg)
5. Between 40 and 60 lb (18 and 27 kg)
6. Above 60 lb (27 kg)

Other comments:

For Authorized Use Only
SCORE

2.4. The object's load
1. Is handled using one hand (e.g., bucket, hose)
2. Is shared equally between the hands
3. Is not shared equally between the hands (i.e., one side is heavier than the other)
4. Can shift suddenly (i.e., unstable contents such as liquid, powder)

Other comments:

2.5. The object has
1. Handles
2. No handles

Other comments:

2.6. The area of contact between hands/arms and object
1. Presents no hazard/difficulty
2. Is cold/hot
3. Has sharp edges

Other comments:

Total of Section 2

Section 3

3. Workplace characteristics

3.1. The task is performed in
1. An unrestricted space
2. A restricted space (places constraints on posture)
3. A very restricted space (places severe constraints on posture)

Other comments:

3.2. The task is performed in
1. A closed environment with no strong air movement
2. An open environment with strong air movementA

Other comments:

3.3. The ambient temperature and humidity of the workplace is
1. Within the comfortable zone (temperature between 19° and 26 °C [66° and 79 °F] and humidity between 35% and 65%)
2. Outside the comfortable zone (hot/cold/humid conditions)

Other comments:

3.4. The workplace is considered to have (consider day and night work)
1. Adequate lighting for the job
2. Inadequate lighting for the job

Other comments:

3.5. The floor surface (multiple answers are possible)*
1. Is even
2. Has variation in levels (i.e., uneven, steps)
3. Is slippery

Other comments:

3.6. The workplace has
1. Good housekeeping
2. Poor housekeeping, such as crowded or waste on the floor

Other comments:

3.7. The conditions where manual materials handling tasks are performed require
1. Standard protective clothing
2. Additional protective clothing (e.g., chemical suits, respirators)
3. Special protective clothing such as thermal coverall and gloves

Other comments:

Total of Section 3

For Authorized Use Only
SCORE

Summary

Relative Scores

	Total (max)	% Severity
Section 1	(116)	
Section 2	(22)	
Section 3	(20)	
Total	(158)	

Comments/recommendations:

Appendix 6.2.2 Musculoskeletal ergonomics program gap analysis checklist

Checklist Item	Confirmed (✓)
Program Feature	
Implementation	
The program is implemented throughout the site	
Employees/local unions support the program	
A system is in place for employees to report concerns	
Feedback is given to employees regarding reported concerns	
Concerns are addressed in a timely manner	
The program includes the following components:	
Planning	
Selection/priority assignment of worker groups for screening	
Screening for detailed assessment	
Detailed risk assessment	
Training for the general population and assessors	

Continued

Checklist Item	Confirmed (✓)
Control development	
Measurement/continuous improvement	
Changes to tasks triggers re-evaluation	
Members of worker group are interviewed	
Jobs are divided into tasks for assessment	
Training and qualifications of ergonomics resources are commensurate with role and level of involvement	
A program is in place to address office ergonomics	
The program is part of everyday practice/awareness	
Management Support	
The program has visible support by site management and line management	
Roles in the program are clearly defined	
Enough personnel to effectively implement and sustain the program	
Enough time to effectively implement and sustain the program	
There is enough budget to effectively implement and sustain the program	
The program is linked to other site systems such as incident investigation, near miss reports, procedures development and training, risk assessments, management of change and medical system	
Training	
All employees are trained with an introduction to ergonomics	
Supervisors/management/medical directors trained in the screening process (overview)	
Training includes consideration for equipment design that affects musculoskeletal ergonomics	
Refresher training scheduled as needed	
Training given new employees	
Verification method in place to determine if training is adequate and understandable	
General awareness training includes identification of symptoms and signs	
Training addresses procedures for reporting work-related musculoskeletal disorders and concerns	
Workers who change program roles have sufficient training to fill new role	
Training enables managers, supervisors, and employees to identify aspects of a job that increase risk of developing work-related musculoskeletal disorders	
Assigning Priorities	
Jobs assigned priority prior to the screening process	
Symptom surveys used across the site to assess trends	
Health/illness data used to assess trends and set priorities	
Absenteeism data used to assess trends and set priorities	
Safety incident data used to assess trends and set priorities	
Communication with similar sites used to identify potentially high-risk tasks	
Outcome of employee concerns are used in assigning priorities	
Screening Tools	
Screening tools are easy to use and understand	
Screening tools identify tasks that require detailed assessment	
Person conducting screening process is qualified	
Screening based on medical data and concern of management, employees, and evaluator	
All field jobs are screened based on prioritization process	

Checklist Item	Confirmed (✓)
Detailed Assessment Tools	
Whole body posture assessed using a standardized tool	
Tool/method: _____	
Upper extremity exposure assessed using a standardized tool	
Tool/method: _____	
Vehicle operation assessed using a standardized tool	
Tool/method: _____	
Lifting and lowering assessed using a standardized tool	
Tool/method: _____	
Whole body pushing and pulling forces assessed using a standardized tool	
Tool/method: _____	
Energy expenditure assessed using a standardized tool	
Tool/method: _____	
Results for each tool mapped to an action level rating to identify relative risk	
Trained evaluator observes the worker performing the task	
Tasks are videotaped for later detailed analysis	
Workstation measurements taken (surface heights, clearance space, etc.)	
Forces and hand positions measured	
Exposure to heat, cold, and vibration considered	
Repetition and duration of task considered	
Hand tool usage assessed	
Coefficient of friction on walking surface (slippery floor) considered	
Person conducting the detailed assessment is qualified	
Intervention and Controls	
Solutions evaluated with considerations of relative risk and relative cost/benefit analysis	
Effectiveness of solutions implemented measured for musculoskeletal risk factors and work efficiency	
Solution development documented	
Solutions developed and implemented in a timely manner, depending on severity of issue	
Solutions and learning effectively communicated throughout the organizational levels and shifts/work groups, across similar sites, and a central resource	
Employees empowered to make changes to their work environment	
Employees empowered to make suggestions for changes to their work environment	
Solution evaluated before implementation by a qualified skilled resource	
Control strategies reported through management of change procedures	
Changes in work procedures or required work effort (force, shift length, container size, etc.) trigger re-evaluation	
Assessed tasks given priority for improvement according to relative risk	
Continuous Improvement/Measurement	
The program stewarded on a regular basis	
The program has been continuously improved	
Benchmarking data (i.e., health illness data, absenteeism data) collected	
Measures of program performance in place and include safety incident data, health illness data, absenteeism data, and restricted work.	
Internal and external quality assurance processes in place to evaluate data and the program	

6.2.7 Review questions

Test your understanding of the material in this chapter.

1. Which of the following is not a type of cumulative trauma disorder?
   (a) Low back pain.
   (b) Broken wrist.
   (c) Carpal tunnel syndrome.
   (d) Tendinitis.
2. What five risk factors are associated with cumulative trauma disorders?
3. How would you determine whether the weight of a box of spare parts is too heavy to be safely lifted by the warehouse attendant?
4. Which dimension, distance of the load away from the body or vertical location of the load, is the more significant contributor to the effect of the load?
5. What general instructions would you give someone on how to lift safely?
6. How would you identify which tasks need a detailed ergonomic assessment?
7. What topics should be covered in a general employee awareness course on preventing musculoskeletal injuries?

References

Ayoub, M. M., Mital, A., and Nicholson, A. S. (1992) *A Guide to Manual Materials Handling*. London: Taylor and Francis.

Canadian Association of Petroleum Producers (CAPP) and Canadian Petroleum Products Institute (CPPI). (2000) *Ergonomic Risk Identification and Assessment Tool*. Calgary, Alberta: CPPI.

Deeb, J. (1994) Personal Communication.

Dempsey, P. G., Burdorf, A., Fathallah, F., Sorock, G., and Hashemi, L. (1991) "Influence of Measurement Accuracy on the Application of the 1991 NIOSH Equation." *Applied Ergonomics* 32, pp. 91-–99.

Eastman Kodak Company. (1983) *Ergonomic Design for People at Work*. New York: Van Nostrand Reinhold Company.

Hignett, S., and McAtamney, L. (2000) "Rapid Entire Body Assessment." *Applied Ergonomics* 31, pp. 201–205.

Karhu, O., Harkonen, R., Sorvali, P., and Vepsalainen, P. (1981) "Observing Working Postures in Industry: Examples of OWAS Application." *Applied Ergonomics* 12, no. 1, pp. 13–17.

Karhu, O., Kansi, P., and Kuorinka, I. (1977) "Correcting Working Postures in Industry: A Practical Method for Analysis." *Applied Ergonomics* 8, no. 4, pp. 199–201.

Kivi, P., and Mattila, M. (1991) "Analysis and Improvement of Work Postures in the Building Industry: Application of the Computerized OWAS Method." *Applied Ergonomics* 22, no. 1, pp. 43–48.

Kraus, J. F., Brown, K. A., McArthur, D. L., Peek-Asa, C., Samaniego, L., Kraus, C., and Zhou, L. (1996) "Reduction of Acute Low Back Injuries by Use of Back Supports." *International Journal of Occupational and Environmental Health* 2, pp. 264–273.

Magnusson, M., and Pope, M. (1996) "Does a Back Support have a Positive Biomechanical Effect?" *Applied Ergonomics* 27, No. 3, pp. 201–205.

Mattila, M., Karwowski, W., and Vilkki, M. (1993) "Analysis of Working Postures in Hammering Tasks on Building Construction Sites Using the Computerized OWAS Method." *Applied Ergonomics* 24, no. 6, pp. 405–412.

McAtamney, L., and Corlett, N. (1993) "RULA: A Survey Method for the Investigation of Work-Related Upper Limb Disorders." *Applied Ergonomics* 24, no. 2, pp. 91–99.

NIOSH. (1981) *Work Practices Guide for Manual Lifting*. Cincinnati, OH: U.S. Department of Health and Human Services.

NIOSH. (1994) *Workplace Use of Back Belts*. DHHS (NIOSH) Publication No. 94-122. Cincinnati, OH: U.S. Department of Health and Human Services.

NIOSH. (1997) *Elements of Ergonomics Programs*. DHHS Publication No. 97-117. Cincinnati, OH: U.S. Department of Health and Human Services.

Occupational Safety and Health Administration. (1995) *OSHA Proposed Ergonomic Protection Standard*. Washington, DC: Bureau of National Affairs.

Scott, G. B., and Lambe, N. R. (1996) "Working Practices in a Perchary System, Using the OVAKO Working Posture Analysing System (OWAS)." *Applied Ergonomics* 27, no. 4, pp. 281–284.

Silverstein, B. A., Fine, L. J. and Armstrong, T. J. (1987) "Occupational Factors and Carpal Tunnel Syndrome." *American Journal of Industrial Medicine* 11, pp. 343–358.

Snook, S. H., and Ciriello, V. M. (1991) "The Design of Manual Handling Tasks: Revised Tables of Maximum Weights and Forces." *Ergonomics* 34, no. 9, pp. 1197–1213.

University of Michigan, Center for Ergonomics. (1995) *3D Static Strength Prediction Program Manual*. Ann Arbor: University of Michigan.

Waters, T. R., Putz-Anderson, V., and Garg, A. (1994) *Applications Manual for the Revised NIOSH Lifting Equation*. DHHS (NIOSH) Publication No. 94-110. Cincinnati, OH: U.S. Department of Health and Human Services.

Waters, T. R., Putz-Anderson, V., Garg, A., and Fine, L. J. (1993) "Revised NIOSH Equation for the Design and Evaluation of Manual Lifting Tasks." *Ergonomics* 36, no. 7, pp. 749–776.

Wiktorin, C., Selin, K., Ekenvall, L., Kilbom, A., and Alfredsson, L. (1996) "Evaluation of Perceived and Self-Reported Manual Forces Exerted in Occupational Materials Handling." *Applied Ergonomics* 27, no. 4, pp. 231–239.

Woldstad, J. D., McMulkin, M. L., and Bussi, C. A. (1995) "Forces Applied to Large Hand Wheels." *Applied Ergonomics* 26, no. 1, pp. 55–60.

Chapter 6.3

6.3

Environmental factors

Dennis Attwood, Joseph Deeb and Mary Danz-Reece

6.3.1 Introduction

This chapter discusses four environmental factors or features that arise from the surroundings (i.e., illumination and temperature) and mechanical features of the workplace (i.e., noise and vibration) that could be encountered at work and affect behavior. Other examples of environmental features that could be encountered at work and may affect behavior are dangerous chemicals, radioactive materials, atmospheric pressure, and the like. These other environmental features are not discussed in this chapter.

This chapter discusses the four environmental features that can affect people in three ways: (1) health, (2) performance, and (3) comfort. The affects of these three aspects are usually combined. For example, poor health can lead to both poor performance and reduced comfort and, thus, reduced work satisfaction. Stressors, arising from illumination, temperature, noise, vibration, or any other aspect of the environment can adversely affect people when they reach a certain level, although the effect may not be apparent either to the person being affected or an observer. The ideal range for performance and comfort is narrow. Therefore, trying to adapt to conditions outside the ideal range can make people use more effort, which can lead to reduced performance and comfort; for example, a person trying to see fine details when illumination levels are too low or too high.

This chapter ends with a case study examining the lighting quantity and quality in a control room. In addition, a set of review questions are included to help you check your understanding of the material covered in this chapter.

6.3.2 Illumination

Light stimulus is characterized by its intensity and wavelength. The intensity of light is expressed in terms of the amount of luminous flux (energy) it generates. Flux is measured in candlepower (cd) or lumen, where 1 candle (or candela) is equal to 1 lumen. The wavelength of light is measured in terms of nanometer (nm). One nanometer is one-billionth of a meter (10^{-9}). For example, the eye discriminates between different wavelengths in the range of 380 and 780 nm by the sensation of color.

In our daily environment, the amount of illumination (also called *illuminance)* that falls on a surface (e.g., desk) depends on the energy of the light source, its distance from the surface, and the angle of the surface to the light source. Most light falls on a surface perpendicular to the light. Illumination is described in terms of the rate of flux (lumens) produced by the source and the surface area over which it spreads. One lumen per square meter is defined as 1 lux and 1 lumen per square foot is defined as a foot-candle (FC). Illumination meters are usually calibrated in both measures. One foot-candle is about 10.76 lux (there are 10.76 square feet in a square meter). Light may fall on a surface from multiple sources or, for example, as reflections from walls, ceilings, and other objects in the immediate environment. These reflections can be considered individual light sources, and the efficiency of each depends on the angle at which it reflects light on the surface. In summary, the amount of light falling on a surface depends on the:

1. Light source and its luminous intensity.
2. Distance between the light source and the surface.

Ergonomic Solutions for the Process Industries; ISBN: 9780750677042
Copyright © 2003 Elsevier Inc. All rights of reproduction, in any form, reserved.

3. Angle of the light source to the surface.
4. Number of light sources and reflecting sources in the immediate environment.

We do not see the light falling on a surface but we see the light reflected from a surface. Luminance is a measure of the light reflected from the surface and is associated with the subjective sensation of brightness. Keep in mind that, depending on their surface characteristics, different bodies absorb and reflect different amounts and qualities of light. Luminance is expressed in terms of candela per meter squared (cd/m^2).

The subjective sensation of brightness (that is, the luminance qualities of an object) is perceived after the light has stimulated the retinal cells of the eye and the information passed on to the optic cortex in the brain. Then, the person reports the intensity of the object's brightness. Therefore, the intensity is determined by the luminance of the object, the amount of light passing through the eyes and to the brain, and the subjective experiences of the observer. In turn, the subjective experiences depend on the observer's past experiences and the brightness of other bodies in the visual field. This process is illustrated in Figure 6.3.1. This section discusses how illumination affects performance in terms of lighting quantity and quality and provides guidelines on improving lighting in the workplace.

Visual performance in the workplace is primarily a function of the overall quantity and quality of illumination, task factors, age of the individual, glare, and contrast.

6.3.2.1 Lighting quantity

Studies in the literature demonstrate that increasing illumination can certainly lead to an increase in visual performance in terms of productivity and efficiency. However, the increase depends highly on the type of task performed (Barnaby, 1980; Jaschinski, 1982; Knave, 1984) and age of person (Bennett, Chitlangia, and Pangrekar, 1977; Ross, 1978). One must be careful in prescribing higher and higher levels of illumination: a continuous increase in the level of illumination results in smaller and smaller improvements in visual performance, until that performance levels off.

6.3.2.2 Task factors

The level of illumination needed generally depends on the task performed. The more visually intense the task (i.e., smaller details or lower contrast), the higher is the illumination level required, up to a maximum, where performance levels off no matter how much more the illumination is increased. The point at which this leveling off occurs is different for different tasks. Another approach to enhance visual performance is through changing the features of the task; for example, increasing the size of the object seen or the contrast between the object and its background (discussed in Chapter 6.1). The visual tasks that people perform can vary in difficulty from simple ones, such as finding a car in a parking lot at night, to the highly demanding, such as calibrating an instrument. Within this wide range of task difficulties are wide ranges of illumination levels that can help increase performance. For example, Table 6.3.1 presents ranges of illumination levels (in lux) recommended for different types of visual work activities and areas. The more demanding the visual task and the longer it is sustained, the higher is the

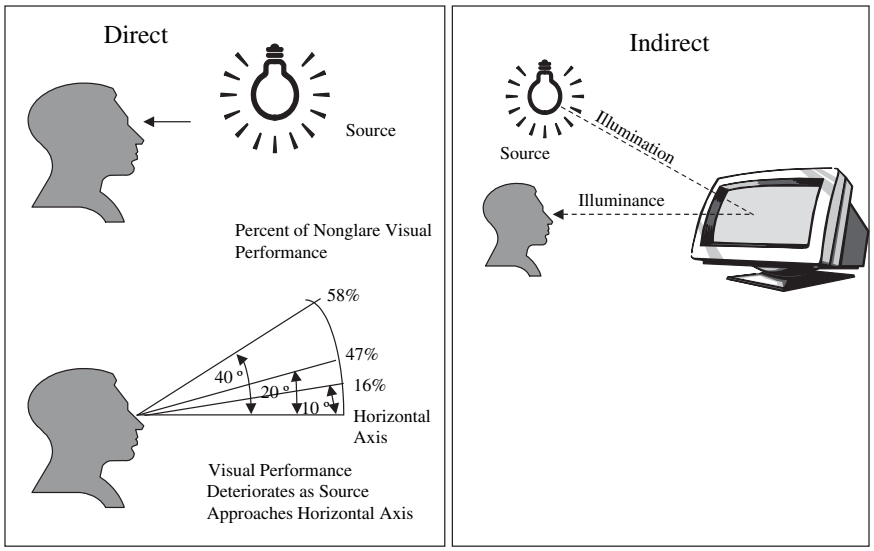


Fig. 6.3.1 Quality of illumination—glare.

Table 6.3.1 Illumination levels for different types of activities

Activity	Illumination range (lux)
Public area navigation	20–50
Occasional visual tasks	100–200
Visual tasks with large objects and high contrast, reading printed material	200–500
Reading poorly printed material with medium contrast or small objects	500–1000
Reading very poorly printed material with low contrast and very small objects for a short period	1000–2000
Reading very poorly printed material with low contrast and very small objects for a long period	2000–5000
Difficult visual tasks such as inspection or fine assembly	5000–10,000
Difficult visual tasks with very low contrast and very small objects	10,000–20,000

Note: To obtain the values in foot-candles, simply divide the values in lux by 10.76; for example 20 lux = 1.86 or about 2 foot-candles.

Table 6.3.2 Illumination levels for selected tasks in process areas of a petrochemical environment

Process area	Illumination level (lux)
Control room	
Desk area and console	325
General	215
Manual sampling points	215
Compressor houses	160
Loading racks	
Loading point	110
General area	32
Platforms	
Main operating area	55
Ordinary area	22
Stairways and ladders (frequently used)	55
Exchanger area	32
Streets	4

Note: To obtain the values in foot-candles, simply dividet; the values in lux by 10.76; for example, 215 lux = 19.98 or about 20 foot-candles.

illumination level required. It is recommended that local task lights be used to provide illumination levels above 1000 lux. Table 6.3.2 lists illumination levels for selected tasks in process areas of a petrochemical environment.

6.3.2.3 Age factors

The aging process has its affects on the visual system. As we grow older, in general, over 40 years old, we experience the following changes (Hughes and Neer, 1981; Wright and Rea, 1984):

- The lens of the eye atrophies and hardens, decreasing its ability to conform to the shape necessary for quick and efficient focus on objects at different distances from the eyes. These changes are also the contributors to the improvement of the far vision (farsighted) and the deterioration of the near vision (nearsighted).
- The muscles controlling the diameter of the pupil also atrophy, reducing the size of the pupil and decreasing the range and speed of adjusting and adapting quickly and efficiently to different levels of illumination.
- A lower level of illumination reaching the retina leads to a reduction in the level and rate of dark adaptation.
- As the lens hardens, it yellows, leading older people to make errors in sorting or matching colors in the blue-green and red regions.
- The changes in both the lens and pupil flexibility and efficiency lead to a decrease in visual acuity; the ability to see fine details.

These changes should encourage us to account for the elderly population in the design of lighting to ensure visual comfort and efficient visual performance. For example, we should provide:

- *More light for elderly people*. This can be achieved through providing adjustable personal light.
- *Higher contrast*. For example, for older people to see as well as people 20–35 years old (baseline), the following multiplication factors are needed to ensure higher contrast and maximize visual performance (Blackwell and Blackwell, 1968; Grandjean, 1988):
  40 years old, 1.17.
  50 years old, 1.58.
  65 years old, 2.66.

Table 6.3.3 Most common lamp types and their color rendering index (CRI), color temperature (K), and life in hours

Lamp type	Color rendering index (CRI)	Color temperature (K)	Life (h)
Incandescent filament lamp	100	2600–3000	1000–2000
Cool white deluxe fluorescent	89	4200	10,000–20,000
High-pressure sodium	22–75	1600	10,000–15,000
Warm white deluxe fluorescent	73	2900	10,000–20,000
Metal halide	65–80	3300–5700	6000–20,000
Cool white fluorescent	66	4200	10,000–20,000
Warm white fluorescent	52	3000	10,000–20,000
Low-pressure sodium	40	1600	16,000
White fluorescent	60	3500	10,000–20,000
Tungsten halogen lamp	100	3000–3400	100–2000
Mercury lamp	15–55	55–60	12,000–24,000
Sunlight at sunrise	—	1800	—
Sunlight at noon	—	5000	—

6.3.2.4 Lighting quality

A number of factors have a direct effect on task visual performance and subjective impressions of people. These are the lighting color, glare, and illumination ratio. Each factor is discussed briefly next.

6.3.2.4.1 Lighting color

The objects in the visual field and the various types of lamps we interact with daily differ in color. In terms of lighting, color is also referred to as *characteristic spectral distribution*. The color of an object is determined by the:

- Wavelengths of visible light that it absorbs and reflects. For example, a red object absorbs most colors in the incident light except red. Incident light is the light falling or striking on something.
- Color of the incident light. Two properties of the incident light determine the resulting color of the object: the color rendering index (CRI) and the color "temperature" of the light.

The color rendering index was developed to quantify the accuracy with which the light source renders (reproduces) the natural colors in an object. The index varies from 0 to 100. As the CRI increases, color judgment errors tend to decrease. Sources with CRIs below 60 should be avoided when accurate color judgment is important (Collins and Worthey, 1985; Wotton, 1986). However, CRIs greater than 80 indicate good color rendering.

Color temperature refers to the numerical assessment of a lamp's color. The numerical assessment is based on a comparative color appearance of glowing of light sources known as *blackbodies*. This is referred to as *color temperature*, expressed in kelvins (K). For example, color temperatures up to 3000 K are "warm" and make reds appear more vibrant. Color temperatures between 3000 and 4000 K enhance most colors equally. Color temperatures greater than 4000 K enhance greens and blues. Table 6.3.3 provides CRI, color temperature, and life values for various types of common lamps.

6.3.2.4.2 Glare

Glare occurs whenever one part of the field of vision is brighter than the level to which the eye has become adapted, causing decrease in visual performance, annoyance, and discomfort. Glare can be characterized as direct or indirect (refer to Figure 6.3.1). Direct glare is caused when the light source appears directly in the field of view (discussed in Chapter 6.1), such as a lightbulb beside a hard to read instrument, the sun during the day, or car headlights, especially high beam, at night. Indirect glare, also referred to as *reflected* or *specular glare*, is caused by bright reflections from

polished or glossy surfaces, such as screens of visual display terminals (VDTs), glossy paper, or white walls. In addition, glare is commonly classified into disability and discomfort glare.

Disability glare occurs when there is direct interference with visibility and visual performance and often is accompanied by discomfort. Discomfort glare produces discomfort, annoyance, irritation, or distraction but does not necessarily affect visibility or visual performance. The distinction between the terms *disability* and *discomfort* is often unclear. The same lighting conditions can produce disability and discomfort glare simultaneously.

Reducing glare is an important ergonomic consideration when designing workplaces. The guidelines presented in Table 6.3.4 should be considered.

6.3.2.4.3 Luminance ratio

The ratio of the brightness (luminance) of the object of interest to the luminance of the surrounding visual field is termed the *luminance ratio*. Our eyes adapt to the luminance of the immediate visual field. The surroundings can have different distribution of luminance and, therefore, create high and low luminance ratios, which can cause transient adaptation problems. Transient adaptation occurs when the visual system moves from a bright to a dark area or vice versa (e.g., entering a dark movie theater on a bright sunny day or coming back outside after the movie). The greater the difference in luminance level (ratio of change), the greater the problem of transient adaptation, which can affect visibility, visual comfort, and performance. An example of large differences in luminance levels is found in the control room. The VDT screen has low luminance level as compared with background windows but about the same luminance as a background wall. So, it is best to mount the VDT against the wall rather than against the window.

Contrast (or relative luminance) is defined by specifying the ratio of the two luminances and is calculated using the following formula:

$$C = (L_t - L_b)/L_b$$

where C = contrast, L_t = target luminance, L_b = background luminance.

The following guidelines are widely accepted:

- The distribution of luminance on the major surfaces in the field of view should be uniform with the working visual field brightness—brightest in the middle and darker toward the edges.
- The luminance ratio (brightness: contrast) between the middle of the visual field and the immediate surroundings should not exceed 1:3, where 1 is the task or object viewed (middle of the visual field) and 3 is the immediate surroundings. The ratio between the middle of the visual field and remote areas should not exceed 10:1.
- The maximum luminance ratio within an office-like setting (includes control room) should not exceed 40:1.

Reducing high luminance ratios can be accomplished by eliminating bright walls, reflective and glossy surfaces, and dark or black furniture, machines, or filing cabinets. For example, Grandjean (1988) recommends the following reflectance for an office environment:

Ceiling, 80–90%.
Walls, 40–60%.
Furniture, 25–45%.
Machines and equipment, 30–50%.
Floor, 20–40%.

Table 6.3.4 Reducing and controlling glare

Reduce or control direct glare	Reduce or control indirect glare
Light source should not be directly in the field of vision or line of sight	Position the light source or work area so reflected light is not in the direction in which the person normally needs to look to achieve maximum visibility
Use shades, hoods, or glare shields to reduce the brightness (luminance) of the light	Avoid using reflective colors and materials, such as bright metal, glass, glossy paper
Use several low-intensity lamps in place of few high-intensity lamps	Use surfaces that diffuse light, such as textured finishes, nonglossy surfaces, flat paint
When using fluorescent tubes, they should be at right angles to the line of sight	Provide several low-intensity lamps in place of few high-intensity lamps and use indirect lighting
Use indirect lighting	The brightness of the lamps (luminaires) should be kept as low as feasible

6.3.3 Temperature

The human body's thermal regulation system tries to maintain a relatively stable internal (core) temperature of between 36.1 and 37.2 °C (97 and 99 °F). The core temperature must stay within a narrow range to prevent serious damage to health and performance. When physical work is performed, additional body heat is generated. If we add relatively high humidity to the ambient temperature, the resulting condition can lead to fatigue and potential health risks.

The human body maintains heat balance by increasing blood circulation to the skin; therefore, we sweat on hot days. When it is cold, the body reduces the blood circulation to the skin and we shiver to keep the extremities warm. The body generates heat through metabolism and physical work. To maintain the internal heat balance, the body exchanges heat (gain or loss) with the environment in four ways:

1. C*onvection*. This process depends on the difference between the air and skin temperatures. If the air temperature is hotter than the skin, then the skin absorbs heat from the air, which becomes heat gain to the body. If the air temperature is cooler than the skin, then the body loses heat.
2. C*onduction*. This process relates to the difference between the temperature of the skin and the surface contacted directly. For example, touching a hot stove, the skin gains heat and may burn.
3. *Evaporation*. This process depends on the difference between the water vapor pressure of the skin and water vapor in the environment (or relative humidity).
4. *Radiation*. This process refers to the difference between the temperature of the skin and surfaces in the environment. For example, standing in sunlight, we receive radiation from the sun.

This section discusses how temperature affects performance and health. It also looks into the comfort and discomfort zones and, finally, discusses the tools used to measure energy demands and how to reduce heat stress. Excellent information on this subject can be found in the literature (ISO, 1994, 1995; Parson, 1995).

6.3.3.1 Effects of heat on performance

Many studies over the last five decades investigated the effects of thermal conditions (heat) on performance of cognitive and physical tasks. No firm conclusions can be drawn. For example, while many studies have shown some performance decrement, others demonstrated no differences or even improvements in performance (Kobrick and Fine, 1983). The reasons may be related to the following interacting factors:

- Most studies occur in a laboratory environment.
- Subjects are mainly young healthy men.
- Different techniques and exposure conditions are used.
- Wide differences among the tasks performed.
- Subjects have different skill levels.
- Individual differences (biologically and physiologically) not controlled.

These reasons create difficulty generalizing the results to the general working population. However, despite the differences, some reasonable generalized conclusions can be used in the work environment. These are summarized next. A good review of task performance in heat is provided in Ramsey (1995).

6.3.3.1.1 Cognitive tasks

The effect of heat on performing cognitive (higher mental function) activities depends on the type of tasks at hand. For example, simple cognitive tasks, such as visual or auditory reaction time, mathematical problem solving, and short-term memory tasks, respond differently to heat exposure than complex cognitive tasks that process control operators typically do, such as monitoring DCS screens for out-of-limit conditions (Ramsey and Kwon, 1988). In simple cognitive tasks, heat does not affect performance negatively. In fact, an improvement in performance was obtained when these simple cognitive tasks were performed during a brief exposure to heat. On the other hand, in complex cognitive tasks, heat negatively affects performance, starting between 30 and 33 °C or 86 and 91.4 °F (Ramsey and Kwon, 1988).

6.3.3.1.2 Physical activities

Physical activities, especially heavy activities, performed in hot, humid conditions may cause fatigue and exhaustion much sooner than in more moderate conditions. For example, sitting at rest, men can tolerate 38 °C (100 °F) for 3 or more hours. However, the tolerance is reduced to 30 minutes (Lind, 1963) if working moderately at 4.67 kilocalories per minute (kcal/min). Kilocalories is the heat required to raise the temperature of 1 kg water from 15 to 16 °C. In another study by Wyndham (1974) on men shoveling rocks, the productivity level was maintained until the temperature exceeded 28 °C (82.4 °F). At about 30 °C (84 °F), productivity dropped to 90% and to 50% when the temperature reached 34 °C (93 °F). It is worth adding here that an increase in temperature can affect people's safe behavior. A large study observing the behavior of over 17,000 people at work found that the incidence of unsafe behavior increased as the climatic

conditions fell outside the range 17–23 °C or 63–73 °F (Ramsey et al., 1983).

6.3.3.2 Effects of cold on performance

As an increase in core body temperature can lead to heat stress (hyperthermia), so a small decrease in core body temperature can produce cold stress (hypothermia). However, the human body is better able to tolerate increased heat than compensate for heat loss. Examples of jobs where people work in a cold environment are oil and gas extraction, electric line repair, gas line repair, sanitation, trucking, warehousing (especially cold storage), and military (Sinks et al., 1987). Similar to the effects of heat on performance, the studies of the effects of cold on performing different functions are not well understood and therefore lead to no firm conclusions. The reasons may be related to the following interacting factors:

- Differences in air flow, temperature, and humidity.
- Differences in the duration of exposure.
- The difference in exposure level to the different parts of the body.
- Individual differences (biologically and physiologically).

In spite of the interacting factors and differences in the results of the studies, some generalized conclusions can be used from the literature. These conclusions are summarized next.

6.3.3.2.1 Cognitive tasks

The effect of cold on cognitive tasks has been detected in studies on reaction time and complex mental activities. However, the effect depends on the severity of the temperature, task performed, and skill and experience of subject with the task and with performing in the cold. The following is a summary of the results:

- Mental performance decreases when performing complex, demanding tasks requiring high levels of concentration and significant use of the short-term memory (Bowen, 1968; Enander, 1987, 1989).
- Cold does not affect the cognitive efficiency of well-motivated subjects (Baddeley et al., 1975).
- Cold acts as a distractor that interferes with some types of mental performance tasks (Enander, 1989).

6.3.3.2.2 Physical activities

Cold significantly affects physical work. The reduction in limb or whole body temperature reduces physical capacity. For example, reduced limb temperature (i.e., arm and hand) affects motor ability, causes a loss of cutaneous (touch) sensitivity, affects the limb's muscular control, and reduces dexterity, muscular strength, and endurance abilities (Lockhart, 1968; Morton and Provins, 1960; Enander, 1989). The reduction in whole body temperature reduces the rate of metabolism within the muscles of the extremities and, in turn, the performance due to shivering. Indeed, Lockhart (1968) suggested that part of the reason performance decreases in tasks involving dexterity activities may be due to shivering.

In practice, gloves and heavy thermal boots and clothing can protect operators in cold weather. In severely cold temperatures, the clothes and gloves can affect the ability of operators to perform tasks that require strength or dexterity by limiting body motion.

6.3.3.3 Effects of heat on health

6.3.3.3.1 Hot environment

Exposure to high levels of heat can affect a person's health in two ways. First, increased temperature on the skin can lead to tissue damage from burning (e.g., over 45 °C or 113 °F). This situation is usually manageable because it is observable and the people react to it by distancing themselves from the heat. The second way heat can affect health is more serious and dangerous. It deals with increased core body temperature to about 42 °C (108 °F), where heat stroke (hyperthermia) occurs. This rise in core body temperature results in an increase in metabolism, which in turn produce heat that needs to be dissipated. If the cycle is allowed to continue, death can ensue. This cycle is referred to as Q10 effect (for every 1 °C increase in deep body temperature, metabolism rises 10%). There are two possible reasons why the heat generated does not leave the body: the body is exposed to heat and humidity that reduces the sweat evaporation; and some protective clothing acts as insulation.

The continuous exposure to hot environmental conditions and the heat stress it generates can lead to several forms of heat illnesses, with heat stroke being the most serious:

- Heat rash, also known as *prickly heat*—where the skin erupts with red pimples with intense itching and tingling caused by inflammation around the sweat gland ducts.
- Heat cramps—spasms of the muscles result from loss of salt.
- Heat exhaustion—results from dehydration, with a general feeling of headache, loss of performance through muscular weakness, dizziness, possible vomiting.
- Heat stroke—a significant and alarming increase in core body temperature and the failure of the body to loose the heat. The general symptoms are headache, dizziness, vomiting, shortness of breath, strange behavior, and eventually loss of consciousness and death.

Heat stroke is caused mainly by strenuous physical work or exercise in hot conditions, duration of exposure to extreme heat sources, and highly motivated and competitive individuals (i.e., sports).

In addition, the exposure and tolerance to heat stress is a function of individual differences (Burse, 1979; Strydom, 1971; Sanders and McCormick, 1993):

- *Gender*. Generally, women are less tolerant to hot environmental conditions and, therefore, more vulnerable to heat illnesses than men.
- *Age*. Older people are more vulnerable to heat illnesses than younger people.
- *Physical fitness*. The more the person is physically fit, the longer is the tolerance to hot environmental conditions, up to a point.
- *Body fat*. Body fat acts as an insulating layer blocking the release of heat. It also acts as an extra weight to move around, increasing the level of physical activity leading to higher body heat and energy expenditure.
- *Alcohol*. Alcohol affects peripheral nervous functioning and produces dehydration.

6.3.3.3.2 Cold environment

Exposure to cold can lead to a drop in core or deep body temperature, likely to produce a risk to health. When the core body temperature falls below 35 °C (95 °F), the state is referred to as *hypothermia* or *cold stress*. With core body temperature below 35 °C (95 °F), the risk of disorientation, hallucination, and unconsciousness increases, leading to cardiac arrhythmia as the core temperature drops even further and subsequent death from cardiac arrest.

As the body is exposed to cold and the core body temperature starts to drop, the body automatically and rapidly uses its regulation system and produces two physiological reactions to cold stress:

- Shivering—Shivering is characterized by increased muscular activities and contractions to produce heat rapidly; the colder it is and the faster the core body temperature drops, the faster and more intense shivering becomes, leading to exhaustion, until core body temperature reaches 30–33 °C (86–91 °F), where shivering gradually is replaced by muscular rigidity.
- Constriction of blood vessels, also known as *vasoconstriction*—Vasoconstriction is characterized by an increase in the constriction of the peripheral (skin and extremities) blood vessels and an eventual shut off of blood flow to these regions as the core body temperature drops further, leading to an increase in blood pressure. This process serves two main purposes: First, the warm blood is diverted from the cold skin, such as hands, fingers, and face, and distributed to the internal organs, so less body heat is lost. Second, the insulating capacity of the skin is increased up to six times by cutting off the flow of blood (Sanders and McCormick, 1993). As the constriction of blood vessels increases and blood flow is shut off, the temperature of all exposed areas (hands, fingers, toes, face, nose) rapidly approaches the air temperature, causing freezing of body tissues (cold injury) or frostbite.

In addition, the exposure and tolerance to cold exposure is a function of individual differences (Timbal, Loncle, and Boutelier, 1976):

- Body composition—amount of subcutaneous (under the skin) fat around the body, which acts as good insulating material. This is especially important when blood vessels constrict in response to the cold and blood is diverted away from the body surface toward the internal organs for survival.
- Individual size and weight—since the degree of heat loss is proportional to the body's surface area, the bigger an individual is in terms of height and weight, the greater is the amount of heat that can be generated by shivering. Note that the heat generated through shivering depends on the mass of active muscular tissues in the body.
- Physical fitness—the better the physical fitness, the more efficient the production of heat through the shivering process and the longer it is maintained with delayed physical exhaustion.

6.3.3.4 Comfort and discomfort zones

Comfort in the comfort zone is perceived when the body's physiological responses to the environmental temperature and humidity are within the normal regulatory responses. However, when those physiological responses exceed the normal regulatory responses, individuals perceive discomfort. People's perception of comfort, in the comfort zone, is influenced by the following environmental factors:

Air temperature—comfort zone, 20–25 °C (68–78 °F).
Relative humidity—comfort zone, 30–70%.
Air velocity—comfort zone, 0.1–0.3 m/s (20–60 ft/s).

Other factors that contribute to the perception of comfort are

- Workload. The heavier the work load, the lower the upper limits of the temperature and humidity must be.
- Radiant heat. Examples are furnaces, sun, ovens, radiant-heat lamps.
- Clothing. The type and amount of clothing worn and its insulation value determine how much the upper

limits of temperature and humidity must be lowered.
- Age. Elderly people prefer higher temperatures than younger people, since the metabolism rate decreases slightly with age.
- Gender. Women have a lower metabolic rate than men and, in general, might prefer a warmer environment.
- Color. Color of the surroundings can influence an individual's thermal comfort. This influence is psychological in nature; for example, when red predominates, individuals feel warmer. The color blue elicits the opposite response.

6.3.3.5 Work tolerance in a hot environment

A key factor to examine and quantify a warm working environment is the energy that the human body consumes to perform the job. Through the metabolism process, the body uses oxygen to convert food and body fat into mechanical energy and heat. In turn, the body consumes energy to stay alive. The total energy consumption (measured in kilocalories, kcal, per unit time) is composed of three components:

1. Basic metabolism—the number of kcal used during sleeping, breathing, and the like to keep the organs alive.
2. Leisure calories—the number of kcal used when performing activities such as reading or playing a musical instrument.
3. Work calories—the number of kcal used when performing different physical movements to accomplish a task, such as typing a manuscript on the computer, lifting objects, walking up or down stairs (Table 6.3.5).

A tool to predict energy expenditure was developed by the University of Michigan (1988). The tasks this tool analyzes are lifting and lowering; holding, walking, and carrying; horizontal arm work forward; lateral arm work; pushing and pulling; general hand and arm work; throwing loads; and climbing steps.

Table 6.3.5 Examples of energy expenditure for some occupations

Occupation	Energy expenditure (kcal/min)
Light assembly work (seated)	2.2
Work in laboratory	2.5
Walking on level (4.5 km/h)	4.0
Walking on level (6 km/h)	5.2
Chopping wood	8.0
Mowing the lawn	8.3
Shoveling (16 lb weight)	10.2

Energy consumption is certainly affected by (1) the ambient temperature and the workload, (2) the age of the individual, and (3) rest duration.

The interaction of temperature and workload is illustrated in Table 6.3.6.

The effect of age on permissible energy consumption is given in Grandjean (1988). If we set the maximum energy consumption of a 20–30 year old as the 100% level, then the relative maximum energy that can be consumed by older workers is given by the following:

Age (years)	Maximum energy consumption
40	96%
50	80%
60	80%
65	75%

The rest duration is determined in terms of the energy consumed to perform a particular task. For example, if we take an individual who works in a warehouse, lifts and stacks bags of chemical powder over the day, and consumes energy at an estimated rate of 7.5 kcal/min, then using the following equation we can determine the rest

Table 6.3.6 Maximum working temperature by workload

Energy consumption, kcal/min	Type of work	Maximum effective temperature, °C (°F)
6.7	Heavy manual work (walking with a 66-lb load (30 kg)	26–28 °C (78.8–82.4 °F)
4.2	Moderate heavy work at 2.5 miles/h (4 km/h)	29–31 °C (84.2–87.8 °F)
1.7	Light sedentary work	33–35 °C (91.4–95 °F)

time needed to be able to continue the job without fatigue:

$$\text{Rest time (as \% of work time)} = (\text{kcal/min} - 1) \times 100\% = 87.5\%$$

The results suggest that, for 10 min (600 s) of steady work, he or she may need about 525 s (about 9 min) rest.

6.3.3.6 Recommendations to improve working conditions

The following recommendations for improving both working conditions in hot and cold environments are summarized from several sources in the literature (Grandjean, 1988; Sanders and McCormick, 1993; Eastman Kodak, 1983; Konz, 1990).

6.3.3.6.1 Guidelines for heat conditions

The recommended guidelines to achieve a level of comfort working in hot conditions follow a systematic approach by making changes to the ambient conditions, the task, the personal protective equipment, and finally the individual:

- A worker, new to a hot environment, should be acclimatized to heat by initially spending only 50% of the working time in the heat and increasing this by 10% each day.
- Train workers to recognize signs of heat illness, and provide and locate drinking water close to workers and the task performed.
- The greater is the heat load and the greater the physical effort performed under heat stress, the longer and more frequent should be the rest or cooling periods.
- If the limit of heat tolerance is being exceeded, the working day must be shortened.
- Where the radiant heat is excessive (i.e., blast furnaces), the worker must be protected by special goggles, screens, and special protective clothing against the risk of burning the eyes and hands. Special protective clothing includes such equipment as ice vests and vortex cooling suits.
- Where radiant heat is excessive (sun heat in the middle of July in traditionally warm states and countries), provide cooling facilities (cool rooms) for individuals to use for their work breaks or when needed.
- Match the workload to ambient temperature and relative humidity, where possible.
- Increase the flow of air in the workplace, where possible, using, for example, fixed or portable fans.

6.3.3.6.2 Guidelines for cold conditions

The recommended guidelines to achieve a level of comfort working in cold conditions are

- Design to ensure that the work is performed in a protected environment (e.g., provide windproofing around a task normally performed outside).
- Block the source of air flow with a shield.
- Change the work area location, if possible, away from the air flow source.
- Increase the ambient temperature where possible. If not, provide adjustable auxiliary heaters to accommodate individual differences. For continuously moving and changing locations, this may not be possible.
- Recommend and provide proper warm clothing to extend the tolerance to cold conditions.
- In extremely cold environments, if the limit of cold tolerance is being exceeded, the working periods must be shortened.
- Provide warming facilities (warm rooms) for individuals to use for their work breaks or when needed.
- Use proper gloves to protect hands and fingers. When finger dexterity is required, use partial gloves that cover the hands but expose the fingers.
- Train people to recognize signs of cold stress.

6.3.4 Noise

Noise is an aspect of our work and daily living environment produced by equipment, machines, and tools. Noise is commonly referred to as *unwanted sound*. Burrows (1960) clearly defines *noise* as "an auditory stimuli bearing no informational relationship to the presence or completion of the immediate task." However, it depends on the individual's subjective reaction whether to consider a sound source as noise or not. In addition, sound labeled *noise* on one occasion may not be noise in a different occasion or environment.

Sound is defined in terms of two parameters, frequency and intensity, because a sound is a vibration stimulus experienced through the air. Frequency is an acoustic energy reflecting the number of vibrations per second, expressed in hertz (Hz). One hertz is equal to 1 cycle per second. The range of frequency between 2 and 20,000 Hz defines the frequency limits of the ear. People subjectively perceive frequency as pitch. Intensity is related to the sound pressure variation caused by the sound source. People subjectively perceive intensity as loudness. Intensity is specified in terms of sound pressure level (SPL) and expressed in the logarithmic units of decibels (dBA). Being logarithmic, a

Table 6.3.7 Examples of sources of noise

Sound	dBA level	Sound power increase
Rock concert 30 m from speakers	120	1,000,000,000,000 (10^{12})
Pneumatic hammer	100	10,000,000,000 (10^{10})
Average traffic	80	100,000,000 (10^{8})
Vacuum cleaner (10 ft)	70	10,000,000 (10^{7})
Conversational speech	60	1,000,000 (10^{6})
Quiet office	40	10,000 (10^{4})
Bedroom at night	30	1000 (10^{3})
Normal breathing	10	10 (10^{1})

10-fold increase in power (intensity) occurs with each 10 dBA increase. For example, 60 dBA on the decibel scale is not twice as intense as 30 dBA but 1000 times as intense as 30 dBA. Table 6.3.7 presents some examples of sources of noise. This section discusses how noise affects performance and health. It also provides guidelines on how to reduce and control noise in living and working environments.

6.3.4.1 Effects of noise on performance

The effects of noise on performance have been studied for many years with mixed results. While some studies showed performance decrements, others showed no effects or even an improvement in performance (Kryter, 1970; Gawron, 1982). Similar to the studies on the effects of temperature on performance, the mixed results of noise effects are related to the following interacting factors:

- Most studies are in a laboratory environment.
- Different techniques used to produce noise (e.g., office noise to rocket noise) and different exposure conditions (i.e., definition of quiet ranged from 0 to 90 dBA).
- Wide differences among tasks, from reaction time and number checking all the way through to accuracy of shooting; because of the different nature of difficulty, these tasks require relatively different demands on memory, cognitive, and motor capabilities of those tested.
- Individual differences (biologically and physiologically) in the way humans respond to noise or a noise and task combination.
- Different techniques of experimental design used to test the effects of noise.

Despite the mixed results, reasonable conclusions can be made about the effects of noise on performance. These are summarized next. It is safe to say at this point that noise has little or no effect on manual work, and this is not discussed further.

6.3.4.1.1 Speech and communication

Speech and communication depend highly on the listener's ability to receive and decode the sounds. This means that the listener must hear the signal (speech or auditory display) and understand the message. Therefore, the speech and communication comprehension depends on the loudness of the voice as well as the level of background noise.

Hearing and understanding the signal may be masked (blocked) by a noisy environment. If the signal intensity remains constant and the masking intensity increases, then more masking occurs. The reverse is also true. Similarly, as the frequency of the masker approaches that of the signal, more masking occurs. From a practical side, it is important to know the intensity and frequency of both the signal and masker. For example, if the intensity and frequency of an auditory display are much higher than those of the noise, then masking does not occur. Otherwise, performance reductions (not hearing and understanding) due to masking occur and corrective measures must be considered (this is discussed in detail in the section on how to control noise).

6.3.4.1.2 Cognitive performance

The relationship between noise and cognitive performance is similar to that between heat or cold and cognitive performance. The results are again mixed. Even so, some reasonable conclusions can be made about the effects of noise on cognitive performance for specific types of tasks:

- Generally speaking, noise clearly has an effect on overall performance, especially when it reaches a level over 90 dBA.
- Noise levels over 100 dBA can affect monitoring performance over relatively long duration (Jerison, 1959).
- Tasks with large short-term memory component (i.e., mathematical calculations) are affected by high noise levels (Poulton, 1976, 1977). The reasoning here is that we cannot hear ourselves think in noise.
- Noise produces more errors in complex rather than simple tasks (Boggs and Simon, 1968).
- Tasks performed continuously without rest pauses between responses are affected by noise (Davies and Jones, 1982).

- Tasks requiring high perceptual or information processing capacity (i.e., high mental demands, complexity, and considerable detail) are affected by noise (Eschenbrenner, 1971).
- Noise does not reduce performance on simple routine tasks or motor performance (i.e., manual work). Positive noise, in the eyes of the beholder, such as music, may facilitate an improvement in performance and satisfaction. The reasoning here is that music can aid in reducing boredom and fatigue associated with these tasks (Fox, 1971).

6.3.4.1.3 Nuisance and distraction

Noise is considered a nuisance (annoying or unpleasant) when it interferes with our ability to carry out an activity. Thus, the response is subjective. This subjective nature indicates that what is annoying to one individual may not be to another. Furthermore, what is a nuisance on one occasion may not be at another time. For example, people are not disturbed by the noise generated by their own activities. A few observations can be made regarding this point:

- Loud, high-frequency noises are considered a nuisance or annoying. They contribute to performance decrements by creating a distraction and affecting concentration and effective communication.
- More familiar noises are perceived as less annoying (Grandjean, 1988).
- Sound features that contribute to the different degrees of "noise perceived annoyance" are intensity, frequency, duration, and the meaning to the listener. For example, the annoyance level increases as the intensity and frequency of sound increases, the longer it remains, the less meaningful it is to the listener (Kryter and Pearsons, 1966).
- Overhearing conversations is also considered, by most people, annoyance (Waller, 1969).
- Intermittent noise can reduce performance on mental tasks more than continuous noise.

6.3.4.2 Effects of noise on health

The most important and obvious effect of noise on health is hearing loss and deafness. Hearing loss and deafness can have two causes: conduction deafness and nerve deafness. Conduction deafness is caused by the inability of the sound waves to reach the inner ear and vibrate the eardrum adequately. Conditions that may lead to conduction deafness are wax in the ear canal and infection or scars in the middle ear or eardrum. Conduction deafness is not the product of a noisy environment and does not result in complete hearing loss. Nerve deafness is caused by reduced sensitivity in the inner ear due to damage or degeneration of nerve cells. The hearing loss here is in the higher frequencies, whereas conduction deafness is more even across frequencies.

Establishing a strong relationship between hearing loss and exposure to a specific noise is not easy. This is simply because during the normal aging process we are exposed to nonoccupational noises (vehicles, TV, vacuum cleaners, etc.) as well as occupational noises. The combined effects and the cumulative nature of the process make it difficult to point to a specific cause. The degree to which hearing is affected depends on the type, intensity, frequency of noise, exposure duration, and individual sensitivity and susceptibility. Two types of hearing loss are discussed here: aging, which is a normal process, and noise-induced temporary and permanent hearing loss.

6.3.4.2.1 Aging hearing loss

As a result of the normal aging process, hearing loss can occur due to changes in the structure of the auditory system. The change is more noticeable in men than in women. This hearing loss is also more noticeable at the higher frequencies. The decline in hearing loss is not consistent in all individuals simply because of individual biological and physiological differences and susceptibility. Taking a frequency of 3000 Hz as standard, Grandjean (1988) reports the loss of hearing to be expected at various ages as

Age (years)	Expected Hearing Loss (dBA)
50	10
60	25
70	35

6.3.4.2.2 Noise-induced hearing loss

Noise-induced hearing loss or deafness can be related to continuous or noncontinuous exposure to noise. This noise-induced deafness can be temporary (temporary threshold shift, TTS) or permanent (permanent threshold shift, PTS). Even though the initial TTS is not damaging, continuous experiences of TTS lead to PTS (Kryter, 1970). This is because the TTS and PTS affect the same areas of the ear.

Temporary hearing loss is characterized by the ability to recover normal hearing in a few hours or days after an exposure to continuous high-intensity noise (hence, the term TTS). For example, being exposed to 90 dBA noise can cause about a 10-dBA shift in the hearing threshold. However, exposure to 100 dBA can cause about a 60-dBA shift. Permanent hearing loss or deafness is characterized by an irreversible deafness after repeated

Table 6.3.8 Permissible noise exposure according to OSHA

Permissible time (h)	Sound level (dBA)
16	85
8	90
6	92
4	95
2	100
1	105
0.5	110

exposure to high-intensity noises (hence, the term PTS). This PTS gradually appears as the number of years of noise exposure increases, and it occurs first at 4000 Hz, which is within the region of frequencies to which the ear is most sensitive.

The United States National Institute of Occupational Safety and Health (NIOSH, 1972) indicates that a continuous exposure, about 8 h, to noise levels above 85 dBA can cause hearing loss. Therefore, personal protective equipment (i.e., hearing protection) and engineering controls must be implemented in the workplace. This is discussed later. The U.S. Occupational Safety and Health Administration (OSHA, 1983) established permissible levels of noise exposure depending on the level of noise and the duration of the exposure. This is shown in Table 6.3.8.

6.3.4.3 Guidelines to control noise

Despite the recommendations by NIOSH and OSHA, in some work areas, workers must perform their jobs in high-noise situations for long hours. This section presents an overview of noise control methods. Noise control methods can be applied at the noise source, in the path of noise transmission and the surrounding area, and to the individual receiving the noise. In many cases, a combination of methods must be applied. Haslegrave (1995) provides a good review of noise assessment.

Before embarking on solving a noise issue, we need to determine the level, location, and extent of the noise issue confronting the human and the task to be accomplished. The following are some principles to follow:

- Use proper, well-calibrated equipment.
- When doing measurements, keep away from reflecting surfaces (i.e., machines, walls) and hold the sound pressure level meter away from your body, at about arm's length, in the hearing zone.
- Sketch the area where noise level measurements are taken. This may not be necessary for simple noise sources.
- Map the noise level measurements on sketched area.
- Depending on the human activities performed at and around the noise source, determine what noise level is acceptable according to local regulations.

4.4.3.1 Noise control at the source

Controlling noise at its source is the ultimate goal. However, this method can prove expensive. The following are some examples on reducing noise at the source:

- Enclose the source area with sound-absorbing material. This can reduce radiated noise by 20–30 dBA. High-frequency noise can be better contained and absorbed by barriers than low-frequency noise; for example, barriers around the flare of an offshore platform, baffles for air intakes such as in furnaces, and mufflers such as on steam vents with high-volume air.
- Routinely maintain, lubricate, and align equipment to reduce vibration and therefore noise levels.
- Choose and use tools and equipment that generate low-frequency noise. This may not always be possible because such tools may not exist commercially and a special order may be costly.

6.3.4.3.2 Noise control in the path of noise transmission

This method can be used either in addition to or after all the possibilities of controlling noise at the source have been exhausted. Some of the applied principles here are

- Place a barrier, such as a shield or wall, between the human and noise source to deflect and absorb noise. Please note that the reduction of low-frequency noise (below 1000 Hz) is not possible, since it travels over and around barriers.
- Cover walls, ceilings, and floors with sound-absorbing panels to reduce noise reflection back into the work area. It is worth noting that the overall noise reduction, using this technique, is rather limited; however, it does further reduce noise.
- Take advantage of distance from the source, since the level of noise reduces with the square of the distance traveled.

6.3.4.3.3 Noise control at the receiver

If the two preceding methods fail to reduce noise to a safe level, then this final method involves the use

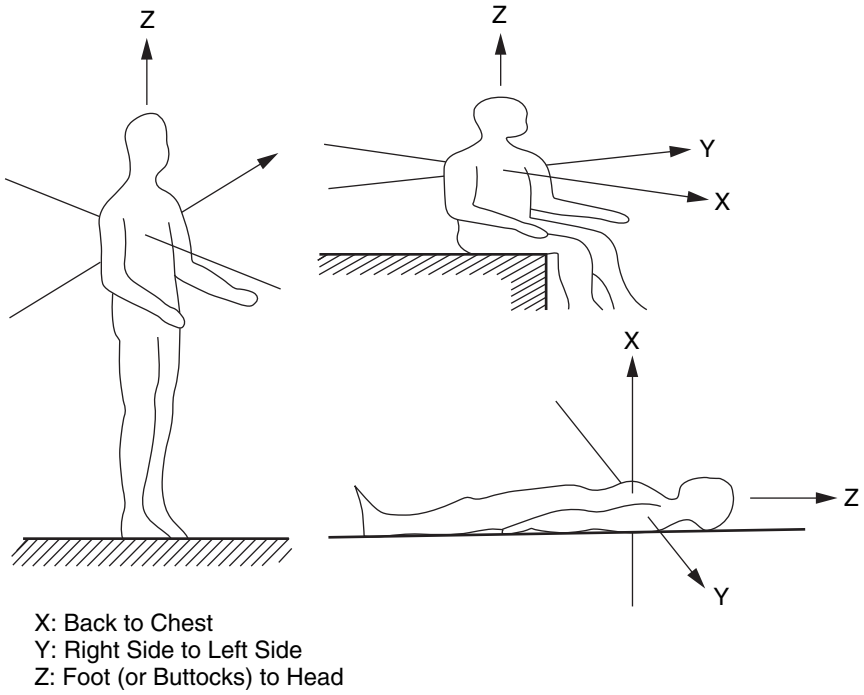


Fig. 6.3.2 Direction of vibration relative to torso.

of personal hearing protection. The applied principles here are

- Make available and properly train individuals on the use of hearing protection devices. These can include ear plugs, which can reduce noise level by up to 30 dBA and are suggested use for noise levels of 85 dBA and above; ear muffs, which can reduce noise level by about 40–50 dBA and are suggested use for noise levels over 100 dBA; a combination of the use of ear plugs and muffs; and soundproof helmets, which cover the head and ears.
- Provide instructions and train people on how to use hearing protection.
- Rotate jobs to reduce exposure time.

6.3.5 Vibration

Vibration is any regular or irregular movement of a body about its fixed position. Regular vibration (such as experienced in machinery like pumps, compressors, or power saws) is known to have a predictable waveform that repeats itself at regular intervals. Random vibration is irregular, unpredictable, and the most common type encountered in the real world; for example, vibration experienced in trucks, tractors, trains, and airplanes.

Vibration is characterized by its direction and amount (frequency and acceleration). The body can vibrate and move in one or more directions. The movement direction is described in terms of three orthogonal components: X, Y, and Z, where

X direction is back to chest.
Y direction is right side to left side.
Z direction is foot or buttocks to head.

For example, looking at Figure 6.3.2, if the person is standing or sitting on an up and down vibrating platform, the vibration is in the Z direction. If the person is lying on his or her back, the vibration is in the X direction.

Frequency of movement expresses the speed of vibration. Frequency is measured in Hz, where 1 Hz = 1 cycle per second. For example, 1 cycle is the body movement from its fixed reference point to its highest point, to its lowest point, back to its reference point. Therefore, the higher the speed and the faster the body vibrates, the more cycles take place per second.

Intensity of vibration is expressed in terms of the maximum amount (amplitude or displacement) the body moves from its fixed reference point, measured in inches or meters. Intensity can also be specified in terms of acceleration. Acceleration can be expressed in terms of inch per second squared (in./s^2) or meter per second squared (m/s^2). Acceleration can also be expressed in terms of gravity (g) where 1 g is the amount of acceleration needed to lift the body off the earth's surface. The metric system defines 1 g as equal to 9.81 m/s^2. This section discusses how vibration affects performance and health. It also provides guidelines on how to reduce and

control vibration. Griffin (1990) provides a full review of the subject.

6.3.5.1 Effects of vibration on performance

The principal effects of vibration on performance are reduced motor (i.e., hand steadiness) and visual (i.e., eye fixation) control. No strong evidence exists to suggest vibration affects information processing, reaction time, pattern recognition, or any intellectual (cognitive) tasks.

6.3.5.1.1 Motor control

Considerable research has consistently demonstrated the effects of vibration on motor performance (Hornick, 1973; Levison, 1976; Lewis and Griffin, 1978, 1979a). A summary of the findings follows:

- Vibration has residual effects that may last up to half an hour after exposure.
- At low vibration frequencies (less than about 15 Hz), performance reduction is related to the amount of vibration experienced by the controlling limb.
- As the vibration intensity transmitted to the related limb increases, tracking performance decreases.
- Manual performance degradation depends primarily on the intensity of the vibration.
- Manual performance degradation occurs where the task demands movement in the axis in which the individual is being vibrated. For example, performing vertical tracking is affected more by the vibration in the *Z* axis than the *Y* or *X* axis.

6.3.5.1.2 Visual performance

Examples of the effects of vibration on visual performance are observed in work situations such as driving trucks or tractors or riding trains. The most commonly reported problems associated with these situations are blurry images in the visual field and reduced visual acuity (ability to see fine details). This is probably due to the instability of the image on the retina. A summary review of the findings (Griffin, 1976; Meddick and Griffin, 1976; Barnes, Benson, and Prior, 1978; Griffin and Lewis, 1978; Lewis and Griffin, 1979b; Moseley and Griffin, 1986) follows:

- When object and person are vibrating, both
  - May be vibrating at different frequencies and intensities.
  - May or may not be moving in phase.
  - May or may not be moving in the same axis.
- The effects of vibration on visual performance depend on the size of object observed and its distance from the observer.
- Visual performance degradation is determined mainly by the vibration frequency. For example, visual performance impairment occurs in the frequency range 10–25 Hz.
- Low frequency vibration affects the ability of the eye to track targets and maintain a stable image on the retina (e.g., reading a book on a train ride).
- High-frequency vibration creates blurry images.

6.3.5.2 Effects of vibration on health

Exposure to vibration, especially long term, leads basically to two health issues: damage to body organs due to whole-body vibration, which is a function of high-intensity vibration; and damage to blood vessels, body tissues, and nervous system to the fingers, hands, and arms due to segmental vibration, which is a function of high-frequency vibration. The majority of studies in the literature focus on tractor drivers, truck drivers, and workers using power tools (Rosseger and Rosseger, 1960; Ramsey, 1975; Abrams and Suggs, 1977; Whitham and Griffin, 1978; Seidel and Heide, 1986; Grandjean, 1988; Sanders and McCormick, 1993).

Whole-body vibration is mainly the product of vertical vibration (*Z* axis), when standing or sitting (Figure 6.3.2), absorbed through the feet when standing and the buttocks and feet when sitting. The most common situations where individuals are exposed to long-term, high-intensity, whole-body vibration are driving trucks, buses, tractors, and operating construction and heavy equipment. The most common physiological effects and complaints of whole body vibration are

- Stomach or abdominal pain and spinal disorders at the low back and neck regions. In addition, traces of blood in the urine have been found, probably due to kidney damage.
- Chest and abdominal pain from operating within the frequency range 4–10 Hz.
- Back and neck aches within the range 8–12 Hz.
- Headaches, muscular tension, and irritation in the intestines and bladder within the range 10–20 Hz.

Segmental vibration is mainly the product of long exposure to high-frequency vibrating objects absorbed through the fingers, hands, and arms. The most common situations where individuals are exposed to long-term, high-frequency segmental vibration come from examples of using handheld power tools such as grinders, chain saws, pneumatic road drills and hammers, or polishing machines. Further, the design, size, and weight of the power tool also affects its vibration level (Abrams and Suggs, 1977). The most common

physiological affects and complaints of segmental vibration are

- Structural damage to blood vessels and nerves, mainly in the fingers and hands, with vibration frequencies between 40 and 500 Hz. The arms could also be affected over time.
- Damage to bones, joints, tendons, and muscles of hands and arms with vibration frequencies less than 40 Hz.
- Symptoms associated with these situations are referred to as *white fingers condition*, *Raynaud's disease*, or *vibration-induced white fingers* and consist of
  - Pain, stiffness, and numbness of the fingers and hands.
  - Blanching (white or bluish color) of either all or part of the hands and fingers.
  - Loss of muscular control and strength.

6.3.5.3 Guidelines to reduce or control vibration

There are basically two ways to reduce or control the effects of vibration on the individual: engineering and administrative controls (Eastman Kodak, 1983; Grandjean, 1988, Sanders and McCormick, 1993). Engineering controls can basically be achieved through two approaches:

- Reducing vibration at the source by
  - Proper equipment maintenance.
  - Mounting equipment on pads or springs.
  - Modifying equipment to reduce speed and acceleration.
- Preventing vibration from reaching the individual by
  - Placing special spring suspensions under seats and platforms to absorb vibration.
  - Using rubber mats for standing tasks.
  - Designing hand-operated power tools with damping-isolating materials, especially between the hand and hand-grip of the tool.
  - Using special vibration-absorbing gloves and, reducing the time spent working in cold conditions.
  - Designing the size and weight of the hand-operated power tools (i.e., using counterweight tools) to minimize the force needed to grip, balance, and operate them.

Administrative controls focuses on three approaches to reduce exposure to vibration:

1. Train individuals on the issues associated with vibration in terms of performance and health effects.
2. Provide the proper personal protection equipment, such as special gloves, and invest in properly designed equipment.
3. Rotate individuals among work tasks to control and reduce exposure. For example, Bovenzi, Petroniv, and DiMarino (1980) recommend the use of a chipping hammer (used to remove caked material from metal surfaces) should not exceed 4 h per person per shift.

6.3.6 Case study

The proper quality and quantity of lighting in a control room environment is critical to the efficiency and safety of the petrochemical plant. This case study deals with the evaluation of an existing lighting system in a process control center, and recommendations for follow-up.

6.3.6.1 Method

The evaluation methods include

- Recording the illumination levels at each writing surface of the control room (with the brightness adjusted to a "personnel preference" level by each control room operator). The levels are typically selected to minimize glare reflected from the screens and the stainless steel panels.
- Discussion with individual operators about the lighting in their areas.
- Photos of qualitative HF concerns.

6.3.6.2 Results

We determine the following:

- Illumination levels are too low for reading tasks.
- Stainless steel panels are highly reflective. The ratio of luminance from stainless steel panels to the luminance of the screens in most cases is greater than the maximum recommended luminance ratio, 10:1.
- Small cell parabolics (Figure 6.3.3) are not located or used correctly:
  - Control equipment is located beneath the luminaires that provide illumination for paper-and-pencil tasks. The operators decrease the brightness of these lights to reduce glare.
  - The luminaires located around the room periphery cause less screen glare, but they are not lit, since they would cast shadows on the writing surfaces.
- The large circular light (Figure 6.3.4) structure in the center of the room is a glare source for each console. It is composed of "egg crate" filters over recessed fluorescent lights. The cut-off angle for the egg crates is poor, resulting in spillover of illumination.

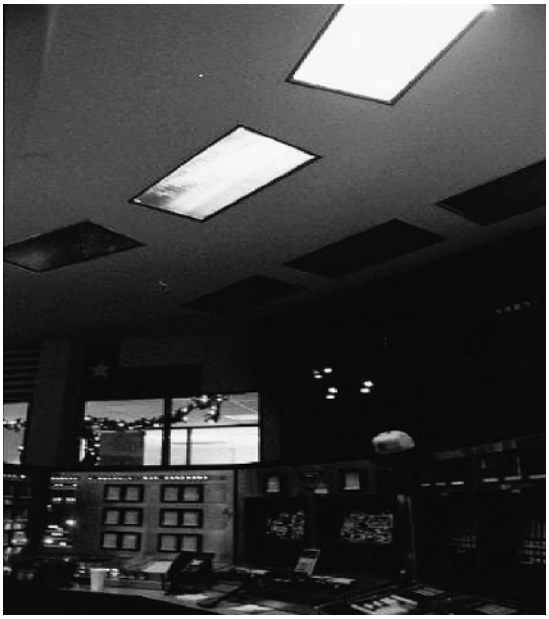


Fig. 6.3.3 Parabolic light.

6.3.6.3 Recommendations

The recommendations for the new lighting system are

- *Task lights*. Ceiling mounted spotlights should be installed to light the writing surfaces in each of the console areas.
- *Ambient illumination*. Drop-down indirect lighting units are recommended to eliminate glare from overhead fixtures.

Recommendations for existing lighting systems are

- Replace the large circular illuminated area in the center of the control room by small-cell parabolics (at the current location of the egg crates) to eliminate screen glare, ceiling-mounted track lights, or inexpensive portable task lights located on the work surfaces that surround the center pole in the room.
- Diffuse the glare on the control system stainless steel panels by sandblasting the existing panels; painting them with a flat, matte-finish paint (an electrostatic paint process may work here); or install a matte finish covering over the panels.

6.3.6.4 Installation of a pilot lighting system

The goal here is to design and install a retrofit lighting system for evaluation. The system consists of two components.

The first component is an indirect lighting luminaire that hangs about 24 in. from the ceiling. The luminaire is inverted and throws its light up, onto the ceiling, from which it is reflected onto the surfaces below. The intensity of the light is continuously adjustable from off to full intensity.

Owing to the even dispersion of the light on the ceiling, no "hot spots" reflect on the panels below, and the light from the luminaire is evenly dispersed across the control space.

The second component of the lighting system is a "spot" task (track) light recessed into the ceiling. The light is equipped with a lens that shines a focused, 3-ft

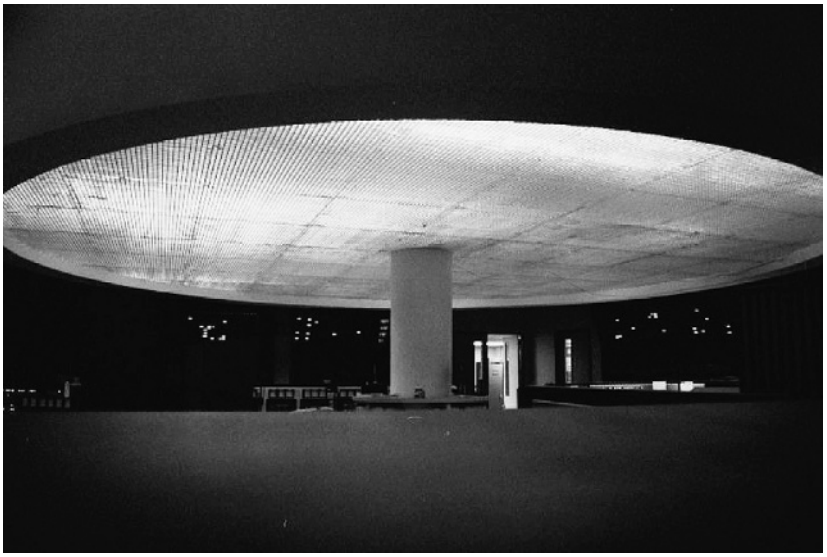


Fig. 6.3.4 Large circular light.

diameter spot of light on the worktable located in the center of the control space. The brightness of the light is continuously adjustable from off to full intensity.

The illumination levels provided by the new lighting system are obtained at several locations around the control space. Each reading is recorded with the indirect light source at full intensity. Readings are obtained with the spotlight off or the light adjusted to full intensity.

6.3.6.5 Final results

The illumination levels provided by the combination of an indirect, ceiling-hung luminaire and track lights are acceptable for the operators' tasks. Levels at each console location are very similar, 2–3 FC (20–30 lux), whether the spotlight is on or off. With the spotlight adjusted to full intensity, the levels for reading at the operator's worktable are higher than required.

The overhead indirect light causes little glare on the console and the task (track) light is so directional that it provides virtually no spillover from a spot of light about 3 ft in diameter on the reading table.

Operators report being very pleased with the result and management has decided to install the indirect lighting system throughout the space. The number and size of the indirect luminaires will be determined from lighting surveys by the contractor. The number and location of the track lights will be determined by where the operators want reading levels of illumination.

6.3.7 Review questions

1. Light stimulus is characterized by its intensity and wavelength. Define light intensity and wavelength.
2. What lighting factors affect performance in the workplace?
3. How is age affected by the light quantity and quality?
4. What types of glare do people experience?
5. What is the luminance ratio and why is it important?
6. How does heat affect physical performance?
7. How does cold affect physical performance?
8. What are the forms of heat illness?
9. What does shivering do to the human body?
10. Define the human comfort zone.
11. How do you calculate rest duration when working in the heat?
12. Define noise and its parameters.
13. How does noise affect people's speech and communication?
14. How can noise lead to a hearing disorder?
15. How can noise be controlled?
16. What is vibration?
17. What are the signs of Raynaud's disease (also known as *white fingers disease)?*
18. What are the effects of vibration on visual performance and motor control?
19. How can vibration be controlled?

References

Abrams, C. F., and Suggs, C. W. (1977) "Development of a Simulator for Use in the Measurement of Chain Saw Vibration." *Applied Ergonomics* 8, no. 2, pp. 130–134.

Baddeley, A. D., Cuccaro, W. J., Egstrom, G. H., Weltman, G., and Willis, M. A. (1975) "Cognitive Efficiency of Divers Working in Cold Water." *Human Factors* 17, pp. 446–454.

Barnaby, J. F. (1980) "Lighting for Productivity Gains." *Lighting Design and Application* 10, no. 2, pp. 20–28.

Barnes, G. R., Benson, A. J., and Prior, A. R. J. (1978) "Visual Vestibular Interaction in the Control of Eye-Movement." *Aviation, Space and Environmental Medicine* 49, pp. 557–564.

Bennett, C., Chitlangia, A., and Pangrekar, A. (1977) "Illumination Levels and Performance of Practical Visual Tasks." Proceedings of the Human Factors Society 21st annual meeting, Santa Monica, CA, pp. 322–325.

Blackwell, H. R., and Blackwell, O. M. (1968) "The Effects of Illumination Quantity upon the Performance of Different Visual Tasks. *Illumination Engineering* 63, pp. 143–152.

Boggs, D. H., and Simon, J. R. (1968) "Differential Effect of Noise on Tasks of Varying Complexity." *Journal of Applied Psychology* 52, pp. 148–153.

Bovenzi, M. L., Petroniv, L., and DiMarino, F. (1980) "Epidemiological Survey of Shipyard Workers Exposed to Hand–Arm Vibration." *International Archives of Occupational and Environmental Health* 46, pp. 251–266.

Bowen, H. M. (1968) "Driver Performance and the Effects of Cold." *Human Factors* 10, pp. 445–464.

Burrows, A. A. (1960) "Acoustic Noise, and Informational Definition." *Human Factors* 2, no. 3, pp. 163–168.

Burse, R. (1979) "Sex Differences in Human Thermoregulatory Response to Heat and Cold Stress." *Human Factors* 21, no. 6, pp. 687–699.

Collins, B., and Worthey, J. (1985) "Lighting for Meat and Poultry Inspection." *Journal of the Illuminating Engineering Society* 15, pp. 21–28.

Davies, D., and Jones, D. (1982) "Hearing and Noise." In W. Singleton (ed.), *The Body at Work*. New York: Cambridge University Press.

Eastman Kodak Co. (1983) *Ergonomic Design for People at Work*, vol. 1. New York: Van Nostrand Reinhold Company.

Enander, A. (1987) "Effects of Moderate Cold on Performance of Psychomotor and Cognitive Tasks." *Ergonomics* 30, no. 10, pp. 1431–1445.

Enander, A. (1989) "Effects of Thermal Stress on Human Performance." *Scandinavian Journal of Work and*

Environmental Health 15, supplement 1, pp. 27–33.

Eschenbrenner, A. J. (1971) "Effects of Intermittent Noise on the Performance of a Complex Psychomotor Task." *Human Factors* 13, no. 1, pp. 59–63.

Fox, J. G. (1971) "Background Music and Industrial Productivity—A Review." *Applied Ergonomics* 2, pp. 70–73.

Gawron, V. (1982) "Performance Effects of Noise Intensity, Psychological Set, and Task Type and Complexity." *Human Factors* 24, pp. 225–243.

Grandjean, E. (1988) *Fitting the Task to the Man.* London: Taylor and Francis.

Griffin, M. J. (1976) "Eye Motion during Whole-Body Vibration." *Human Factors* 18, pp. 601–606.

Griffin, M. J. (1990) *Handbook of Human Vibration.* London: Academic Press.

Griffin, M. J., and Lewis, C. H. (1978) "A Review of the Effects of Vibration on Visual Acuity and Continuous Manual Control: Part I. Visual Acuity." *Journal of Sound and Vibration* 56, pp. 383–413.

Haslegrave, C. M. (1995) "Auditory Environment and Noise Assessment." In J. R. Wilson and E. N. Corlett, *Evaluation of Human Work.* London: Taylor and Francis.

Hornick, R. (1973) *Vibration in Bioastronautics Data Book.* Second Ed. NASA Sp-3006. Washington, DC: National Aeronautics and Space Administration.

Hughes, P., and Neer, R. (1981) "Lighting for the Elderly: A Psychobiological Approach to Lighting." *Human Factors* 23, pp. 65–85.

ISO. (1994) *Moderate Thermal Environment—Determination of the PMV and PPD Indices and Specification of the Conditions for Thermal Comfort.* ISO 7730. International Standards Organization.

ISO. (1995) *Hot Environment—Estimate of the Heat Stress on Working Man, Based on the WBGT-Index (Wet Bulb Globe Temperature).* ISO 7243. International Standards Organization.

Jaschinski, W. (1982) "Conditions of Emergency Lighting." *Ergonomics* 25, pp. 363–372.

Jerison, H.J. (1959) "Effects of Noise on Human Performance." *Journal of Applied Psychology* 43, pp. 96–101.

Knave, B. (1984) "Ergonomics and Lighting." *Applied Ergonomics* 15, no. 1, pp. 15–20.

Kobrick, J. L., and Fine, B. J. (1983) "Climate and Human Performance." In D. J. Oborne and M. M. Gruneberg (ed.), *The Physical Environment at Work.* Chichester, UK: John Wiley.

Konz, S. (1990) *Work Design: Industrial Ergonomics.* Third Ed. Worthington, OH: Publishing Horizon.

Kryter, K. D. (1970) *The Effects of Noise on Man.* New York: Academic Press, pp. 545–585.

Kryter, K. D., and Pearson, K. S. (1966) "Some effects of spectrum content duration on perceived noise level." *Journal of the Acoustical Society of America* 39, pp. 451–464.

Levison, W. H. (1976) *Biomechanical Response and Manual Tracking Performance in Sinusoidal, Sum of Sines, and Random Variation Environments.* Army Medical Research Laboratory Report AMRL-TR-75–94.

Lewis, C. H., and Griffin, M. J. (1978) "A Review of the Effects of Vibration on Visual Acuity and Normal Control. Part II. Continuous Manual Control." *Journal of Sound and Vibration* 56, pp. 415–457.

Lewis, C. H., and Griffin, M. J. (1979a) "Mechanics of the Effects of Vibration Frequency, Level, and Duration on Continuous Manual Control Performance." *Ergonomics* 22, pp. 855–889.

Lewis, C. H., and Griffin, M. J. (1979b) "The Effect of Character Size on the Legibility of a Numeric Display during Vertical Whole-Body Vibration." *Journal of Sound and Vibration* 67, pp. 562–565.

Lind, A. R. (1963) "Tolerable Limits for Prolonged and Intermittent Exposure to Heat." In J. D. Hardy (ed.), *Temperature—Its Measurement and Control in Science and Industry,* vol. 3. New York: Reinhold.

Lockhart, J. M. (1968) "Extreme Body Cooling and Psychomotor Performance." *Ergonomics* 11, pp. 249–260.

Meddick, R. D. L., and Griffin, M. J. (1976). "The Effect of Two-Axis Vibration on the Legibility of Reading Material." *Ergonomics* 19, pp. 21–33.

Morton, R., and Provins, K. A. (1960) "Finger Numbness after Acute Local Exposure to Cold." *Journal of Applied Psychology* 15, pp. 149–154.

Moseley, M., and Griffin, M. J. (1986) "Effects of Display Vibration and Whole-Body Vibration on Visual Performance." *Ergonomics* 29, pp. 977–983.

NIOSH. (1972) *Criteria for a Recommended Standard—Occupational Exposure to Noise.* Cincinnati, OH: U.S. Department of Health, Education, and Welfare. National Institute for Occupational Safety and Health.

OSHA. (1983) "Occupational Noise Exposure: Hearing Conservation Amendment." *Federal Register* 48, pp. 9738–9783.

Parson, K. C. (1995) "Ergonomics Assessment of Thermal Environments." In J. R. Wilson and E. N. Corlett, *Evaluation of Human Work.* London: Taylor and Francis.

Poulton, E. C. (1976) "Continuous Noise Interferes with Work by Masking Auditory Feedback and Inner Speech." *Applied Ergonomics* 7, pp. 79–84.

Poulton, E. C. (1977) "Continuous Intense Noise Masks Auditory Feedback and Inner Speech." *Psychological Bulletin* 84, pp. 977–1001.

Ramsey, J., Burford, C., Beshir, M., and Jensen, R. (1983) "Effects of Workplace Thermal Conditions on Safe Work Behavior." *Journal of Safety Research* 14, pp. 105–114.

Ramsey, J., and Kwon, Y. (1988) "Simplified Decision Rules for Predicting Performance Loss in the Heat." In *Proceedings on Heat Stress Indices.* Luxembourg: Commission of the European Communities. Cited in M. S. Sanders and E. J. McCormick, *Human Factors in Engineering and Design.* Seventh Ed. New York: McGraw-Hill, 1993, pp. 570–571.

Ramsey, J. D. (1975) "Occupational Vibration." In C. Zenz (ed.), *Occupational Medicine: Principles and Practical Applications,* pp. 553–562.

Ramsey, J. D. (1995) "Task Performance in Heat: A Review." *Ergonomics* 38, no. 1, pp. 154–165.

Ross, D. (1978) "Task Lighting—Yet Another View." *Lighting Design and Application* pp. 37–42.

Rosseger, R., and Rosseger, S. (1960) "Health Effects of Tractor Driving." *Journal of Agricultural Engineering Research* 5, pp. 241–275.

Sanders, M. S., and McCormick, E. J. (1993) *Human Factors in Engineering and Design.* Seventh Ed. New York: McGraw-Hill, pp. 551–588.

Seidel, H., and Heide, R. (1986) "Long-Term Effects of Whole-Body Vibration: A Critical Survey of the Literature." *International Archives of Occupational and Environmental Health* 58, pp. 1–26.

Sinks, T., Mathias, C., Halpern, W., Timbrook, C., and Newman, S. (1987) "Surveillance of Work-Related Cold

Injuries Using Workers' Compensation Claims." *Journal of Occupational Health and Safety* 29, pp. 505–509.

Strydom, N. (1971) "Age as a Causal Factor in Heat Stroke." *Journal of the South African Institute of Mining and Metallurgy* 72, pp. 112–114.

Timbal, J., Loncle, M., and Boutelier, C. (1976) "Mathematical Model of Man's Tolerance to Cold Using Morphological Factors." *Aviation, Space and Environmental Medicine* 47, pp. 958–964.

University of Michigan. (1988) *Energy Expenditure Prediction Model.* Ann Arbor: University of Michigan.

Waller, R. A. (1969) "Office Acoustics—Effect of Background Noise." *Applied Acoustics* 2, pp. 121–130.

Whitham, E. M., and Griffin, M. J. (1978) "The Effects of Vibration Frequency and Direction on the Location of Areas of Discomfort Caused by Whole-Body Vibration." *Applied Ergonomics* 9, no. 4, pp. 231–239.

Wotton, E. (1986) "Lighting the Electronic Office." In R. Lueder (ed.), *The Ergonomics Payoff, Designing the Electronic Office.* Toronto: Holt, Rinehart and Winston, pp. 196–214.

Wright, G., and Rea, M. (1984) "Age, a Human Factor in Lighting." In D. Attwood and C. McCann (ed.), *Proceedings of the 1984 International Conference on Occupational Ergonomics.* Rexdale, Ont., Canada: Human Factors Association of Canada, pp. 508–512.

Wyndham, C. (1974) "Research in the Human Sciences in the Gold Mining Industry." *American Industrial Hygiene Association Journal*, pp. 113–136.

Section **Seven**

Reliability and maintainability

This page is intentionally left blank

Chapter 7.1

7.1

Quality and reliability

J.D. Booker, M. Raines and K.G. Swift

7.1.1 Statement of the problem

In order to improve business performance, manufacturing companies need to reduce the levels of non-conformance and attendant failure costs stemming from poor product design and development. Failure costs generally make up the largest cost category in a manufacturing business and include those attributable to rework, scrap, warranty claims, product recall and product liability claims. This represents lost profit to a business and, as a result, it is the area in which the greatest improvement in competitiveness can be made.

The effect of failure cost or 'quality loss' on the profitability of a product development project is shown in Figure 7.1.1. High levels of failure cost would produce a loss on sales and would probably mean that the project fails to recover its initial level of investment.

In an attempt to combat high quality costs and improve product quality in general, companies usually opt for some kind of quality assurance registration, such as with BS EN ISO 9000. Quality assurance registration does not necessarily ensure product quality, but gives guidance on the implementation of the systems needed to trace and control quality problems, both within a business and with its suppliers. The adoption of quality

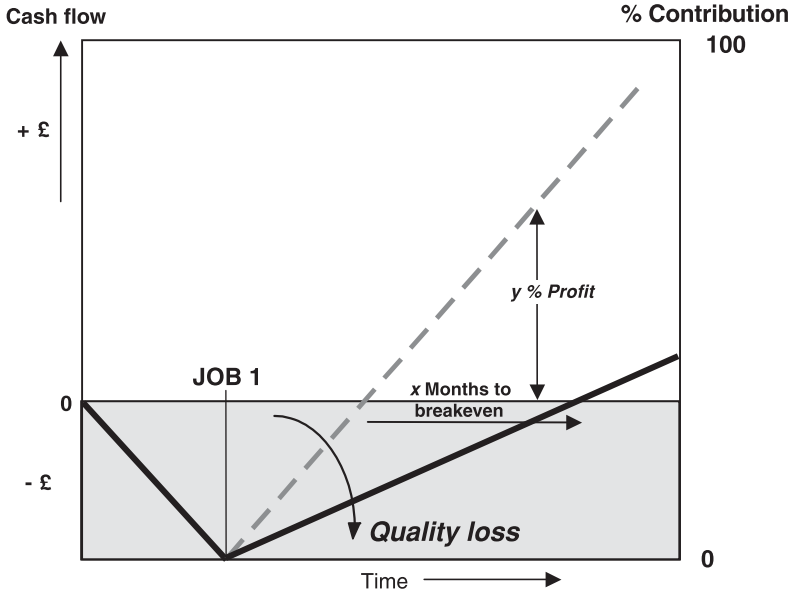


Fig. 7.1.1 Effect of quality loss on the profitability of a product development project.

Designing Capable and Reliable Products; ISBN: 9780750650762
Copyright © 2001 Elsevier Ltd. All rights of reproduction, in any form, reserved.

standards is the only first step in the realization of quality products and also has an ambiguous contribution to the overall reduction in failure costs. A more proactive response by many businesses has been to implement and support long-term product design and development strategies focusing on the engineering of the product.

It has been realized for many years that waiting until the product is at the end of the production line to measure its quality is not good business practice. This has led to an increased focus on the integration of quality into the early design stages of product development. Subsequently, there has been a gradual shift away from the traditional 'on-line' quality techniques, such as Statistical Process Control (SPC), which has been the main driver for quality improvement over the last 50 years, to an 'off-line' quality approach using design tools and techniques.

The focus on quality improvement in design is not misplaced. Studies have estimated that the majority of all costs and problems of quality are created in product development. Focusing on the generation of product faults in product development, we find that typically 75% originate in the development and planning stages, but compounding the problem, around 80% of faults remain undetected until final test or when the product is in use (see Figure 7.1.2). The consequences of a design fault can be crippling: massive recalls, costly modifications, loss of reputation and sales, or even going out of business! Engineers and designers sometimes assume that someone else is causing product costs, but it is the details of how a product is designed that generate its costs in nearly every category.

The most significant cost savings can result from changes in product design rather than, say, from changes in production methods. The costs 'fixed' at the planning and design stages in product development are typically between 60% and 85%, but the costs actually incurred may only be 5% of the total committed for the project. Therefore, the more problems prevented early on, through careful design, the fewer problems that have to be corrected later when they are difficult and expensive to change. It is often the case that quality can be 'built in' to the product without necessarily increasing the overall cost. However, to achieve this we need to reduce the 'knowledge gap' between design and manufacture as illustrated in Figure 7.1.3.

Design is recognized as a major determinant of quality and therefore cost. It is also a driving factor in determining the 'time to market' of products. Historically, designers have concerned themselves with product styling, function and structural integrity. Now the designer has the great responsibility of ensuring that the product will conform to customer requirements, comply to specification, meet cost targets and ensure quality and reliability in every aspect of the product's use, all within compressed timescales.

From the above, it is clear that the designer needs to be aware of the importance of the production phase of product development. As far as quality is concerned, the designer must aim to achieve the standards demanded by

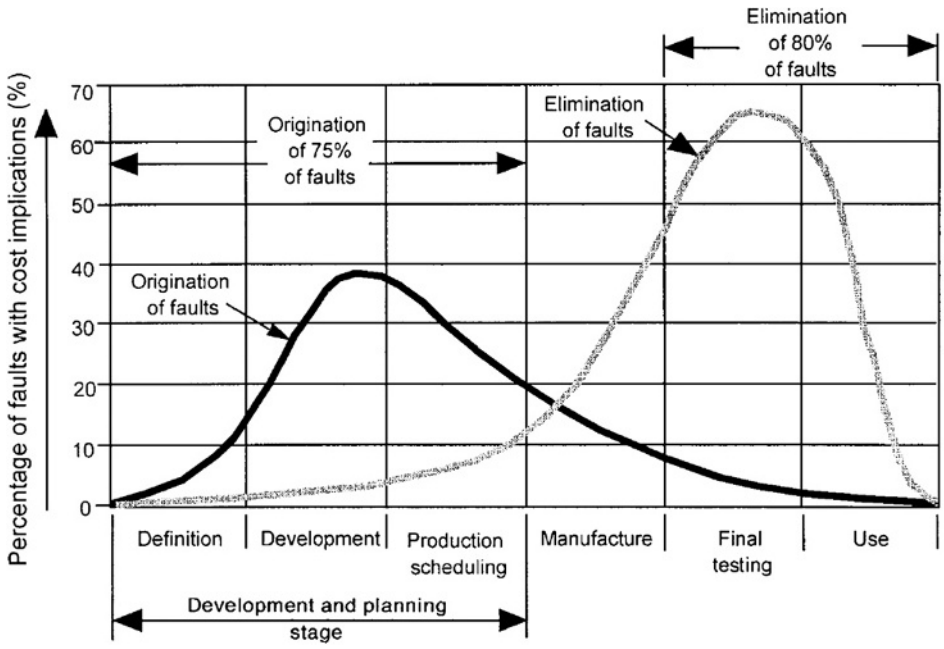


Fig. 7.1.2 Origination and elimination of faults in product development (DTI, 1992).

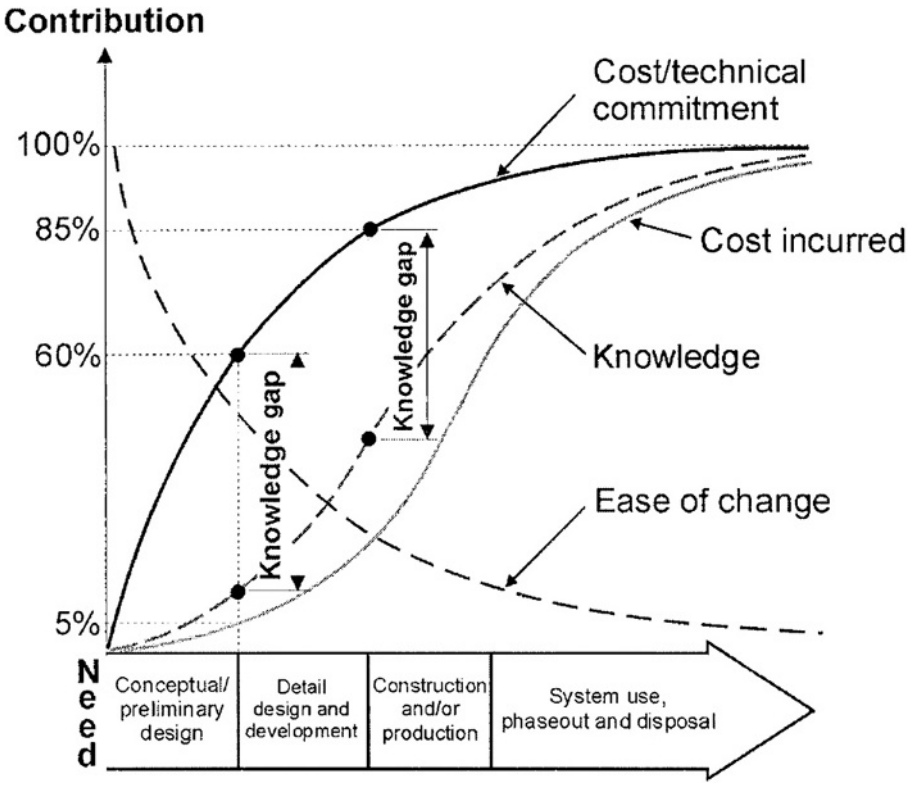


Fig. 7.1.3 Commitment and incursion of costs during product development and the 'knowledge gap' principle (adapted from Fabrycky, 1994).

the specification, but at the same time should be within the capabilities of the production department. Many designers have practical experience of production and fully understand the limitations and capabilities that they must work within. Unfortunately, there are also many who do not. From understanding the key design/manufacture interface issues, the designer can significantly reduce failure costs and improve business competitiveness. One of the most critical interface issues in product development is that concerning the allocation of process capable tolerances.

There is probably no other design improvement effort that can yield greater benefits for less cost than the careful analysis and assignment of tolerances. The effects of assigning tolerances on the design and manufacturing functions are far reaching, as shown in Figure 7.1.4. Product tolerances affect customer satisfaction, quality inspection, manufacturing and design, and are, therefore, a critical link between design, manufacture and the customer. They need to be controlled and understood!

Each product is derived from individual pieces of material, individual components and individual assembly processes. The properties of these individual elements have a probability of deviating from the ideal or target value. In turn, the designer defines allowable tolerances on component characteristics in anticipation of the manufacturing variations, but more often than not, with limited knowledge of the cost implication or manufacturing capability in order to meet the specification. When these variations are too large or off target, the usability of the product for its purpose will be impaired. It therefore becomes important to determine if a characteristic is within specification, and, if so, how far it is from the target value.

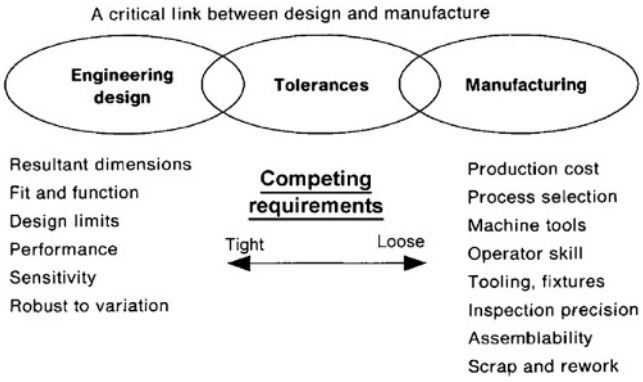


Fig. 7.1.4 Tolerances – the critical link between design and manufacture (Chase and Parkinson, 1991).

Improperly set tolerances and uncontrolled variation are one of the greatest causes of defects, scrap, rework, warranty returns, increased product development cycle time, work flow disruption and the need for inspection. If manufacturing processes did not exhibit variation, quality problems would not arise, therefore reducing the effects of variability at the design stage, in a cost-effective way, improves product quality.

A significant proportion of the problems of product quality can directly result from variability in manufacturing and assembly. However, the difficulties associated with identifying variability at the design stage mean that these problems are detected too late in many cases, as indicated by a recent study of engineering change in nine major businesses from the aerospace, industrial and automotive sectors. On average, almost 70% of product engineering rework was due to quality problems, that is failure to satisfy customer expectations and to anticipate production variability on the shop floor. The need for more than 40% of the rework was not identified until production commenced.

The reasons for the rework, described in Figure 7.1.5, can be classified into four groups:

- Customer driven changes (including technical quality)
- Engineering science problems (stress analysis errors, etc.)
- Manufacturing/assembly feasibility and cost problems
- Production variability problems

This indicates that customer related changes occurred throughout concept design, detailing, prototyping and testing with some amendments still being required after production had began. Engineering science problems, which represented less than 10% of the changes on average, were mostly cleared before production commenced. The most disturbing aspect is the acceptance by the businesses that most of the manufacturing changes, and more so manufacturing variability changes, were taking place during production, product testing and after release to the customer. Because the cost of change increases rapidly as production is approached and passed, the expenditure on manufacturing quality related rework is extremely high. More than 50% of all rework occurred in the costly elements of design for manufacture and production variability.

Further evidence of the problems associated with manufacturing variability and design can be found in published literature. Here, an investigation in the automotive industry showed that of the 26 quality

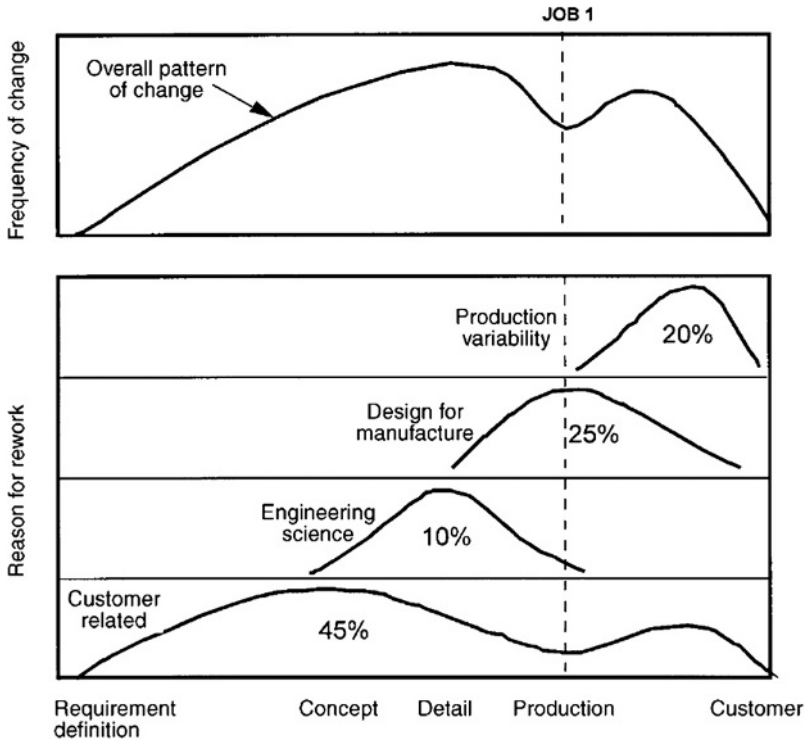


Fig. 7.1.5 Disposition of rework in product development (Swift et al., 1997).

problems stated, 12 resulted from process integrity and the integrity of assembly. Process integrity was defined as the correct matching of the component or assembly design to either the current manufacturing process or subsequent processes. Integrity of assembly was defined as the correct matching of dimensions, spatial configuration of adjacent or interconnecting components and subassemblies.

Variability associated with manufacturing and assembly has historically been considered a problem of the manufacturing department of a company. It is now being recognized that there is a need to reduce such variations at the design stage, where its understanding and control may lead to:

- Easier manufacture
- Improved fit and finish
- Less work in progress
- Reduced cycle time
- Fewer design changes
- Increased consistency and improved reliability
- Better maintainability and repairability.

Variation is an obvious measure for quality of conformance, but it must be associated with the requirements set by the specification to be of value at the design stage. Unfortunately, difficulty exists in finding the exact relationship between product tolerance and variability. Approximate relationships can be found by using process capability indices, quality metrics which are interrelated with manufacturing cost and tolerance.

The first concern in designing process capable products is to guarantee the proper functioning of the product, and therefore to satisfy technical constraints. Dimensional characteristics reflect the spatial configuration of the product and the interaction with other components or assemblies. Tolerances should be allocated to reflect the true requirements of the product in terms of form, fit and function in order to limit the degradation of the performance in service. Ideally, designers like tight tolerances to assure fit and function of their designs. All manufacturers prefer loose tolerances which make parts easier and less expensive to make.

Tolerances alone simply do not contain enough information for the efficient manufacture of a design concept and the designer must use process capability data when allocating tolerances to component characteristics. Process capability analysis has proven to be a valuable tool in this respect, and is most useful when used from the very beginning of the product development process.

If the product is not capable, the only options available are to either: manufacture some bad product, and sort it out by inspection; rework at the end of the production line; narrow the natural variation in the process; or widen the specification to improve the capability. Post-production inspection is expensive and widening the specification is not necessarily desirable in some applications as this may have an impact on the functional characteristics of the product. However, in many cases the tolerance specification may have been set somewhat arbitrarily, implying that it may not be necessary to have such tight tolerances in the first place. Making the product robust to variation is the driving force behind designing capable and reliable products, lessens the need for inspection and can reduce the costs associated with product failure.

Variability must become the responsibility of the designer in order to achieve these goals. An important aspect of the designer's work is to understand the tolerances set on the design characteristics, and, more importantly, to assess the likely capability of the characteristics due to the design decisions.

Industry is far from understanding the true capability of their designs. Some comments from senior managers and engineers in the industry give an indication of the cultural problems faced and the education needed to improve design processes in this respect.

We will have difficulty meeting those tolerances – it is 'bought-in' so we will get the supplier to do the inspection.

$C_{pk} = 1.33$! We do much better than that in the factory. We are down to $C_{pk} = 0.8$!

I do not see how we make this design characteristic at $C_{pk} = 1.5$. Let us kill it with 100% inspection.

The components are not going to be process capable, but we can easily set the tolerance stack at ± 0.1 mm when we build the assembly machine. Our assembly machine supplier uses robots.

I can see that this design is not likely to be capable, but my new director has said we have to use this design solution because it has the lowest part count.

I can't spend any more time on design. I see the problems, but it will cost the department too much if I have to modify the design.

I have been told that we must not use any secondary machining operations to meet the tolerance requirements. It just costs too much!

Good design practice does not simply mean trying to design the product so that it will not fail, but also identifying how it might fail and with what consequences. Effective understanding of the quality of conformance associated with design decisions requires undertaking a number of engineering activities in the early stages of product development. In addition to understanding the capability of the design, the designer must consider the severity of potential failures and make sure the design is sufficiently robust to effectively eliminate or accommodate defects. Effective failure analysis is an essential part

of quality and reliability work, and a technique useful in this capacity is Failure Mode and Effects Analysis (FMEA).

FMEA is a systematic element by element assessment to highlight the effects of a component, product, process or system failure to meet all the requirements of a customer specification, including safety. FMEA can be used to provide a quantitative measure of the risk for a design. Because FMEA can be applied hierarchically, through subassembly and component levels down to individual dimensions and characteristics, it follows the progress of the design into detail listing the potential failure modes of the product, as well as the safety aspects in service with regard to the user or environment. Therefore, FMEA provides a possible means for linking potential variability with consequent design acceptability and associated failure costs. The application of a technique that relates design capability to potential failure costs incurred during production and service would be highly beneficial to manufacturing industry.

Conceivably, a number of new issues in product design and development have been discussed in this opening section, but in summary:

- Understanding and controlling the variability associated with design characteristics is a key element of developing a capable and reliable product.
- Variability can have severe repercussions in terms of failure costs.
- Designers need to be aware of potential problems and shortfalls in the capability of their designs.
- There is a need for techniques which estimate process capability, quantify design risks and estimate failure costs.

Next, we review the costs of quality that typically exist in a manufacturing business, and how these are related to the way products fail in service. The remainder of the chapter discusses the important elements of risk assessment as a basis for design. This puts in context the work on designing for quality and reliability.

7.1.2 The costs of quality

The costs of quality are often reported to be between 5% and 30% of a company's turnover, with some engineering businesses reporting quality costs as high as 36%. This figure can be as high as 40% in the service industry! In general, the overall cost of quality in a business can be divided into the following four categories:

- *Prevention costs* – These are costs we expect to incur to get things right first time, for example quality planning and assurance, design reviews, tools and techniques, and training.
- *Appraisal costs* – Costs which include inspection and the checking of goods and materials on arrival. Whilst an element of inspection and testing is necessary and justified, it should be kept to a minimum as it does not add any value to the project.
- *Failure costs* – *Internal failure costs* are essentially the cost of failures identified and rectified before the final product gets to the external customer, such as rework, scrap, and design changes. *External failure costs* include product recall, warranty and product liability claims.
- *Lost opportunities* – This category of quality cost is impossible to quantify accurately. It refers to the rejection of a company product due to a history of poor quality and service, hence the company is not invited to bid for future contracts because of a damaged reputation.

Up to 90% of the total quality cost is due to failure, both internal and external, with around 50% being the average. A survey of UK manufacturing companies in 1994 found that failure under the various categories was responsible for 40% of the total cost of quality, followed by appraisal at 25%, and then prevention costs at 18%. This is shown in Figure 7.1.6. Of the companies surveyed, 17% were unsure where their quality costs originated, but indicated that these costs could be attributable to failure, either internally or externally.

Many organizations fail to appreciate the scale of their quality failures and employ financial systems which neglect to quantify and record the true costs. In many cases, the failures are often costs that are logged as 'overheads'. Quality failure costs represent a direct loss of profit! Organizations may have financial systems to recognize scrap, inspection, repair and test, but these only represent the 'tip of the iceberg' as illustrated in Figure 7.1.7.

A company should minimize the failure costs and appraisal costs, but be prepared to increase investment

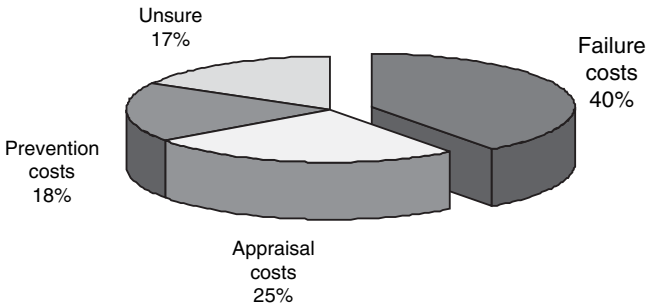


Fig. 7.1.6 The costs of quality in UK industry (Booker, 1994).

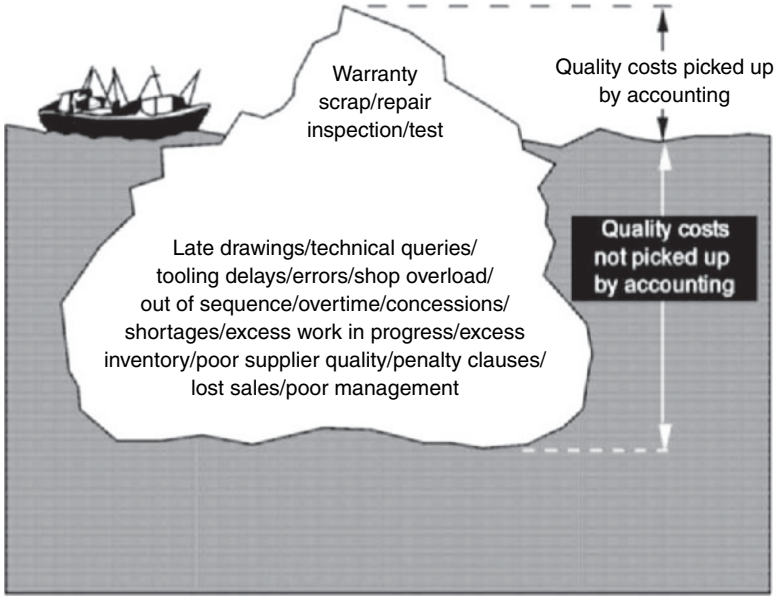


Fig. 7.1.7 The hidden cost of poor quality (Labovitz, 1988).

on prevention. Some quality gurus promote the philosophy of zero defects. Whilst this is obviously the ultimate goal, the prevention costs can become prohibitive. It is possible to determine the optimum from a cost point of view. This value may not be constant across the different business sectors, for instance a machine shop may be prepared to accept a 1% scrap rate, but it is doubtful that the public would accept that failure rate from a commercial aircraft! Each must set its own objectives, although 4% has been stated in terms of a general target as a percentage of total sales for manufacturing companies. A simple quality-cost model that a business can develop to define an optimum between quality of conformance and cost is illustrated in Figure 7.1.8.

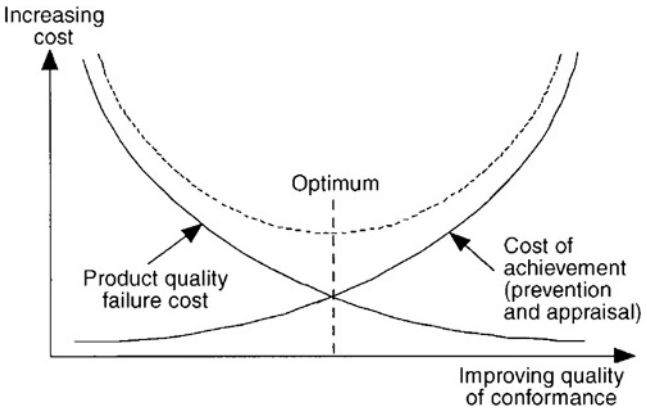


Fig. 7.1.8 The optimization of quality costs.

7.1.2.1 Case studies in failure costs

External failure costs and lost opportunities are potentially the most damaging costs to a business. Several examples commonly quoted in the literature are given below.

This first example applies to UK industry in general. The turnover for UK manufacturing industry was in the order of £150 billion in 1990. If the total quality cost for a business was likely to be somewhere in the region of 20%, with failure costs at approximately 50% of the total, it is likely that about £15 billion was wasted in defects and failures. A 10% improvement in failure costs would have released an estimated £1.5 billion into the economy. IBM, the computer manufacturer, estimated that they were losing about $5.6 billion in 1986 owing to costs of nonconformance and its failure to meet quality standards set for its products and services. A further $2 billion was estimated as being lost as a result of having poor working processes. IBM proceeded on the basis that the company had $7.6 billion of potential savings to be obtained getting things 'right first time'.

In the US over the last 30 years, there has been an increasing trend in product liability claims and associated punitive damages. For example, after a legal battle, General Motors had to pay a plaintiff's family $105 million, when the plaintiff was killed when a poorly positioned petrol tank in a truck exploded during an accident. The UK motorcycle industry in the 1970s suffered greatly due to their Japanese counterparts not only producing more cost-effective bikes, but also of higher

quality. The successful resurgence of Triumph only recently was based on matching and even bettering the Japanese on the quality of it products.

Most producers believe in the adage 'quality pays' in terms of better reputation and sales, customer loyalty, lower reject rates, service and warranty costs. They should also realize that 'safety pays' in terms of reducing the legal exposure and the tremendous costs that this can incur, both directly and indirectly, for example from compensation, legal fees, time and effort, increased insurance premiums, recalls and publicity. Few manufacturers understand all the cost factors are involved, and many take a shortsighted view of the actual situation with regard to the costs of safety.

Measures to minimize safety problems must be initiated at the start of the life cycle of any product, but too often determinations of criticality are left to production or quality control personnel who may have an incomplete knowledge of which items are safety critical. Any potential non-conformity that occurs with a severity sufficient to cause a product or service not to satisfy intended normal or reasonably foreseeable usage requirements is termed a 'defect'. The optimum defect level will vary according to the application, where the more severe the consequences of failure the higher the quality of conformance needs to be.

The losses that companies can face are influenced by many factors including market sector, sales turnover and product liability history. It is not easy to make a satisfactory estimate of the product liability costs associated with quality of non-conformance, and losses due to safety critical failures in particular are subject to wide variation.

It is known that product liability costs in the US have risen rapidly in recent years and this trend is set to continue. It has been predicted that US tort costs would reach $300 billion by the year 2000, which would then represent 3.5% of US economic output. Tort is a term used in common law for a civil wrong and for the branch of law dealing with the liability for these wrongs. It is an alleged wrongful act for which the victim can bring a civil action to seek redress. Examples of individual torts are negligence, nuisance, and strict liability. Tort cost growth far outstripped Gross National Product (GNP) growth since 1930, increasing 300 times over this period, as shown in Figure 7.1.9, compared with a 50-fold increase for GNP.

The way in which tort costs are moving provides valuable evidence of the costs of 'getting it wrong'. Product liability experts believe that while the US system has its differences, as lawyers in the UK become more attuned to the US system, a similar situation may occur here also. Some background to the situation in the UK is shown in Figure 7.1.10, indicating that product liability costs could reach £5 billion annually. A further

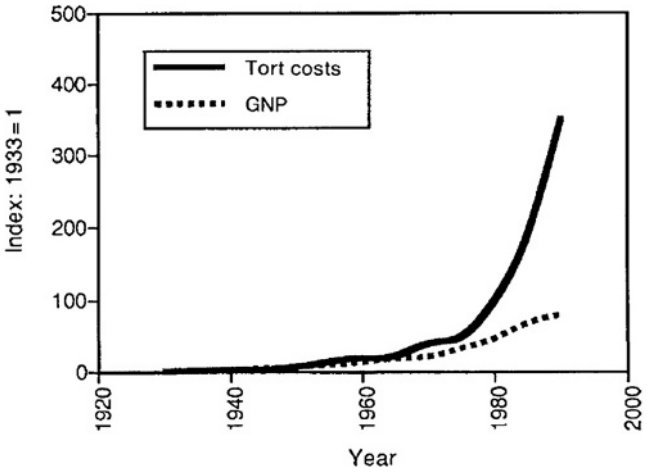


Fig. 7.1.9 US tort cost escalation compared with GNP growth (Sturgis, 1992).

escalation in product liability claims could result in higher insurance premiums and the involvement of insurance companies in actually defining quality and reliability standards and procedures.

Insurance appears to be the safest solution for companies to defend themselves against costly mistakes, but there are problems, notably the cost of the premiums and the extent of the cover provided. The insurance sector must address some of the above issues in assessing their exposure. Case histories provide some insight into the costs that can accrue though.

For a catastrophic failure in the aerospace industry with a high probability of loss of life, which relates to an FMEA Severity Rating (S) = 10, a business could quite possibly need insurance cover well in excess of

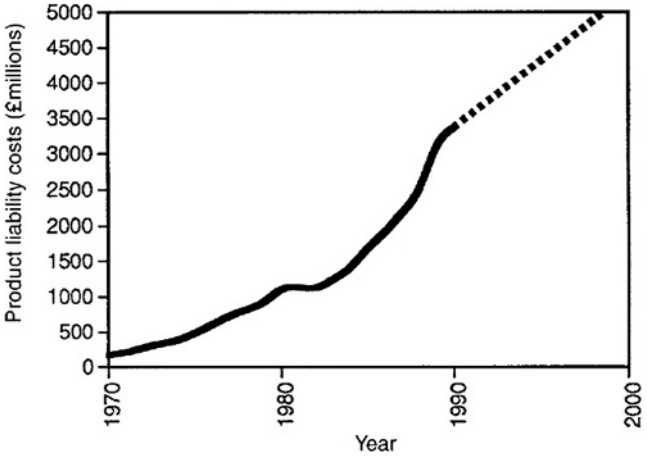


Fig. 7.1.10 Trend of product liability costs in the UK (Sturgis, 1992).

£100 million. This will allow for costs due to failure investigations, legal actions, product recall and possible loss of business. High failure costs are not only associated with the aerospace industry. Discussions relating to the automotive sector suggest that for a failure severity of $S = 9$, complete failure with probable severe injury and/or loss of life, a business could well face the need for cover in excess of £10 million. Less safety critical business sectors and lower severity ratings reduce the exposure considerably, but losses beyond £1 million have still been recorded. The relationship between safety critical failures and potential cost is summarized in Figure 7.1.11. It is evident that as failures become more severe, they cost more, so the only approach available to the designer is to reduce the probability of occurrence.

Little progress towards reducing product liability exposure can be made by individuals within a business unless top management are committed to marketing safe products. Safety is one aspect of the overall quality of a product. While most producers and suppliers realize the importance of quality in terms of sales and reputation, there is sometimes less thought given to the importance of safety in terms of legal liabilities. Where 'quality' is mentioned it should be associated with 'safety'. Management strategy should be based on recognizing the importance of marketing a safe product, and the potential costs of failing to do so, and that failures will occur and plans must be made for mitigating their effect.

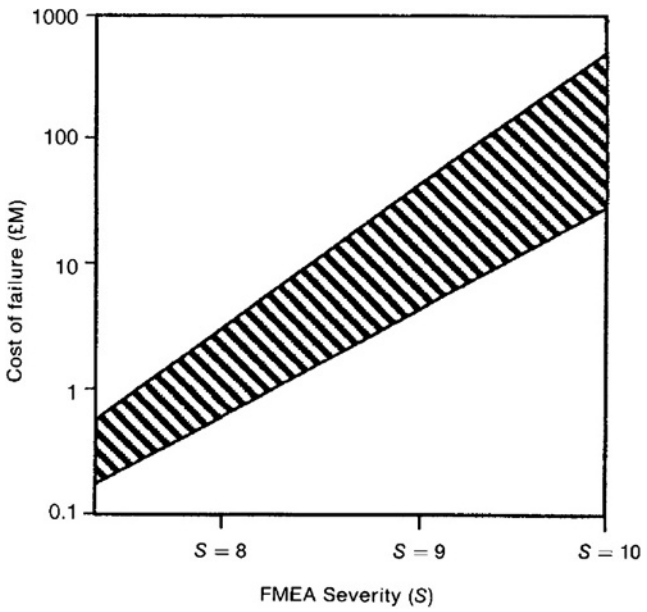


Fig. 7.1.11 The potential cost of safety critical failures.

7.1.2.2 Quality-cost estimating methods

Quality-cost models can help a business understand the influence of defect levels on cost during product development. More importantly, designers should use models to predict costs at the various stages. These results make the decision-making process more effective, particularly at the design stage. The estimation of quality costs in the literature is commonly quoted at three quite different levels:

- Economic quality-cost models which are 'global' or 'macro-scaling' top down methods, show general trends in quality costs which are predicted based on varying some notion of time or quality improvement. The model in the latest standard BS 6143, as shown in Figure 7.1.12, is such a model. It is perhaps closest to the view of some quality experts, but surprisingly infers that prevention costs reduce substantially with increased quality improvement. A model published recently also combines failure and appraisal costs, two distinct categories. Quality managers believe that many of the widely publicized quality-cost models are inaccurate and may even be of the wrong form. A valid model that could be used to audit business performance and predict the effects of change would be most helpful. However, since the modelling of quality costs range from inaccurate to questionable, they are unlikely to provide a rigorous basis for product engineering to connect design decisions to the costs of poor quality.
- 'Micro-scaling' or bottom up approach to quality costs, where it is possible to calculate the cost of losses involved in manufacture and due to returns

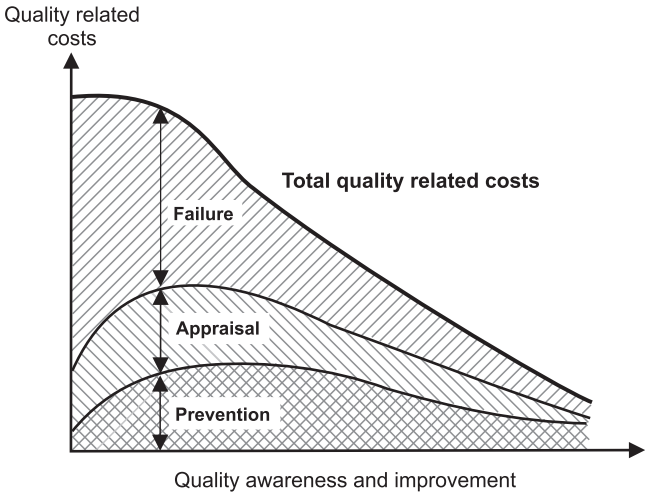


Fig. 7.1.12 Global quality-cost model (BS 6143,1990).

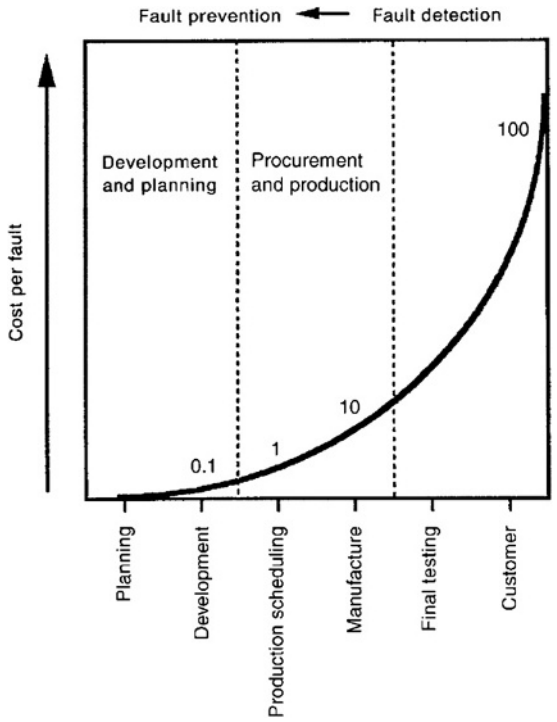


Fig. 7.1.13 The 10 x rule of fault related costs by percentage (DTI, 1992).

and/or claims. This method requires a great deal of experience and relies on the availability of detailed cost data throughout a product's life-cycle. While this is a crucial activity for a business, it is also not a practical approach for estimating the quality cost for product in the early stages of product development.

- The Conformability Analysis (CA) method and Taguchi's Quality Loss Function are what might be called 'meso-scaling' or quality-cost scaling. Here past failure costs are scaled to new requirements allowing for changes in design capability. It gives more precision than the global approach, but would clearly lack the accuracy of the bottom up approach possible once a product is in production. However, these methods are more useful at the design stage.

We have already seen elements of the CA approach when considering the costs due to safety critical failures. A further insight into the way that failure costs can be estimated for non-safety critical failures is also used to support the CA methodology. Estimates for the costs of failure in this category are based on the experiences of a sample of industrial businesses and published material as follows.

Consider a product whose product cost is *Pc*. The costs due to failure at the various stages of the product's life cycle have been investigated, and in terms of *Pc*, they have been found to be:

0.1 *Pc* – internal failure cost due to rework at the end of the production line

Pc – external failure cost for return from customer inspection

10 *Pc* – external failure cost for warranty return due to failure with customer in use

The relationship is commonly known as the '10× rule' and is shown in Figure 7.1.13. The 10× rule demonstrates how a fault, if not discovered, will give rise to 10 times the original elimination costs in a later phase of the life-cycle. In other words, products must be designed in such a way that scarcely any defects develop or if they do, they can be identified as early as possible in the product development process and rectified. Other surveys have found that these costs could be even higher as shown in Figure 7.1.14.

Suppose a particular fault in a product is not detected through internal tests and inevitably results in a failure severity $S = 5$. If around 80% of failures are found by customer testing and 20% are warranty returns, then the expected cost on average for one fault will be 2.8 *Pc*, from Figure 7.1.13. If the product has been designed such that $C_{pk} = 1.33$, or in other words, approximately 30 parts-per-million (ppm) failures are expected for the characteristic which may be faulty, then for a product costing £100 the probable cost of failure per million products produced would be £8400. At $C_{pk} = 1$, or approximately 1300 ppm failures expected, the probable cost of failure per million products would be £364,000. At $C_{pk} = 0.8$ (or 8000 ppm) the probable cost of failure would increase by an order of magnitude to over £2.2 million. These failure costs do not take into account the costs associated with damaged company reputation and lost opportunities which are difficult to assess, but do indicate that failure cost estimates associated with product designs are possible. This aspect of the CA methodology is further developed in Chapter 7.2.

7.1.3 How and why products fail

7.1.3.1 Failure mechanisms

We have already established that variability, or the lack of control and understanding of variability, is a large determinant of the quality of a product in production and service and, therefore, its success in avoiding failure. In addition, understanding the potential failure mechanisms and how these interact with design decisions is necessary to develop capable and reliable products. It is helpful next to investigate the link between the causes and

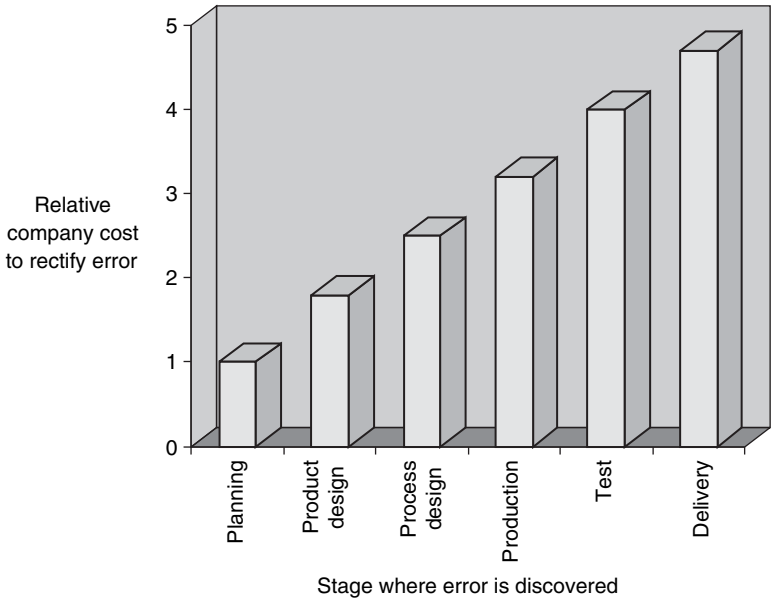


Fig. 7.1.14 Cost escalation of rectifying errors at down stream stages (Ostrowski, 1992).

modes of failure and variability throughout the life-cycle of a mechanical product.

Mechanical failure is any change or any design or manufacturing error that renders a component, assembly or system incapable of performing its intended function. However, it is also possible to suggest several key aspects of failure:

- A failure is said to occur when disappointment arises as a result of an assessment of an outcome of an activity.
- Failure can be a shortfall of performance below a standard, the generation of undesirable effects or the neglect of an opportunity.
- Failure can occur in a variety of forms, namely: catastrophic or minor, overwhelming or only partial, sudden or slow.
- Failure can arise in the past, present or future.
- Failure will be found to be multi-causal, and to have multiple effects.

For the product to fail there must be some failure mechanism caused by lack of control of one or more of the engineering variables involved. Most mechanisms of mechanical failure can be categorized by one of the following failure processes:

- Overload – static failure, distortion, instability, fracture
- Strength degradation – creep failure, fatigue, wear, corrosion.

Figure 7.1.15(a) shows the results of an investigation to find the frequency of failure mechanisms in typical engineering components and aircraft components. By far the most common failure mechanism is fatigue. It has been suggested that around 80% of mechanical failures can, in fact, be attributable to fatigue. Failures caused by corrosion and overload are also common. Although the actual stress rupture mode of failure is cited as being uncommon, overload and brittle fracture failures may also be categorized as being rupture mechanisms, distinct from strength degradation.

Figure 7.1.15(b) provides some insight into the reasons for the mechanical failures experienced. Some root causes of failure are found to be improper material selection, fabrication imperfections and design errors. Other causes of failure are ultimately related to variability in production and service conditions. In general, several primary root causes of failure can be suggested:

- Design errors
- Production errors (manufacturing, assembly)
- Handling/transit damage
- Misuse or operating errors
- Adverse environmental conditions.

From the above, there is a strong indication that the design errors are responsible for many mechanical failures found in the field. This is supported by another study, the results of which are shown in Figure 7.1.16. Through making poor or uneducated design decisions, the designer alone accounted for around 35% of the failures of those investigated.

Historically, the failure of products over their life-cycles, sometimes termed the 'bath-tub curve', can be classified into three distinct regions as shown in

Mechanism	Percentage of failures	
	Engineering components	Aircraft components
Corrosion	29	3
Fatigue	25	61
Brittle fracture	16	–
Overload	11	18
High temperature corrosion	7	2
SCC/ corrosion fatigue/ HE	6	8
Creep	3	–
Stress rupture	–	1
Wear/abrasion/ erosion	3	7

(a) Frequency of failure mechanisms

Cause	Percentage of failures	
	Engineering components	Aircraft components
Improper material selection	38	–
Fabrication imperfections	15	17
Faulty HT	15	–
Design errors	11	16
Unanticipated service conditions	8	10
Uncontrolled environmental conditions	6	–
Inadequate inspection/QC	5	–
Material mix	2	–
Inadequate maintenance	–	44
Defective material	–	7
Unknown	–	6

(b) Frequency of causes of failure

Key: SCC - Stress corrosion cracking, HE - Hydrogen embrittlement, HT - Heat treatment, QC - Quality control

Fig. 7.1.15 The frequency and causes of mechanical failure (Davies, 1985).

Figure 7.1.17. A detailed breakdown of the attributable factors in each region is also given below.

Infant mortality period – Quality failures dominate and occur early in the life of the product. In detail, these can be described as:

- Poor manufacturing techniques including processes, handling and assembly practices
- Poor quality control
- Poor workmanship
- Substandard materials and parts
- Parts that failed in storage or transit
- Contamination
- Human error
- Improper installation.

Useful life period – Stress-related failures dominate and occur at random over the total system lifetime – caused by the application of stresses that exceed the design's strength (most significant period as far as reliability prediction activities are concerned). In detail, these can be described as:

- Interference or overlap of designed-in strength and experienced stress during operation
- Occurrence of higher than expected random loads
- Occurrence of lower than expected random strengths

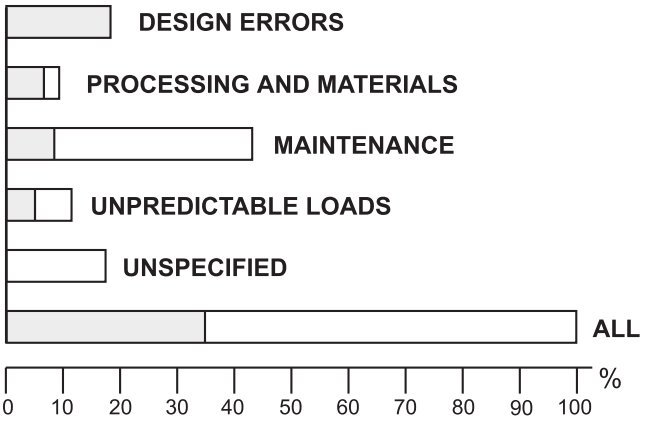


Fig. 7.1.16 Designer's responsibility for mechanical failures (designer's share is shaded) (Larsson et al., 1971).

- Defects
- Misapplication or abuse
- 'Acts of God'.

Wearout period – Failure occurs when the product reaches the end of its effective life and begins to degenerate and wear out. In detail, these can be described as:

- Ageing
- Wear
- Fatigue
- Creep
- Corrosion
- Poor servicing or maintenance.

7.1.3.2 The link between variability and failure

There exists a relationship between the failure characteristics of a product over its life-cycle, as described by the three periods of the bath-tub curve above, and the phenomenon of variability. It has already been established that the potential for variation in design parameters is a real aspect of product engineering. Subsequently, three major sources of undesirable variations in products can be classified, these being:

- Production variations
- Variations in conditions of use
- Deterioration (variation with time and use).

The above fits in with the overall pattern of failure as described by Figure 7.1.17. The first two and sometimes even all the three parts of the bath-tub curve are closely connected to variations.

The manufacturing process has a strong impact on component behaviour with respect to failure, and production variabilities arising from lack of precision or deficiencies in manufacturing processes lead to failures concentrated early in the product's life. A common occurrence is that correctly designed items may fail as a result of defects introduced during production or simply because the specified dimensions or materials are not complied with. Production defects are second only to those product deficiencies created by inadequate design as causes of accidents and improper production techniques can actually create hazardous characteristics in products.

Modern equipment is frequently composed of thousands of components, all of which interact within various tolerances. Failures often arise from a combination of drift conditions rather than the failure of a specific component. For example, typically an assembly tolerance exists only to limit the degradation of the assembly performance. Being 'off target' may involve later warranty costs because the product is more likely to break down than one which has a performance closer to the target value. This again is related to manufacturing variation problems, and is more difficult to predict, and therefore less likely to be foreseen by the designer.

Variations in a product's material properties, service loads, environment and use typically lead to random failures over the most protracted period of the product's expected life-cycle. During the conditions of use, environmental and service variations give rise to temporary

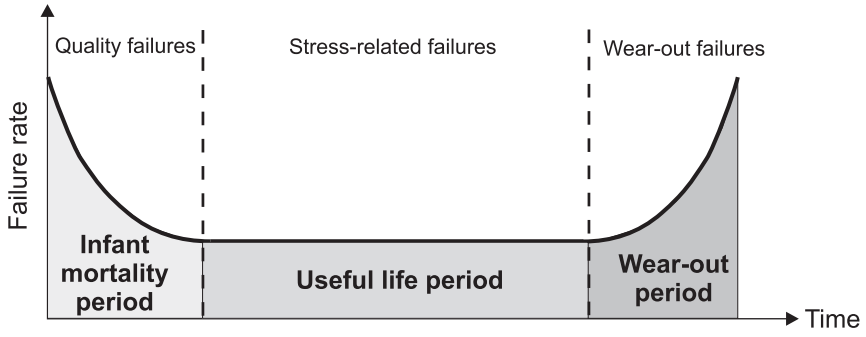


Fig. 7.1.17 'Bath-tub' curve showing typical life characteristic of a product (Priest, 1988).

overloads or transients causing failures, although some failures are also caused by human related events such as installation and operation errors rather than by any intrinsic property of the product's components. Variability, therefore, is also the source of unreliability in a product. However, it is evident that if product reliability is determined during the design process, subsequent manufacturing, assembly and delivery of the system will certainly not improve upon this inherent reliability level.

Wearout attracts little attention among designers because it is considered less relevant to product reliability than the other two regions, although degradation phenomena are clearly important for designs involving substantial operating periods.

Many kinds of failures described above may be reduced by either decreasing variations or by making the product robust against these variations. For example, the smaller the variability associated with the critical design parameters, the greater will be the reliability of the design to deal with unforeseen events later in the product's life-cycle. As stated earlier, with reference to an individual manufacturing process, variation is an obvious measure for quality performance and the link among product failure, capability and reliability is to a large degree embedded within the prediction of variation at the design stage. However, two factors influence failure: the robustness of the product to variability, and the severity of the service conditions. To this end, it has been cited that the quality control of the environment is much more important than quality control of the manufacturing processes in achieving high reliability.

In order to quantify the sometimes intangible elements of variability associated with the product design and the safety aspects in service requires an understanding of 'risk'. The assessment of risk in terms of general engineering practice will be discussed next. This will lead to a better understanding of designing for quality and reliability.

7.1.4 Risk as a basis for design

The science of risk, and the assessment and management of risk, is a very complex subject and one that covers a wide diversity of disciplines. Society is becoming more aware of the risks related to increased technological innovation and industrialization. Recent reports in the media about environment (global warming), health (BSE) and technological (nuclear waste processing) risks have played their part in focusing attention on the problem of how to assess risk and what makes an acceptable risk level. As a result, risk and risk-related matters are becoming as important as economic issues on the political agenda. There is, therefore, an increasing need for a better understanding of the topic and tools and techniques that can be used to help assess product safety and support the development of products and processes that are of essentially low risk. To meet these needs, a new British Standard, BS 6079, has been published to give guidance to businesses on the management of risks throughout the life of a project.

The term 'risk' is often used to embrace two assessments:

- The frequency (or probability) of an event occurring.
- The severity or consequences of the event on the user/environment.

The product of these two conditions equals the risk:

$$\text{Risk} = \text{Occurrence} \times \text{Severity}$$

We can demonstrate the notions of risk and risk assessment using Figure 7.1.18. For a given probability of failure occurrence and severity of consequence, it is possible to map the general relationship of risk and what this means in terms of the action required to eliminate the risk.

For example, if both occurrence and severity are low, the risk is low, and little or no action in eliminating or accommodating the risk is recommended. However, for the same level of occurrence but a high severity, a medium level of risk can be associated with concern in

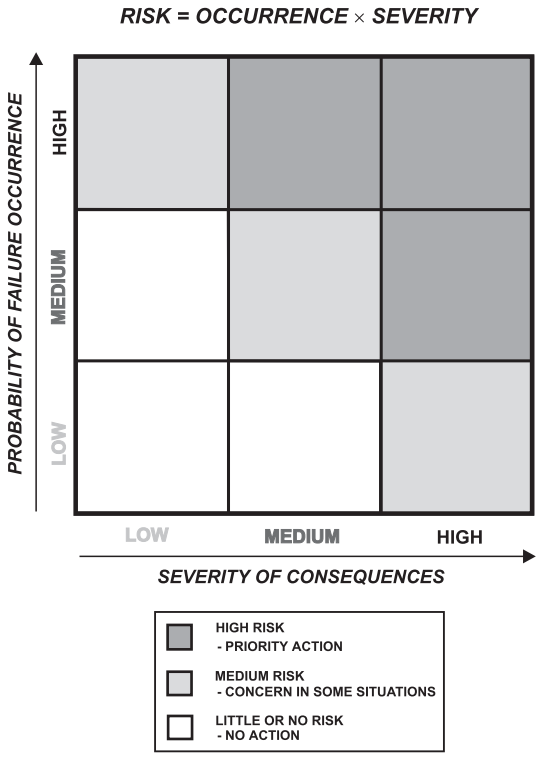


Fig. 7.1.18 Risk and the identification of priority action.

some situations. The level of occurrence, for some unknown reason, changes from low to medium and suddenly we are in a situation where the risk requires priority action to be eliminated or accommodated in the product.

The aim of a risk assessment is to develop a product which is 'safe' for the proposed market. A safe product is any product which, under normal or reasonably foreseeable conditions of use, including duration, presents no risk or only the minimum risk compatible with the product's use and which is consistent with a high level of protection for consumers (DTI, 1994). In attempting to protect products against failure in service and, therefore, the user or environment, difficulty exists in ascertaining the degree of protection most suitable for a given application. First, the choice should be expressed in terms of risk and probability of failure as shown above, but the determination of an acceptable level for a product depends on many factors, such as:

- Safety
- The costs of failure
- Criticality of function that it supports
- Complexity – number of component parts, subsystems
- Operational profile – duty cycle or time it operates
- Environmental conditions – exposure to various environmental conditions
- Number of units to be produced
- Ease and cost of replacement
- 'State of the art' or present state of engineering progress
- Market sector/consumer category.

Suggestions have also been made as to the number of people that are affected by the risk at any one time, and, in principle, it should be possible to link the acceptable level of individual risk to the number of people exposed to that risk. For example, Versteeg provides risk levels associated with three areas: acceptable risk, reduction desired and unacceptable risk, and the number of people exposed to the risk. This further compounds the problem of assigning acceptable risk targets, but implies that safety is of paramount importance.

Businesses make decisions that affect safety issues, but which are only considered implicitly. In an increasingly complex world, the resulting decisions are not always appropriate because the limits of the human mind do not allow for an implicit consideration of a large number of factors. Formal analyses are needed to aid the decision-making process in these complex situations. However, the application of a formal analysis to safety issues raises new questions. The risks perceived by society and by individuals cannot be captured by a simple technical analysis. There are many reasons for this; however, it is clear that decision making needs to account for both technical and public values. The decision to accept risk is not based on the absolute notion of one acceptable risk level, but has some flexibility as the judgement depends on the cost/benefit and the degree of voluntariness. The notion of safety is often used in a subjective way, but it is essential to develop quantitative approaches before it can be used as a functional tool for decision making . A technique which 'quantifies' safety is FMEA.

7.1.4.1 The role of FMEA in designing capable and reliable products

In light of the above arguments, it has been found that there are two key techniques for delivering quality and reliability in new products: process capability analysis and FMEA. FMEA is now considered to be a natural tool to be used in quality and reliability improvement and it has been suggested that between 70% and 80% of potential failures could be identified at the design stage by its effective use. For example, performing a comprehensive FMEA well will alleviate late design changes.

FMEA was first mentioned at the starting of this chapter. In general, an FMEA does the following:

- Provides the designer with an understanding of the structure of the system, and the factors which influence quality and reliability.
- Helps to identify items that are of high risk through the calculation of the Risk Priority Number (RPN), and so gives a means of deciding priorities for corrective action.
- Identifies where special effort is needed during manufacture, assembly or maintenance.
- Establishes if there are any operational constraints resulting from the design.
- It gives assurance to management and/or customers that quality and reliability are being or have been properly addressed early in the project.

Of the many characteristics of a product defined by the dimensions and specifications on a drawing, only a few are critical to fulfilling the product's intended function.

Hence, a critical characteristic is defined as one in which high variation could significantly affect product safety, function or performance. In order to assess the level of importance of the characteristics in a design, a process of identifying the critical characteristics and then using special symbols on the detailed drawing is commonly used. For example, the symbol '▼' is used by some companies to indicate that a particular characteristic should be controlled during manufacture using SPC. The identification process is facilitated by the use of a design FMEA using multi-disciplined teams.

As seen in Figure 7.1.19, the important results from an FMEA in terms of designing capable and reliable products are the potential failure modes, severity rating and critical characteristics for the design. By identifying the capability of the critical characteristics, and the potential failure mode, a statistical analysis can then be performed to determine its reliability. The FMEA Severity Rating (S) is crucial for setting capability and reliability targets because it is a useful indication of the level of the safety required for the application. Although subjective in nature, the effective use of an FMEA in the design process is advocated as it brings significant benefits.

In summary, to reduce risk at the product design stage requires that we do not speculate without supporting evidence on the causes, consequences and solutions for an actual or potential design problem. This means making predictions, where appropriate, based on evidence from testing, experience or other hard facts using statistical probabilities, not vague guesses. The use of FMEA to evaluate all the potential risks of failure and their consequences, both from normal use and foreseeable misuse, is a key element in designing capable and reliable products.

7.1.5 Designing for quality

The improvement of the quality of a design is seen as the primary need of industry, but to facilitate this we need appropriate methods for predicting quality and evaluating the long-term quality of an engineer's design. However, there is relatively little work published in the field of Design for Quality (DFQ) compared to Design for Assembly (DFA), for example, and little methodology exists as yet. One possible reason for this is that DFQ methods should have the objective of selecting the 'technically perfect' from a number of alternative solutions. The word quality, therefore, implies a relative rather than a precise standard from which the designer has to work. This requires a cultural shift of thought in design activities.

Techniques such as FMEA, DFA and Quality Function Deployment (QFD) can enhance the success of a product, but alone they will not solve all product development issues. They provide useful aids in the process of quality improvement, but they do not ensure product quality. There exists an important need for DFQ techniques to aid design and support the product development process. In addition, it has been cited that in order to make further reductions in product development time requires new progress in these techniques.

A substantial review of DFQ and the framework for its application has been proposed by Mørup in 1993. This identifies eight key elements in DFQ which are placed under the headings of preconditions, structured product development and supporting methods/tools and techniques as illustrated in Figure 7.1.20. In thinking about DFQ, Mørup states that it is convenient to divide product quality into two main categories:

- 'Big **Q**' which is the customer/user perceived quality.
- 'Little **q**' which relates to our efforts in creating big **Q**.

Product quality is a vector with several types of quality elements. The **Q** vector relates to issues including reputation, technology, use, distribution and replacement. The term **q** can also be considered as a vector with elements related to variability in component manufacture, assembly, testing, storage, product transport and installation. The notion of little **q** can also be expressed as an efficiency, related to efforts in meeting **Q**. The issues of **q** are met when the product meets those systems that are used to realize quality **Q**. The maintenance of **Q** relies upon the ability of a business to understand and control the variability which might be associated with the process of product realization. The quality in a product is not directly connected to cost. Every single **Q** element

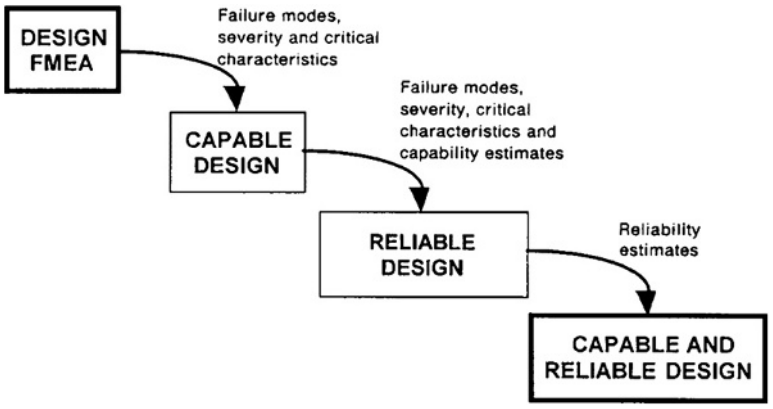


Fig. 7.1.19 The FMEA input into designing capable and reliable products.

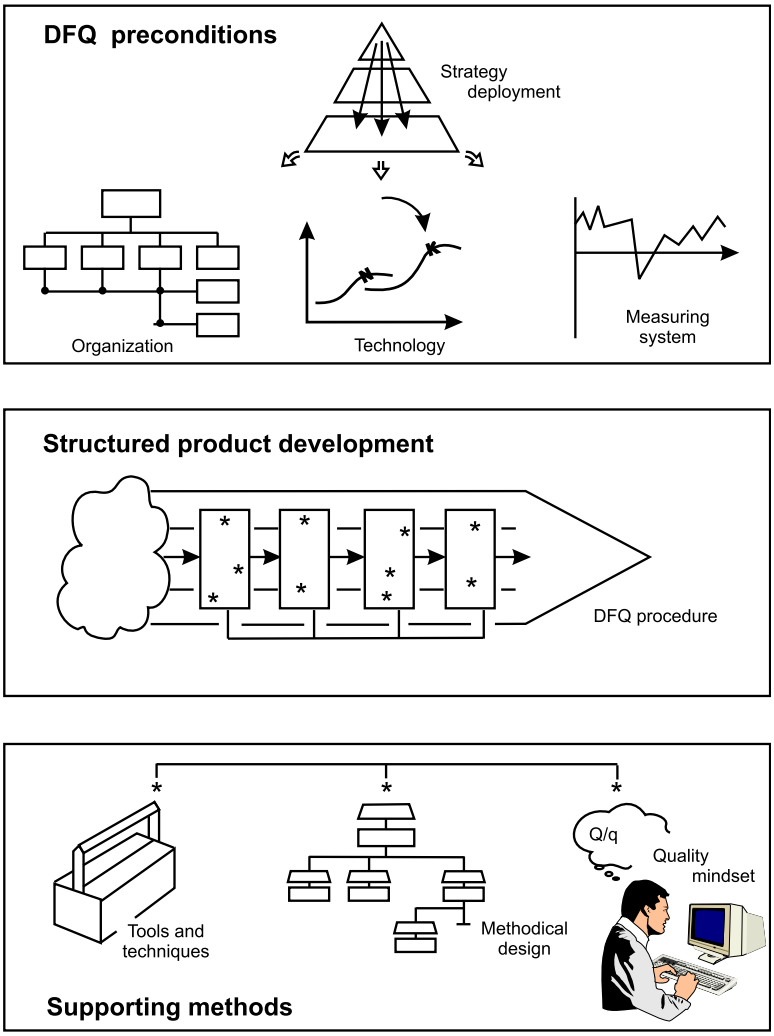


Fig. 7.1.20 Preconditions for and main elements of DFQ (adapted from Mørup, 1993).

$\mathbf{Q}_i$ has corresponding **q** elements that contribute to cost. **Q** is fundamentally connected to selling price.

As can be seen from the above, the DFQ and the **Q**/**q** concepts are extremely broad in perspective. The general model may be used to drive the considerations of the important issues throughout the stages of production development and in the design of individual components and assemblies. The **q** element of quality described by Morup is adopted in the CA methodology.

The link between customer wants/perceived quality, **Q**, and quality of conformance, **q**, is a vague area. A large number of problems created at the design/manufacture interface are also caused by technical quality problems, for example wrong material specification, wrong dimensions, etc. These are essentially design communication deficiencies and so are amenable to an appraisal by a methodology of sorts. The flow of information through these quality disciplines is shown in Figure 7.1.21 where an analogy is made to the design of a simple hole in a plate. Further investigation of technical quality, **Qq**, should be also performed if we are to gain a further understanding of DFQ. For example, the measuring and monitoring of design drawing errors using SPC attribute techniques has been a major step forward in reducing the design changes for an aerospace company.

It is also possible to categorize the different types of DFQ techniques that are required to analyse the several types of quality highlighted above. These are:

- Specification techniques which aid product developers in formulating quality objectives and specifications → **Q**.
- Synthesis techniques which aid the designer in generating ideas and in detailing solutions → **Qq**.
- Verification techniques which verify and evaluate the quality of solutions in relation to the specification → **q**.

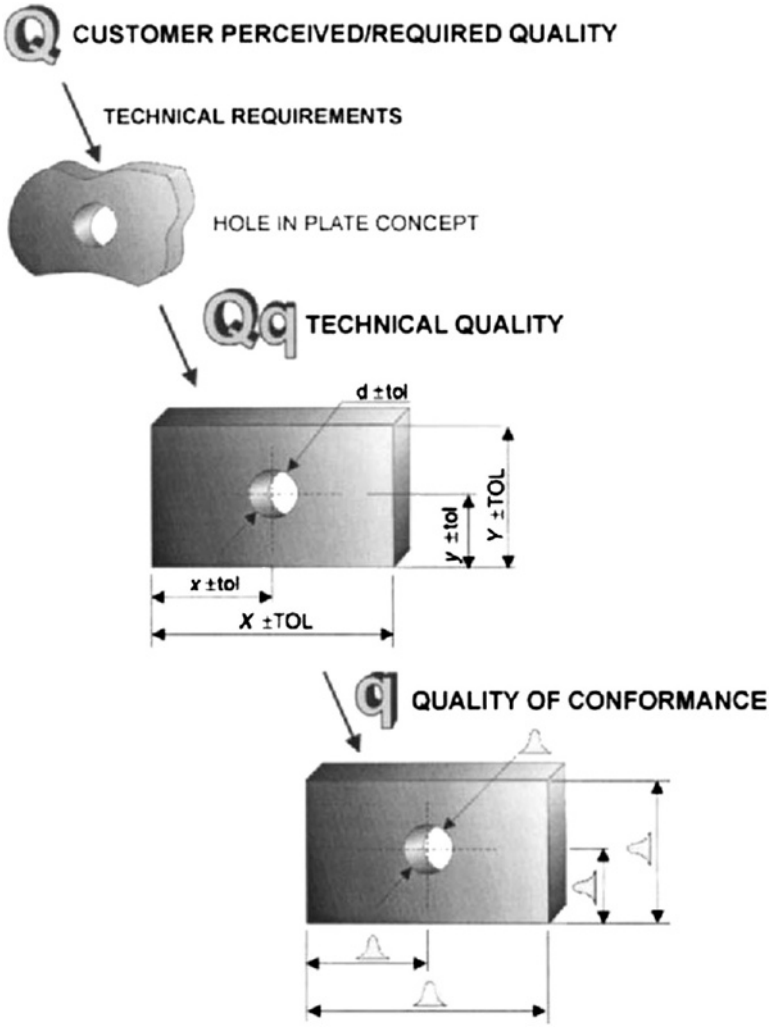


Fig. 7.1.21 Hole in plate analogy to quality and the **Q**/**q** concepts.

There is a need for verification techniques in DFQ that can be used in the early and critical product development phases, where the quality is determined, i.e. can be applied on abstract and incomplete product models. The CA methodology is largely a verification technique that aims to achieve this.

The DFQ approach at the verification level has many elements in common with Design for Manufacture (DFM) techniques. DFM helps create a product design that eases the task of manufacturing and lowers manufacturing cost. This is achieved by invoking a series of guidelines, principles and recommendations at the design stage and providing an understanding of the characteristics, capabilities and limitations of the manufacturing processes employed. These design rules, or 'producibility' guidelines, are more generally applied at the component level than the assembly level, although DFA is sometimes, rather confusingly, associated with DFM. Producibility guidelines are commonly developed by companies for designing products that are similar in nature to the ones for which the guidelines were written. They are therefore limited in their application because they may not apply to innovative design or where process capabilities are taken to the extreme limits by customer requirements.

Designing for robustness has also been associated with the DFM guidelines. Robust design has different meanings to different engineering communities. For example, the three descriptions below focus on three different but connected aspects of product design:

- Robust design creates performance characteristics that are very insensitive to variations in the manufacturing process, and other variations related to the environment and time.
- Robust design is the design of a product or process that results in functionally acceptable products within economic tolerances.

- Robust design improves product quality by reducing the effects of variability.

The first definition focuses on a process orientated design, the second the economic aspects of the design, and the third the impact of variability on the product in use. Although robust design is mentioned as a DFM guideline, no guidance for how to achieve 'robustness' is given. The definition of robust design must be made clear, and more importantly detailed guidance must be given to the designer on what to do to achieve robustness in a practical way. DFM techniques do not specifically answer this question. However, a robust design can be defined as a capable design in the context of the work presented later.

As stated earlier, awareness is growing that cost and quality are essentially designed into products (or not!) in the early stages of product engineering. The designer needs to know, or else needs to be able to, predict the capability of the process used to produce the design and to ensure the necessary tolerance limits are sufficiently wide to avoid manufacturing defects. Furthermore, the designer must consider the severity of potential failures and to make sure the design is sufficiently robust effectively to eliminate or accommodate defects. The major benefit of doing this is to reduce the potential for failure costs. Alternatively, in seeking to control or reduce cost, the safety of the product can be jeopardized, for example allowing production cost to dominate a design decision to the extent that the product does not meet customer expectation. The design phase of a product is crucial because it is here that the product's configuration and, therefore, much of its potential for harm are determined. Those engaged in product design must give a high priority to the elimination or control of hazards associated with the product. This may have to be to the detriment of ease of manufacture, styling, user convenience, price and other marketing factors.

The experiences of industrial collaborators and surveys of UK businesses suggest that failure costs are the main obstacle to reducing the costs of quality. There is, therefore, a need for design methods and guidelines to give businesses the foresight to identify product characteristics that depict potential costly failures, particularly in the case of bought out components and assemblies. Designers need models to predict costs at the various design stages. The provision of these techniques is essential for future business competitiveness. Through DFQ we can change over from merely preventing and eliminating quality problems to actively incorporating the level of quality expected by the consumer into the product. At the same time we can see that the target of high product quality is to a great extent compatible with the target of low costs, and thus with the creation of good business. We need a culture, professionalism and techniques for DFQ.

7.1.6 Designing for reliability

The pursuit of product quality over the last few decades has been intense in many companies around the world, but in many sectors of industry, reliability is considered to be the most important quality attribute of the product. As consumers become more aware than ever of quality, their expectations for reliability are also increasing. Even equipments without obvious safety concerns can have important reliability implications. Most products can, therefore, benefit from the use of sound reliability techniques. Reliability prediction, in turn, has the benefit that it gives a quantifiable estimate of the likely reliability that can be assessed to see if this is appropriate for the market.

Reliability prediction is undoubtedly an important aspect of the product design process, not only to quantify the design in terms of reliability, but also to determine the critical design parameters that go to produce a reliable product. To this end, it is necessary to have a mathematical, quantitative measure of reliability defined by probability. It is, however, a controversial aspect of reliability engineering in terms of accuracy and validity because it relies on detailed knowledge about sometimes unknown design parameters. In addition, the practical 'engineering' of reliability in the product is still seen as a science to many designers and evaluating concept designs for reliability is especially difficult for inexperienced staff. A fundamental reason for this is the supporting use of statistics and probability theory. Designers and engineers have consistently turned a blind eye to the advantages of using these methods; however, they can be trained to use them without being rigorously schooled in their mathematical foundation. Reliability prediction will remain a controversial technique until the statistical methods for quantifying design parameters become embedded in everyday engineering practice. Using statistically based design techniques will not solve the whole problem, although it will be much closer to the desired end result.

The reliability of a product is the measure of its ability to perform its intended function without failure for a specified time in a particular environment. Reliability engineering has developed into two principal areas: part and system. Part reliability is concerned with the failure characteristics of the individual part to make inferences about the part population. System reliability is concerned with the failure characteristics of a group of typically different parts assembled as a system. There are currently three main approaches used in the pursuit of a reliable product :

- Reliability prediction (includes probabilistic design)
- Design techniques (for example, FMEA or Fault Tree Analysis (FTA))
- Pre-production reliability testing (prototyping).

Reliability prediction techniques are a controversial but effective approach to designing for reliability, as discussed later. FTA and FMEA are established techniques that aim to determine the potential causes and effects of failures in components and systems. Although subjective in nature, they are useful at the system level where many interactions of components take place. Finally, prototype testing is a pre-production exercise performed on a few model products tested to determine if they meet the specification requirements. All three approaches have their value in the product development process, although the aim of the first two must surely be to reduce the third, prototype testing, which can be an expensive undertaking. However, you could never remove the necessity for some kind of prototype testing, especially where the aim is to verify the integration of parts and subassemblies. From the above, it would seem that designing for reliability depends on both quantitative results and qualitative processes. However, the collection of reliability tasks might look intimidating causing one to wonder if the benefit gained by performing all is worth the effort.

Virtually all design parameters such as dimensional characteristics, material properties and service loads exhibit some statistical variability and uncertainty that influence the adequacy of the design. Variability may arise from material quality, adverse part geometry and environmental factors. Historically, the designer has catered for these variabilities by using large factors of safety in a deterministic design approach. In probabilistic design, statistical methods are used to investigate the combination and interaction of these parameters, having characterized distributions, to estimate the probability of failure. A key requirement is detailed knowledge about the distributions involved to enable plausible results to be produced. The amount of information available at these early stages is limited, and the designer makes experienced judgements where information is lacking. This is why the deterministic approach is still popular, because many of the variables are taken under the 'umbrella' of one simple factor. If knowledge of the critical variables in the design can be estimated within a certain confidence level, then the probabilistic approach becomes more suitable. It is essential to quantify the reliability and safety of engineering components and probabilistic analyses must be performed.

There are numerous applications for probabilistic design techniques in mechanical engineering. A number of important applications exist in design optimization and reliability engineering, specifically where it would be useful to explore the level of random failure, resulting from the interaction of the distributions of loading stress and material strength. The approach is shown in Figure 7.1.22 again for a simple hole in a plate concept. All the design parameters are shown as statistical distributions rather than unique or nominal values. The final failure prediction through an appropriate probabilistic and failure analysis model reflects the distributional nature of the design parameters.

More plausible representations of stress and strength distributions for a given situation will enable meaningful failure predictions to be produced, and will be particularly useful where test to failure is not a practical proposition, where weight minimization and/or material cost reduction is important, or where development time is crucial. Engineering experience indicates that many devices are overdesigned, that is, they feature excess weight or excess occupied space. When weight and/or space is at a premium, a more realistic design process is required that permits relatively accurate predictions of device performance. A probabilistic design process, taking into account of the uncertainties with typical design inputs, provides the required realism.

In a probabilistic approach, design decisions must reduce the probability of unwanted performance to acceptable levels. In a deterministic approach, the designer can only assure that the performance remains within an acceptable domain. As the deterministic approach provides no firm basis for dealing with variability, it is not pertinent to a reliability approach. As an example of this, Haugen shows that the failure probability for a particular problem can vary from zero to unity because of the variability of the design parameters, but with the factor of safety selected remaining constant. Factors of safety can lead to either an unconservative design with unacceptable high failure rates, or a very conservative design that provides the required performance with unnecessarily high product costs.

Deterministic design is still appealing because of its simplicity in form and application, but since factors of safety are not performance related measures, there is no way of indicating if the design is near optimum. With increasing concern over minimizing the cost of failure, the probabilistic design approach will become more important. Probabilistic design gives the designer a better feel of just how conservative or unconservative the design is. In order to determine this, however, it is important to make decisions about the target reliability level.

To be able to evaluate design reliability estimates using probabilistic methods, the designer needs much more information than for a deterministic evaluation. It can be argued that probabilistic design can be used only when all the needed statistical data are available and it would be dangerous to design to a reliability target when the data are suspect. Because of the lack of statistical data for the strength of materials used and the applied loads in particular, design concepts based on the factor of safety will still dominate the design of some products. However, the probabilistic approach allows us to perform a sensitivity analysis of the design with respect to the various design

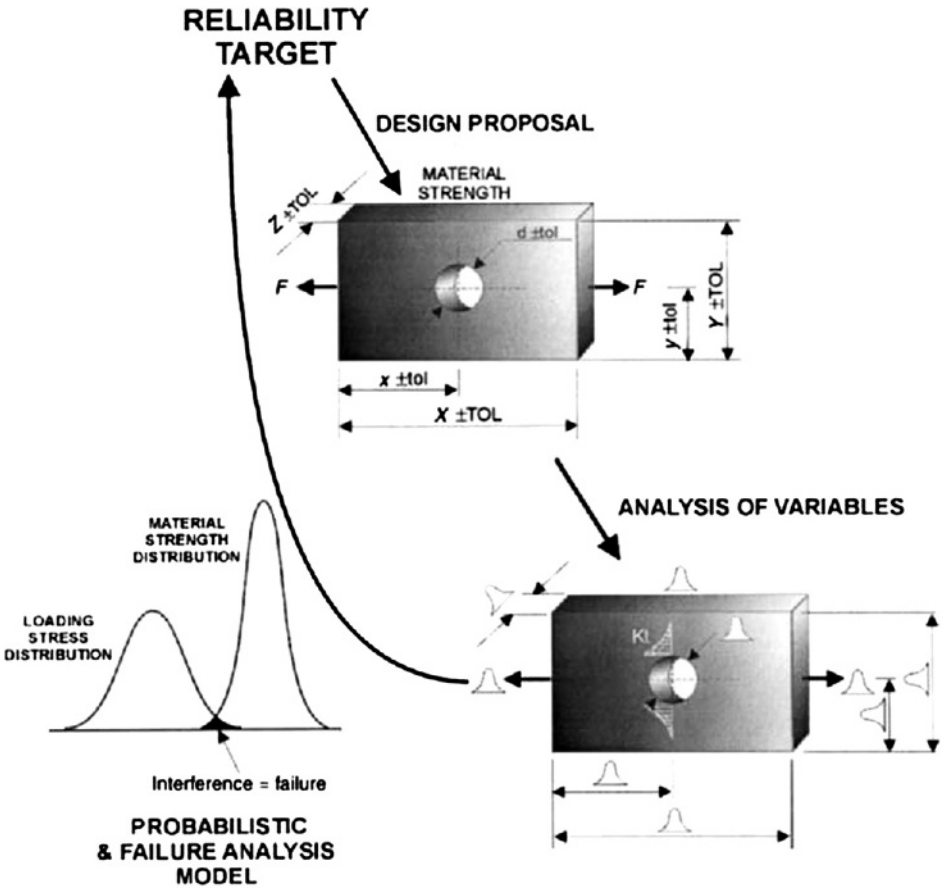


Fig. 7.1.22 Hole in a plate analogy of probabilistic design.

parameters to give an idea of the impact of the variability of dimensions, material strength and loads on performance, and this makes design optimization possible. Probabilistic design is another way of thinking about the design problem which must surely be an improvement over using large factors of safety.

Probabilistic methods have gained increased interest in engineering as judged from the growing community of reliability engineers and from the increasing number of conferences on the subject. Some practitioners in the UK, however, either seem to lose confidence with statistical and probabilistic methods or are just not aware of them. At present, only larger companies seem to be aware of their importance. Some advocates of a statistical approach to engineering design even claim that this is why large chunks of manufacturing have moved to countries like Japan who embrace the use of such techniques. A comment in 1995 by Margetson gives an indication of the situation related to the UK:

> *It is essential to introduce probabilistic design methods into engineering design procedures. I feel that the UK will be faced with a severe skill gap. I also feel there is a lack of appreciation of the need and time scale to introduce the required procedures to engineers … the Japanese have identified probabilistic design as a key technical area.*

A problem may lie in the knowledge that is required as an essential input to such approaches, being both statistical in nature and from the authors' own experiences, often difficult to obtain and interpret generally. Unfortunately, experienced designers will not use statistical methodology, although statistical methods should play an important role in the design and manufacture of reliable products. A principal drawback of the probabilistic design approach is that it requires a good knowledge of probability and statistics, and not every design engineer has this knowledge. Summarizing the above, the main characteristics of the deterministic and probabilistic design approaches are shown in Table 7.1.1.

The provision for reliability must be made during the earliest design concept stage. The more problems prevented early on through careful design, the fewer

Table 7.1.1 Competing issues in deterministic and probabilistic design approaches

Deterministic design	Probabilistic design
Dominated design for 150 years	Application for approximately 40 years
Design parameters treated as unique values	Design parameters treated as random variables
Underlying empirical and subjective nature	Small samples used to obtain statistical distributions
Factor of safety dominates determination of reliability	
Ignorance about the problem being multi-variable	An understanding of statistics is required
Calculations simple and data widely available	Knowledge scattered throughout engineering texts
Inherent overdesign – wasteful, costly and ineffective	Output as a probability of failure or reliability
Culture of deterministic design still exists in industry	Better understanding of the effects of variability

problems that have to be corrected later through a time-consuming and often confusing process of prototype. A principal necessity then is to design to a reliability goal without an inordinate amount of component and prototype testing. This can only be achieved by a rigorous appraisal of the design as honestly and early as possible in the product development process.

7.1.7 Summary

The designer's job is to try to capture customer expectations and translate as many of these expectations as possible to the final product. The functional requirements of the design become detailed into dimensional tolerances or into attributes of the component or assembly. The ability of the manufacturing process, by which these products are made or assembled, to consistently provide dimensions within tolerance may be called its conformance to design. Understanding and controlling the variability associated with these design attributes then becomes a key element of developing a quality product.

Designers rarely fully understand the manufacturing systems that they are designing for, and subsequently they do not understand the variability associated with the design characteristics. Variability can have severe repercussions in terms of failure costs, appearing in production due to rework and scrap, and warranty costs (or worse!) when the product fails in service. There is need to try to anticipate the variability associated with the manufacturing processes used to produce the final product early in the design process. The designer needs to know, or else be able to predict the capability of the process and to ensure the necessary tolerance limits are sufficiently wide to avoid manufacturing defects. However, this has previously been difficult to achieve on concept design or where little detail exists.

Quality assurance systems demand not just tolerances, but process capable tolerances and characteristics that limit the potential failure costs incurred. The level of process capability is not necessarily mandated, but clearly the more severe the consequences of a defect, the lower the probability of occurrence of that defect should be. This 'risk' may be seen to bear a failure cost penalty and the expected cost must be limited. It should be clear that techniques such as SPC can only be applied to enhance the intrinsic capability of a process. Further, any inspection process to remove defective items can be only a compensation for a process with inadequate capability, or perhaps a design which has been specified with tolerances which are just too tight, or assemblies which are not quite practical.

From the above discussion, it follows that the quality and conformance to tolerance of the product characteristics should be 'designed in' and not left to the process engineer and quality engineer to increase to the required level. In order to do this, designers need to be aware of potential problems and shortfalls in the capability of their designs. They therefore need a technique which estimates process capability and quantifies design risks.

Reliability prediction techniques are a controversial but fundamental approach for designing for reliability. A key objective of these methods is to provide the designer with a deeper understanding of the critical design parameters and how they influence the adequacy of the design in its operating environment. These variables include dimensions, material properties and in-service loading. A key requirement is detailed knowledge about the distributions involved to enable plausible results to be produced in an analysis. It is largely the appropriateness and validity of this input information (and failure theory) that determines the degree of realism of the design process and the ability to accurately predict the behaviour and therefore the success of the design.

The determination of an absolute value of reliability is impossible. In developing a product, a number of design schemes or alternatives should be generated to explore each for their ability to meet the target requirements. Evaluating and comparing designs and choosing the one with the greatest predicted reliability, or quality for that matter, will provide the most effective design solution.

The costs of quality experienced by manufacturing businesses are dominated by costs associated with failure, for example rework, scrap, design changes, warranty and product liability claims. These costs represent lost profit and potentially impact on the future opportunities of the business. Decisions made during the design stage of the product development process account for a large proportion of the problems that incur failure costs in production and service. It is possible to relate these failure costs back to the original design intent where variability, and the lack of understanding of variability, is a key failure costs driver. Current quality costing models are useful for identifying general trends in a long-term improvement programme, but are of limited use in the identification of the failure costs associated with actual design decisions. An important aspect of the discussion in the next section is to demonstrate a method whereby the estimation of failure costs at the design stage is useful in selecting the most cost effective, from a failure point of view, from a number of alternatives.

In Chapter 7.2, we introduce the concepts of component manufacturing capability and the relationships among tolerance, variability and cost. The Component Manufacturing Variability Risks Analysis is then introduced, the first stage of the CA methodology, from which process capability estimates can be determined at the design stage. The development of the knowledge and indices used in the analysis is discussed within the concept of an 'ideal design'. The need for assembly variability determination and the inadequacy of the DFA techniques in this respect is argued, followed by an introduction to assembly sequence diagrams and their use in facilitating an assembly analysis. The Component Assembly Variability Risks Analysis is discussed next, which is the second stage of the CA methodology. Finally, a cost–quality model is presented, used to determine the failure costs associated with non-safety critical and safety critical applications in production and service through linkage with FMEA. Comprehensive guidance on the application of the technique is given and a number of case studies and example analyses are used to illustrate the benefits of the approach.

This page is intentionally left blank

Chapter 7.2

7.2

Designing capable components and assemblies

J.D. Booker, M. Raines and K.G. Swift

7.2.1 Manufacturing capability

One of the basic expectations of the customer is conformance to specification, that is, the customer expects output characteristics to be on target with minimum variation. Assessing the capability of designs early in product development therefore becomes crucial. The designer must aim to achieve the standards demanded by the specification, but at the same time not exceed the capabilities of the production department. This may not be an easy task because the determinants of quality are frequently difficult to identify at the design stage. The designer must decide how the specifications are affected by features of the production process, as well as by the choice of materials, design scheme, etc. If the designer knows that a certain manufacturing process can only achieve very poor dimensional tolerances, for example, a decision must be made to decide whether such an process can be used for the new product. If not, an alternative solution must be devised.

The available information about manufacturing processes is both extensive and diverse, but it is not easily accessible for designers, as the structure is not suited for design applications. The designer's job is made easier if standard rules are available for the candidate manufacturing method. A good set of design rules can indicate process capable tolerances for a design. Life is much more problematic when rules are not available or an unfamiliar manufacturing process is to be used. Many designers have practical experience of production methods and fully understand the limitations and capabilities that they must work within. Unfortunately, there are also many that do not have this experience and, quite simply, do not appreciate the systems that they are supposed to be designing for. For example, three aspects of engineering drawing that are often overlooked, but are vital in design practice, are dimensions, tolerances and specification of surface finish. The designer needs to understand when required tolerances are pushing the process to the limit and to specify where capability should be measured and validated.

Tolerances alone simply do not contain enough information for the efficient manufacture of a design concept. At the design stage, both qualitative manufacturing knowledge about candidate manufacturing processes for design features, and quantitative manufacturing knowledge on the influences of design parameters on production costs, are needed. Only with strong efforts to integrate the product design with the selection of the manufacturing process can the desirable performance characteristics be produced with a minimum of variability and minimum cost.

The goal of this first stage of Conformability Analysis (CA), the component manufacturing variability risks analysis, is the provision of support in the early stages of the detailed design process for assessing the tolerance and surface finish capability associated with component manufacture. 'Risk' in this sense is a measure of the chances of not meeting the specification. Recognizing that the relationship between a design and its production capability is complex and not easily amenable to precise scientific formulation, the methodology described has resulted largely from knowledge engineering exercises in manufacturing businesses, including those with expertise in particular manufacturing processes. For example, Poeton in Gloucester, UK, specialize in supplying surface coating technology in the form of design guidance and processing techniques. The Poeton engineers had a major

Designing Capable and Reliable Products; ISBN: 9780750650762
Copyright © 2001 Elsevier Ltd. All rights of reproduction, in any form, reserved.

consultation role in the determination of the main issues related to surface engineering process variability. To support other aspects of the analysis, such as the tolerance capability analysis, manufacturing tolerance data were readily available, albeit in a form not useful to designers. During the evolution of CA, many alterations and improvements were made through exhaustive consultation and validation in industry. The development of only the current aspects of the analysis will be discussed in detail here. However, some background topics will be reviewed before proceeding.

7.2.1.1 Variability factors in manufacturing

Designers have always had to deal with the fact that parts cannot be made perfectly; or if they could, they would not remain perfect for long during use. So defining the 'ideal' component is only one aspect of the designer's job. The designer must also decide by how much a still acceptable component can be from the ideal. A component can vary from the ideal in many ways: in its geometry, its material properties, surface finish – a virtually unlimited list.

In analysing a design at the concept development or early detailing stage, it is only necessary to focus on the main variabilities associated with manufacturing processes. In this way the analysis performs at a level of abstraction which facilitates a rapid assessment without being complex or difficult to comprehend. Engineering judgement and experience are needed to identify potential variability risks or 'noise' associated with manufacturing and assembly routes. Before introducing the main aspects of the component manufacturing variability risks analysis, it is worth exploring variability issues in general. There are two main kinds of variability:

- Common-cause or inherent variability is due to the set of factors that are inherent in a machine/process by virtue of its design, construction and the nature of its operation, for example positional repeatability, machine rigidity, which cannot be removed without undue expense and/or process redesign. When only common-cause variability is present, the process is performing at its best possible level under the current process design.
- Assignable-cause or special-cause variability is due to identifiable sources which can be systematically identified and eliminated (see below).

The last sources of variation highlighted are of paramount concern for allocating and analysing mechanical tolerances.

At the most detailed level, a variation can belong to the basic design properties: form, dimension, material and surface quality for the components, and structure for the relations between components. For example, the levels of dimensional tolerance accuracy and surface roughness associated with industrial manufacturing processes vary widely. In general, tolerances reflect the accuracy specifications of the design requirements. These will inherently reflect the variability in the manufacturing process. That is the design of the tolerance for a certain component characteristic will influence the variability in the measurements of that characteristic. Hence, a tightened tolerance requires higher precision manufacturing devices, higher technical skills, higher operation attention and increased manufacturing steps.

The general factors influencing variation include the following:

- Tool and functional accuracy
- Operator
- Set-up errors
- Deformation – due to mechanical and thermal effects
- Measurement errors
- Material impurities
- Specifications
- Equipment
- Method or job instructions
- Environment

Further variations may arise from the working of the material during the manufacturing process or from deliberate or unavoidable heat treatment. In general, the undesirable and sometimes uncontrollable factors that cause a functional characteristic to deviate from its target value are often called noise factors and are defined below:

- *Outer noise* – environmental conditions such as temperature, humidity and different customer usage.
- *Inner noise* – changes in the inherent properties of the system such as deterioration, wear and corrosion.
- *Product noise* – piece to piece variation due to manufacturing variation and imperfections.

An example of assignable-cause variability is given in Figure 7.2.1. Two cases of milling the same component to finished size are shown. In the first case, the component is relocated a number of times in the tooling to produce the finished product. The tolerance capability of a key dimension is analysed and the results are given in the form of a histogram. In the second case, the number of relocations is reduced and simultaneous operations are carried out on the component in the form of gang milling. A significant capability improvement is experienced in the second case as shown by the comparatively low process spread and accumulation around the target value. This is mainly due to the set-up and operation procedures for the manufacture of the component. It is

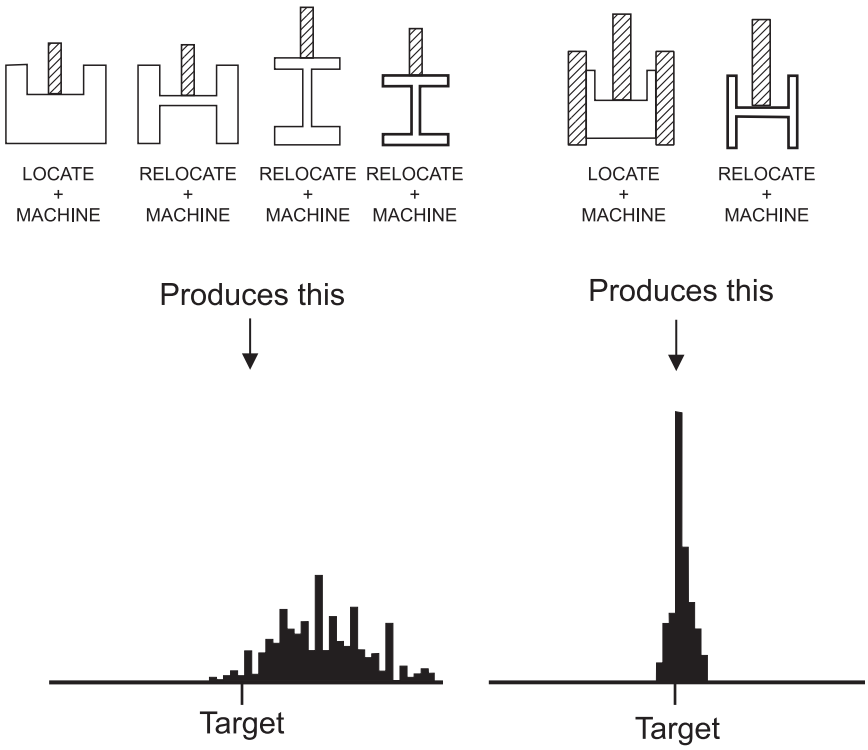


Fig. 7.2.1 Example of process capability improvement. (adapted from Leaney, 1996a).

evident from this that a good process set-up ensures better product quality.

7.2.1.2 Cost–tolerance relationships

Ideally, engineers like tight tolerances to assure fit and function of their design. Designers often specify unnecessary tight tolerances due to the lack of appreciation of cost and due to the lack of confidence in manufacturing component parts which conform to specification. However, a tightened tolerance requires higher precision and more expensive manufacturing processes, higher technical skills, higher operator attention, increased manufacturing steps and is more time consuming to achieve. The configuration and material of a part, as well dimensions, tolerances and surface finishes, can also change the amount of work required in part manufacture.

All manufacturers, on the other hand, prefer loose tolerances which make parts easier and less expensive to produce. The choice of tolerance is therefore not only related to functional requirements, but also to the manufacturing cost. It has been argued that among the effects of design specifications on costs, those of tolerances are perhaps the most significant. Figure 7.2.2(a) shows the relationship between product tolerance and approximate relative cost for several manufacturing processes. Each tolerance assignment results in distinct manufacturing costs.

Cost–tolerance functions are used to describe the manufacturing cost associated with a process in achieving a desired level of tolerance. The shape of the cost–tolerance curve has been suggested by several researchers. Cagan and Kurfess propose a hyperbolic cost function. Speckhart in 1972 and Dong in 1997 suggest an exponential cost function. The inverse quadratic form was advocated by Spotts in 1973. A comparison of various researchers' cost models is provided in Figure 7.2.2(b). Determining the parameters of the cost–tolerance models described, however, is by no means a trivial task.

7.2.1.3 Process capability and tolerances

Product tolerance can significantly influence product variability. Unfortunately, we have difficulty in finding the exact relationship between them. An approximate relationship can be found from the process capability index, a quality metric interrelated to manufacturing cost and tolerance. The random manner by which the inherent inaccuracies within a manufacturing process are generated produces a pattern of variation for the dimension which resembles the normal distribution and

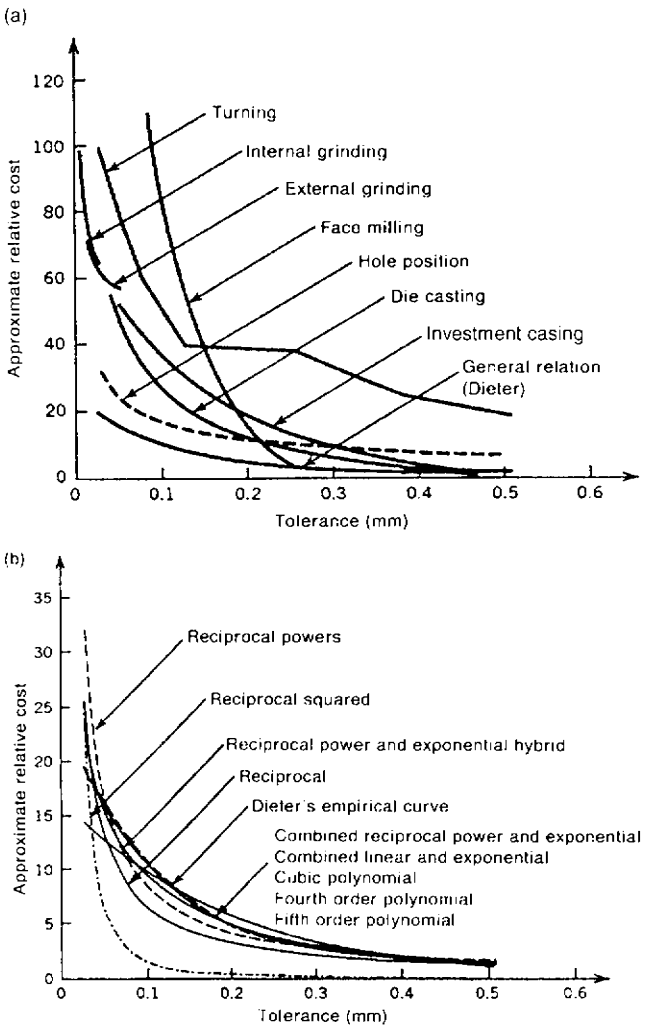


Fig. 7.2.2 (a) Tolerance versus production costs of various processes; (b) comparison of cost–tolerance models (Dong, 1993).

therefore process capability indices, which are based on the normal distribution, are suitable for use.

There are many variations on the two basic process capability indices, C_p and C_{pk}, these being the most commonly used. A process capability index for a shifted distribution at $C_{pk} = 1.33$ is still regarded to be the absolute minimum (which relates to around 30 ppm failing the specification). This increases to $C_{pk} = 1.66$ (which relates to approximately 0.3 ppm failing) for more safety critical characteristics. Motorola stipulate $C_{pk} = 1.5$ (or approximately 3 ppm) to their suppliers as an absolute minimum. However, one weakness of the process capability index is that there is no apparent basis for specifying the optimal value. In practice, most designers do not worry about the true behaviour of a manufacturing process and compensate for the lack of knowledge with larger than necessary process capability indices.

The precise capability of a process cannot be determined before statistical control of the actual process has been established. Designers seldom have sufficient data by which to specify the variability of the manufacturing processes and therefore during the design phase, the designer must use the best available process capability data for similar processes. It is far easier, not to mention less costly, to create robust designs based on known process capabilities than it is to track down and subsequently reduce sources of variation during the manufacturing phase.

7.2.2 Component manufacturing variability risks analysis

In the development of the Component Manufacturing Variability Risk Index, q_m, it was found to be helpful to consider a number of design/manufacture interface issues, including:

- Material to process compatibility
- Component geometry to process limitations
- Process precision and tolerance capability
- Surface roughness and detail capability.

In the formulation of q_m, it has been assumed that there is a basic level of variability associated with an 'ideal' design for a specific manufacturing process, and that the factors listed above can be estimated independently and manipulated according to some function:

$$q_m = f(m_p, g_p, t_p, s_p) \tag{7.2.1}$$

where q_m is the component manufacturing variability risk; m_p the material to process risk; g_p the geometry to process risk; t_p the tolerance to process risk; s_p the surface roughness to process risk. It then becomes possible to define the specific variability risk indices relating to Equation (7.2.1) through the use of a number of charts. Reference to the guidance notes associated with each chart is recommended.

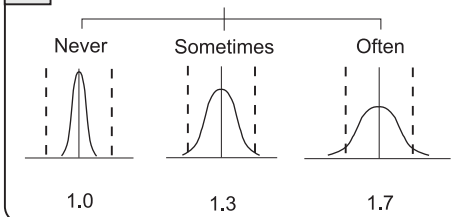
The material to process risk, m_p, uses manufacturing knowledge in the assessment of the component material to process compatibility, for example machinability rating or formability rating (see Figure 7.2.3). These values are greater than unity only when the material to be used in conjunction with the manufacturing process is not defined on the process capability maps used. A thorough discussion of process capability maps is given in Section 7.2.2.1.

The geometry to process risk, g_p, uses knowledge of known geometries that exhibit variability during component manufacture to support the risks analysis, for example parting lines on castings and long unsupported sections (see Figure 7.2.4).

START

Initially, knowledge of the chosen manufacturing process is required. If the type of material is stated on the process capability map, assume m_p=1, otherwise proceed to 'A'.

A Do material properties vary? (1)



= mean material property affecting processing

B Values of 'B' relating to material compatibility for those processes where a material is not defined on the process capability maps given (2)

	MATERIAL \ PROCESS	IMPACT EXTRUSION COLD FORMING COLD EXTRUSION	SHEET METALWORK: CUTTING BLANKING FINE BLANKING	SHEET METALWORK: BENDING/DRAWING ROLL FORMING SPINNING	MACHINING	POWDER METAL SINTERING
STEELS	FREE CUTTING				1.2	
	MILD	1.2	1.4	1.2	1.3	1.1
	MED. CARBON	1.4	1.5	1.4	1.4	1.1
	HIGH CARBON	1.7	1.7	1.5	1.5	1.1
	ALLOY	1.7	1.5	1.5	1.5	1.2
	TOOL				1.7	1.3
	CAST				1.4	
	STAINLESS	1.7	1.5	1.4	1.6	1.1
IRONS	MALLEABLE				1.1	1.0
	NODULAR				1.2	
	PEARLITIC				1.5	
	GREY				1.4	
NONFERROUS ALLOYS	ALUMINIUM	1.0	1.0	1.0	1.0	1.5
	COPPER	1.0	1.2	1.2	1.2	1.7
	MAGNESIUM	1.1	1.3	1.7	1.1	
	NICKEL	1.5	1.4	1.4	1.6	1.4
	ZINC	1.0	1.1	1.5	1.1	
	TIN	1.0	1.2	1.7	1.5	
	LEAD	1.0	1.2	1.7	1.4	
	TITANIUM		1.5	1.5	1.7	1.4
	PLASTICS				1.5	

Notes

① Variations in material properties create out of tolerance variations during processing. Variations can relate to uncertainty in material composition, flaws, cracks, discontinuations, as well as non-uniform processing characteristics.

② There is a material to process compatibility risk for impact extrusion, cold forging, cold extrusion, sheet metalworking, machining and powder metal sintering processes because their respective process capability maps relate to the ideal material case.

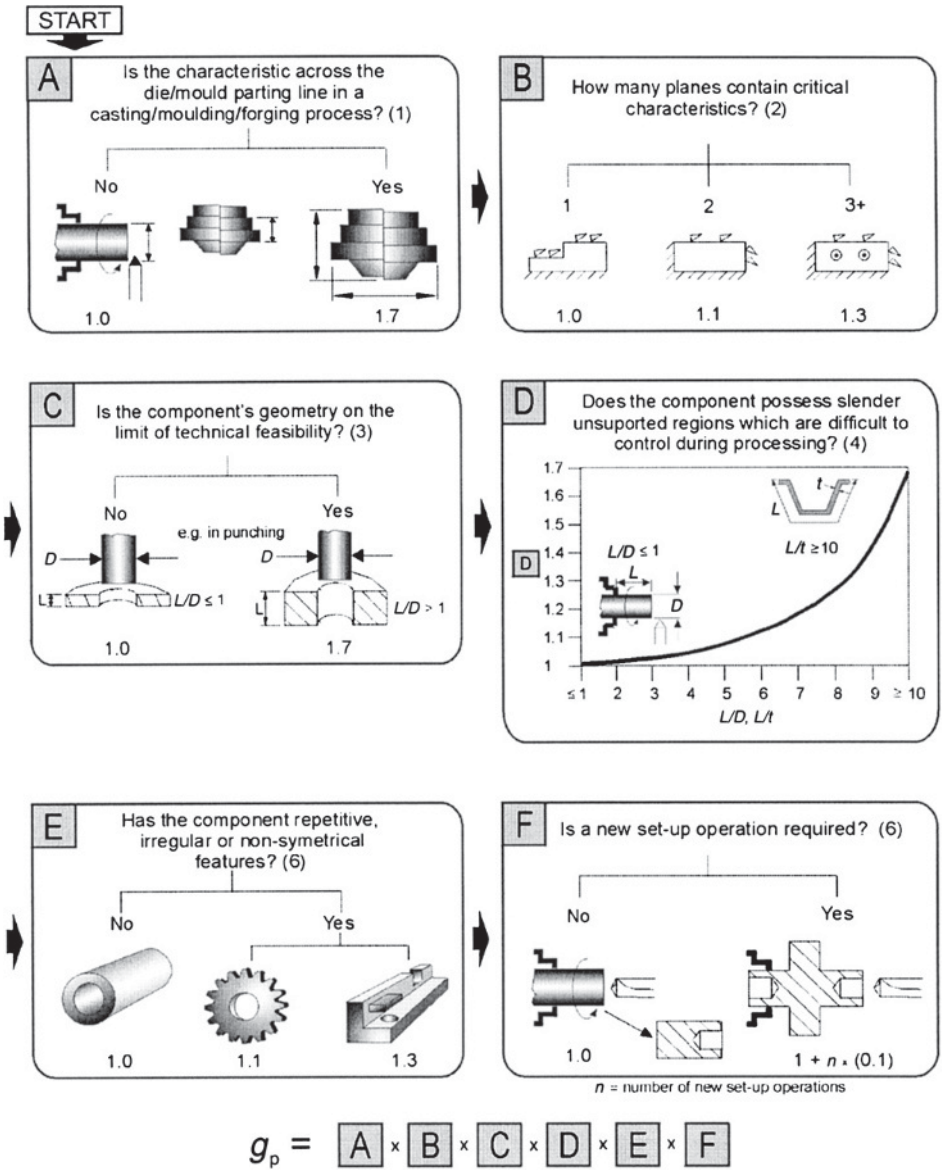
$$m_p = A \times B$$

Fig. 7.2.3 Material to process risk chart, m_p.

The tolerance to process risk chart for t_p (see Figure 7.2.5) uses knowledge of the dimensional tolerance capability for a number manufacturing processes and material combinations in the form of process capability maps (Figure 7.2.6).

Related to tolerance, but distinct from the tolerance to process risk, the surface roughness to process risk, s_p, is a function of the surface roughness specification. Figure 7.2.7 shows the chart used initially. Again, knowledge of the surface roughness values for a number of manufacturing processes is used to facilitate an analysis as provided in Figure 7.2.8. A discussion of the generation procedure for the surface roughness chart is given in Section 7.2.2.2.

Validation studies in manufacturing business and discussion with experts led to the view that knowledge used



Notes

① Allowances should be made for tolerances across the parting line. Allowance may also be required for mismatch of the dies/moulds in some casting and forging processes. Flash thickness allowances may also be required in closed die forging. Note, this only applies to casting, moulding and forging processes.
② Refers to the number of orthogonal axes on which the critical characteristics lie, and which cannot be achieved by processing from a single direction.
③ The processing of components that are on the limits of technical feasibility is likely to result in out of tolerance variation. High forces and flow restriction in metalworking and metal cutting processes can lead to instability. Also, material flow in casting processes, where abnormal sections and complex geometries are present, can lead to variability problems and defects.
④ Slender unsupported regions with large length to thickness ratios are highly liable to distortion during processing.
⑤ Repetitive, irregular or non-symmetrical features require greater process control and complex set-up or tooling requirements. This can be an added source of variability.
⑥ There is all increased risk of variation each time a new set-up operation is required due to changes in the orientation of the part or tooling.

Fig. 7.2.4 Geometry to process risk chart, g_p.

to define m_p, g_p, t_p and s_p, could be structured such that q_m may be formulated as

$$q_m = t_p \times s_p \tag{7.2.2}$$

where

$$t_p = f(\text{design tolerance}, m_p, g_p) \tag{7.2.3}$$

(See Figure 7.2.5 for the complete formulation of t_p.)

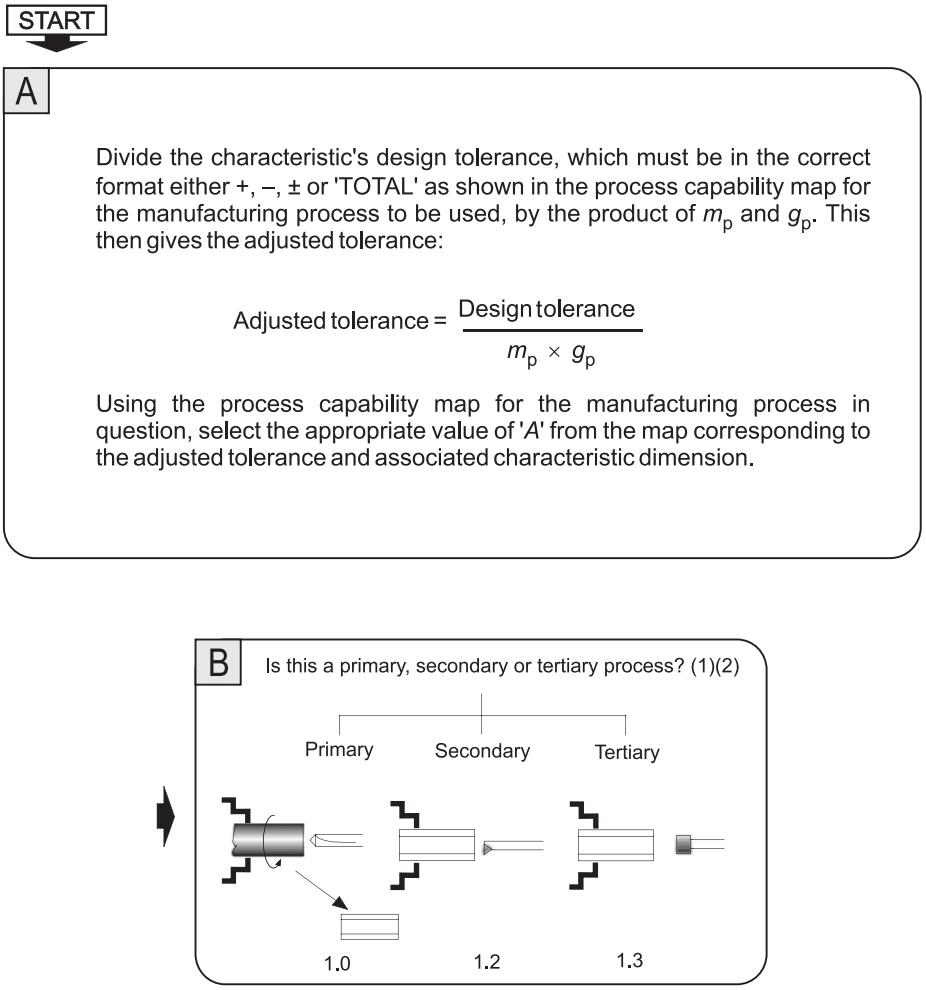


Fig. 7.2.5 Tolerance to process risk chart, t_p.

A link between the material used and the geometry of the component is compounded in the formulation for the tolerance to process risk as shown in Equation (7.2.3). It is recognized that increasing material incompatibility and geometry complexity has the effect of increasing the variability associated with achieving the dimensional tolerance requirement. The above equation relates this notion to the tolerance process risk by dividing the design tolerance for the characteristic, as stated by the designer, by the product of m_p and g_p. The risk index for surface roughness, s_p, usually defaults to unity unless a surface roughness requirement is considered critical, for example a valve seat or lubricated surface.

The underlying notion of the Component Manufacturing Variability Risk Index, q_m, is that an ideal design exists for a component where the risk index is unity, indicating that variability is in control. Risk indices greater than unity exhibit a greater potential for variability during manufacture. The resulting value of q_m indicates the risk of out of tolerance/surface roughness capability when compared to an ideal situation. For the

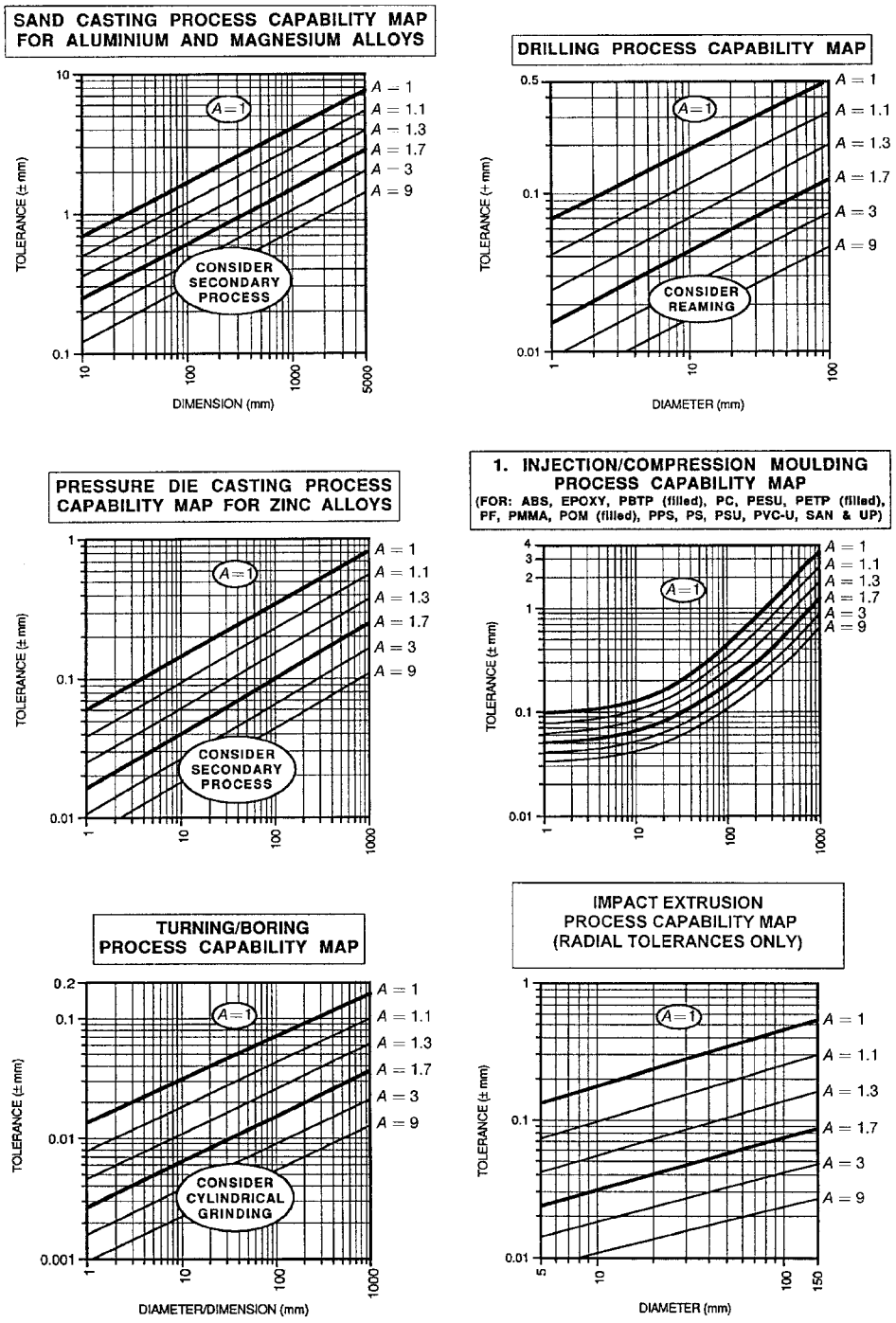


Fig. 7.2.6 A sample set of process capability maps.

ideal design of a component and processing route, each of the quantities is unity and therefore in all cases $q_m \geq 1$. Tolerance, surface roughness, material and geometry designed into a component, which is not matched with the ideal, have an effect on variability. For example, in die casting there is a higher risk when processing copper alloys compared with the tolerance capability resulting from processing zinc-based materials, largely due to temperature effects.

Additionally, there are other manufacturing processes (for example, tempering and nitriding) that must be considered in the analysis if used in the product's manufacturing route. These processes are carried out after the primary/secondary processes have

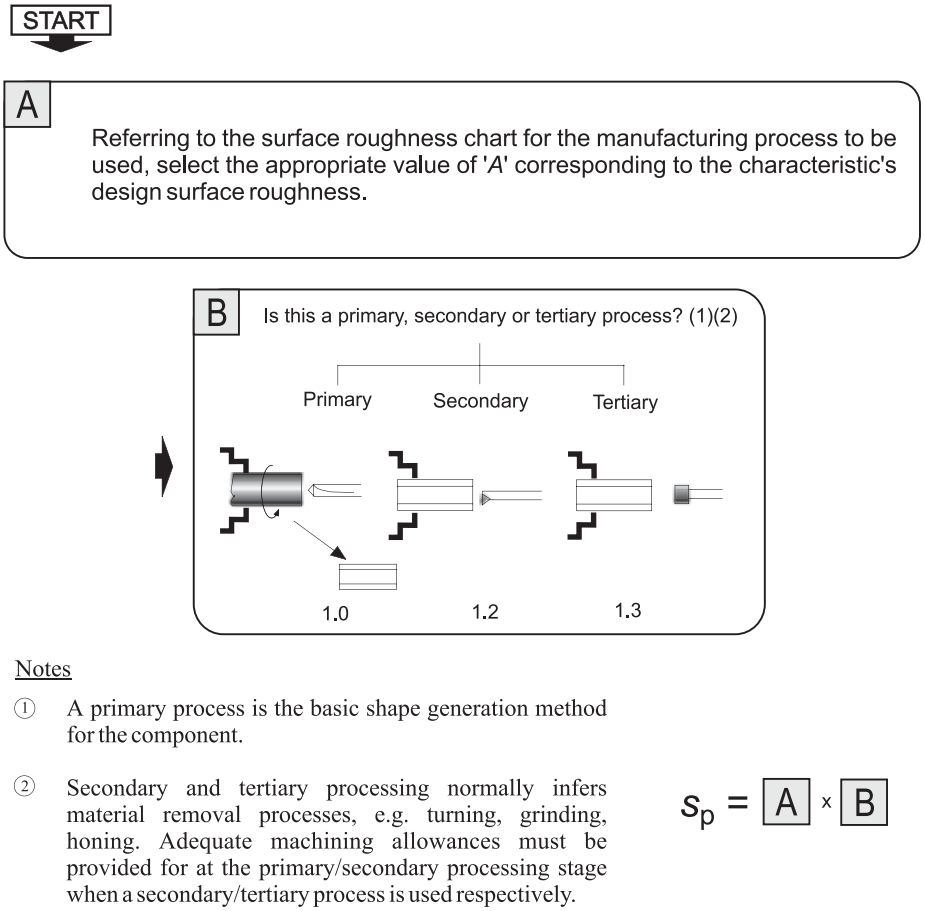


Fig. 7.2.7 Surface roughness to process risk chart, s_p.

been used to manufacture the component and are treated as post-manufacturing processes. The potential for variability in the final component due to these processes is great, due to the possible combination of high temperatures and unsymmetrical sections, which are particularly likely to cause out tolerance variations. This introduces an additional factor to consider, based on the following.

Surface engineering processes (bulk and surface heat treatment/coating processes)

Surface engineering processes are usually performed after the primary shape generation of the component, or post-manufacturing, therefore q_m defaults to k_p when it is considered in the manufacturing route:

$$q_m = k_p \tag{7.2.4}$$

where k_p is the surface engineering process risk.

Figure 7.2.9 shows the surface engineering process risk chart, k_p. It includes the key variability issues related to these types of processes.

Validation of the predictions for process capability through the use of the component manufacturing variability risks analysis, q_m, is given later.

7.2.2.1 Process capability maps

As can be seen from the above, central to the determination of q_m is the use of the process capability maps which show the relationship between the achievable tolerance and the characteristic dimension for a number of manufacturing processes and material combinations. Figure 7.2.6 shows a selection of process capability maps used in the component manufacturing variability risks analysis and developed as part of the research. There are currently over 60 maps incorporated within the analysis covering processes from casting to honing.

Data on the tolerance capability of the manufacturing processes covered were compiled from international standards, knowledge engineering in specialist businesses and engineering texts. The data used in the creation of the maps usually come in the form of tables, such as that given in Figure 7.2.10 for machining using turning and boring. International Organization for Standards (ISO) tolerance grades are commonly used as a straightforward way of representing the tolerance capability of a manufacturing process. The lower the tolerance grade, the more difficult the attainment using the particular manufacturing process.

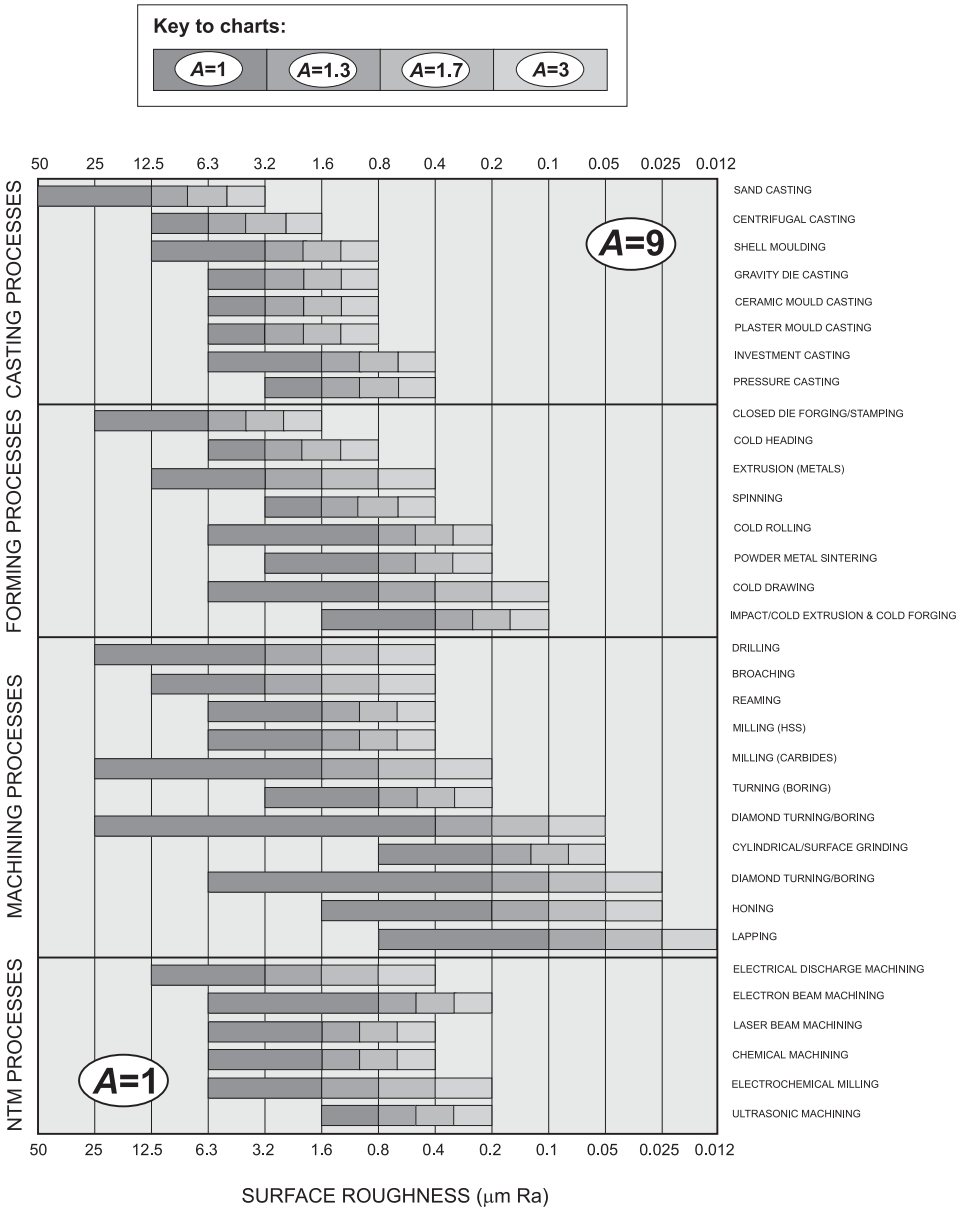
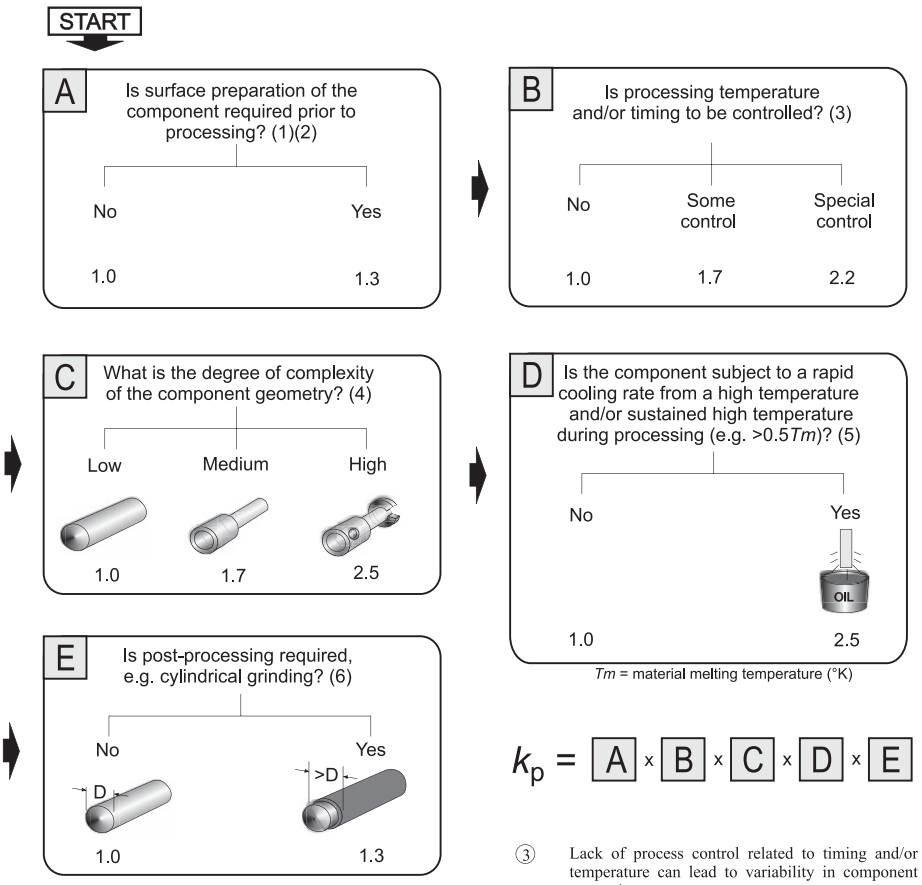


Fig. 7.2.8 Surface roughness risks for a number of manufacturing processes.

The tolerance grades are interpreted using standard tables for conversion into dimensional tolerances. However, the tolerance grades do not take into consideration different materials machined or the complexity of the component being processed.

Both unilateral and bilateral tolerances are encountered in practice. A unilateral tolerance permits variation in only one direction from a nominal or target value; a bilateral tolerance permits variations in both directions. Most tolerances used, unless stated otherwise, in the generation of the maps are bilateral or $\pm t$, where '?' is half of the unilateral tolerance, T. Bilateral tolerances are a common way of representing manufacturing process accuracy, although some processes are more suited to other tolerance representations. For example, forging requires that the total tolerance or unilateral tolerance is divided $+(2/3)T, -(1/3)T$, and drilling has a positive tolerance only, $+T$, the negative tolerance from target being negligible. This is catered for in the representation of the tolerance data in the process capability maps.

After plotting the tolerance data, it is useful, in the first instance, to set the boundary conditions as $A = 1$ corresponding to a dimension/tolerance combination that is of no risk, and $A = 1.7$ on the interface of acceptable/special control region. The data used in the creation of the maps span these two conditions, that is, the region where the process consistently produces the required tolerance. This is shown in Figure 7.2.11 for the



Notes

① Initially, knowledge of the process is required. It is assumed that the component is free from defects, e.g. porosity, as this will affect surface integrity, and free from residual stresses caused by any previous manufacturing process. There is also a risk in the reduction of component fatigue life associated with some surface coating processes. The compatibility between mating surfaces in service must also be addressed because of possible galvanic corrosion failure

② Surface preparation, e.g. polishing, tumbling, blasting, etching or a machine process, may be required due to the nature of the process or due to surface contamination. This can be an added source of variability.

③ Lack of process control related to timing and/or temperature can lead to variability in component properties.

④ Process capability is dependent on the complexity of the component geometry. Avoidance of sharp edges, deep recesses, blind holes, non-uniform sections non-symmetrical sections and slender unsupported sections is important.

⑤ Rapid cooling can cause residual stresses (leading to accelerated corrosion), distortion and quench cracks. The orientation of the component during processing is important in this respect. Sustained high temperatures can cause warping and softening.

⑥ Post-processing, e.g. machining (for dimensional restoration) or shot peening (remove residual stresses), can be an added source of variability.

Fig. 7.2.9 Surface engineering process risk chart, k_p.

turning/boring data taken from the ISO tolerance grades and many other references. The risk index $A = 1.7$ was taken from initial work in this area, where the empirical values for the component manufacturing variability risks determined were compared to historical C_{pk} data.

The intermediate values for '*A*' are derived from the 'squared' relationship that is analogous to that of the relative cost/difficulty trend exhibited by manufacturing processes and their tolerance capability (see Figure 7.2.12).

A target process capability value, $C_{pk} = 1.33$, is aligned to the risk value at $A = 1.7$. Values of '*A*' greater than 1.7 indicated on the maps continue with the squared relationship, therefore $(1.7)^2 = 3$ and $(3)^2 = 9$. It follows then that risk indices of '*A*' greater than 1.7 would not be process capable. In essence, the spacing of the lines $A = 1$ to $A = 9$ represents decreasing percentages of the tolerance band at any given dimension as the value of '*A*' increases. However, the log–log axis as used on the maps shows the difference as a linear step. A further development of the use of the maps is that the '*A*' values can be interpolated between $A = 1$ and $A = 9$ values bounded on the map. This ultimately improves the accuracy in determining the risk value. Therefore, to determine the tolerance risk value '*A*', look along the horizontal axis until the characteristic dimension is found, and locate the adjusted tolerance on the vertical

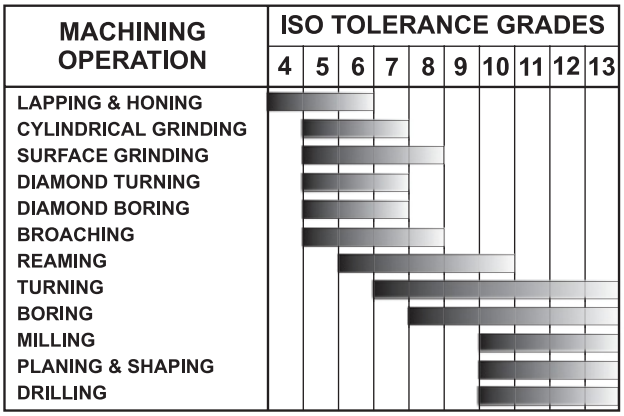


Fig. 7.2.10 ISO tolerance grades for machining processes (adapted from Green, 1992).

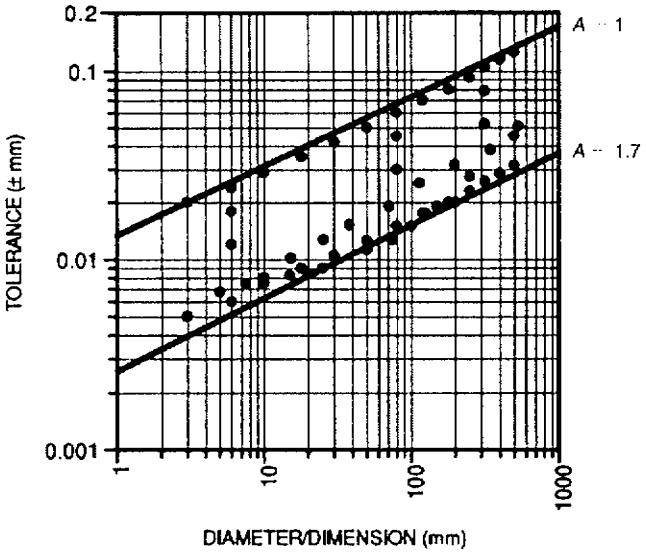


Fig. 7.2.11 Employment of the tolerance data in the generation of a process capability maps.

axis. Read off the '*A*' value in the zone at which these lines intersect on the map by interpolating as required between the zone bands, $A = 1$ to $A = 9$.

The knowledge contained in the maps is also useful in determining the tolerance requirement at an early stage in the detailed design process. In this capacity, the region of process capable tolerance is bounded by two bold lines at $A = 1$ and $A = 1.7$ on the maps. Of course, this does not take into consideration the material and geometry effects initially, for example parting line allowances. Reference to the 1997 work of Swift and Booker can be made for approximate parting line allowances. In most cases, guidance is also given on the maps for the need of a secondary process if the dimension/tolerance combination defined gives a risk index greater than 3 (which is considered to be out of manufacturing control).

7.2.2.1.1 Constructing process capability maps from manufacturing data

The procedure shown in Figure 7.2.13 can be used by a company to construct their own process capability maps. It is necessary only when a new or specialized manufacturing process is to be used and when data from the machine tool supplier are not available. Of course, this activity may become prohibitive when the costs involved and/or time in performing such studies are high.

The data used to generate the maps are taken from a simple statistical analysis of the manufacturing process and is based on an assumption that the result will follow a normal distribution. A number of component characteristics (for example, a length or diameter) are measured and the achievable tolerance at different conformance levels is calculated. This is repeated at different characteristic sizes to build up a relationship between the characteristic dimension and achievable tolerance for the manufacture process. Both the material and geometry of the component to be manufactured are considered to be ideal, that is, the material properties are in specification, and there are no geometric features that create excessive variability or which are on the limit of processing feasibility. Standard practices should be used when manufacturing the test components and it is recommended that a number of different operators contribute to the results.

7.2.2.2 Surface roughness chart

Figure 7.2.8 shows the range of surface roughness values likely for various manufacturing processes. The ranges determined are bounded within the risk index, *A*, in the same way as the process capability maps, because a similar cost–surface finish relationship exists, as suggested for tolerance and cost. This is shown in Figure 7.2.14 for several machining processes. The finer the surface finish required, the longer the manufacturing time, thereby increasing the cost.

For a given manufacturing process, shown on the vertical axis in Figure 7.2.8, and design surface roughness, shown on the horizontal axis, the risk index '*A*' on the shaded band to the right is taken as representative of the risk of obtaining the desired surface roughness. At $A = 1$, this corresponds to a surface roughness that is easily achieved by the manufacturing process. The intermediate values of '*A*' ($A = 1.3$, $A = 1.7$, $A = 3$) correspond to increasing technical difficulty/cost to achieve the design surface roughness. $A = 9$ corresponds to a surface roughness that is technically unattainable using the manufacturing process. A secondary (and possibly a tertiary) process may be required to achieve the design surface roughness. All surface roughness values are in μm Ra, which is the most common format.

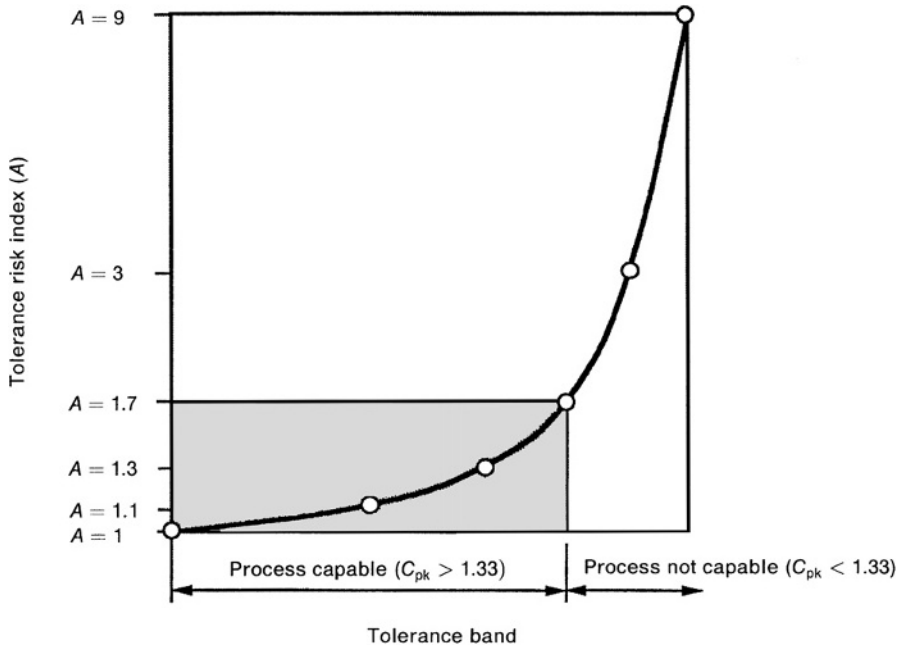


Fig. 7.2.12 Modelling tolerance risk using a 'squared' relationship.

7.2.2.3 Validation of the component manufacturing variability risks analysis

Validation of the component manufacturing variability risk, q_m, is essential to CA in determining C_{pk} estimates for component characteristics at the design stage. Collecting component parts from various industrial sources with known statistical histories was central to this. The components were taken from a number of collaborating companies which had produced the components and had measured a critical characteristic using SPC, therefore the process capability indices C_{pk} and C_p could be determined.

The critical characteristic on each component was analysed, q_m calculated from the analysis and the value obtained was plotted against the process capability indices, C_{pk} and C_p, for the characteristic in question. Note that some components studied have a zero process capability index. This is a default value given if the process capability index calculated from the SPC data had a mean outside either one of the tolerance limits, which was the case for some of the components submitted. Although it is recognized that negative process capability indices are used for the aim of process improvement, they have little use in the analyses here. A correlation between positive values (or values which are at least within the tolerance limits) will yield a more deterministic relationship between design capability and estimated process capability.

Figure 7.2.15(a) shows the relationship between q_m and C_{pk} for the component characteristics analysed. Note, there are six points at $q_m = 9$, $C_{pk} = 0$. The correlation coefficient, r, between two sets of variables is a measure of the degree of (linear) association. A correlation coefficient of ± 1 indicates that the association is deterministic. A negative value indicates an inverse relationship. The data points have a correlation coefficient, $r = -0.984$. It is evident that the component manufacturing variability risks analysis is satisfactorily modelling the occurrence of manufacturing variability for the components tested.

Confidence limits are also drawn on Figure 7.2.15(a) to give boundaries of C_{pk} for a given q_m determined from the analysis, which are within 95%. The relationship between q_m and C_{pk} is described by a power law after linear regression giving

$$C_{pk} = \frac{4.093}{q_m^{2.071}} \tag{7.2.5}$$

This can be approximated to

$$C_{pk} \approx \frac{4}{q_m^2} \tag{7.2.6}$$

Note that the 'squared' relationship which was initially used to model the degree of difficulty in obtaining more capable tolerances for a given manufacturing route and product design is being returned by the power law. Similarly, a relationship between the process capability index C_p and q_m for the components analysed is shown in Figure 7.2.15(b). The data points have a correlation

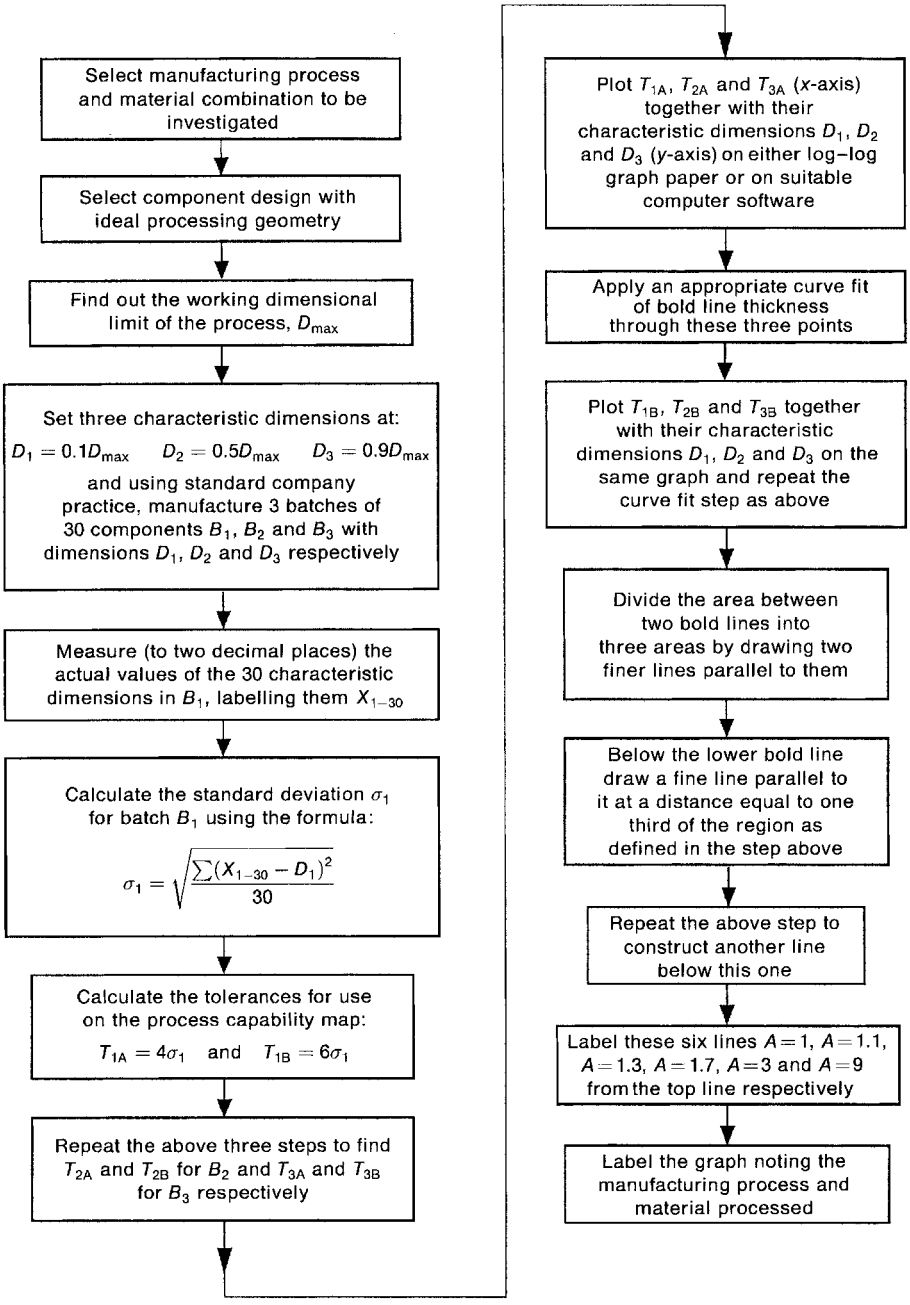


Fig. 7.2.13 Process capability map construction flow chart.

coefficient, $r = -0.956$, and the corresponding power law is given by

$$C_p = \frac{3.981}{q_m^{1.332}} \tag{7.2.7}$$

This can be approximated by the following equation:

$$C_p \approx \frac{4}{q_m^{4/3}} \tag{7.2.8}$$

It is possible, therefore, to determine an estimate for C_{pk} from the formulation given above in Equation (7.2.6), within some confidence. It is assumed that the C_{pk} values given for the component characteristics were based on the mean values for a number of batches produced and that a confidence limit of 95% is adequate for the outcome for the prediction of C_{pk}. These values will be later transferred to the quality-cost model for the prediction of the failure costs for given levels of process capability and failure severity (see Section 7.2.5).

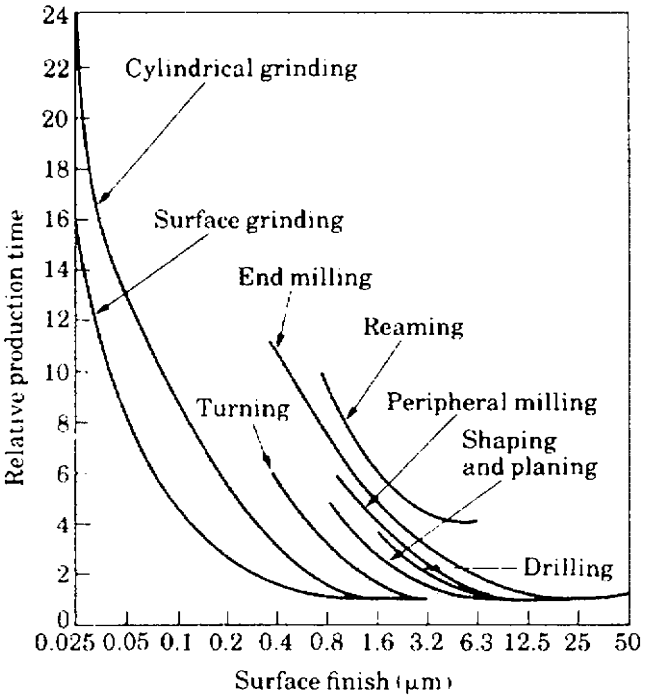


Fig. 7.2.14 Relative manufacturing time as a function of surface roughness for several machining processes (BS 1134,1990).

7.2.3 Assembly capability

When considering assembly processes, many problems are due to relatively minor design details. It must be remembered that a product is often an assembly of sheet metal, pipes, wires and a variety of other components held together by nuts, bolts and rivets. The attention to detail cannot be overemphasized at this stage of product design. The basic question we need to ask is, 'Can we assemble it every time, not just the once, and if not, why?'

Assembly variation (along with manufacturing variation) is a major contributor to poor quality and increased costs. For example, when the assembly of a poorly designed and poorly made product is attempted, faults such as accumulated tolerance error, incompatible dimensions and difficult part installation become apparent. At the same time, the cost of recovering from these problems during the late phase of production is high. Industry has recognized the need to reduce assembly variations.

The three main sources of variation in mechanical assemblies are:

- Dimensional variations (lengths, angles)
- Form and feature (flatness, roundness, angularity)
- Kinematic variations (small adjustments between mating parts).

The above are all closely linked to manufacturing variability depending on the characteristic associated with the product. Manufacturing is an important aspect of assembly too. The design of products for assembly requires careful consideration of many factors that influence the functionality and manufacturability. For example, while stability and relative precision of part positions are often essential for the functional performance of assemblies, these same requirements may make the product difficult to manufacture. At the same time, dimensional clearance among parts is essential to create paths for assembly operations. In general, close (or numerically small) tolerances must be maintained on parts that are to be assembled with other parts, and the closer the tolerances, the greater the ease of assembly.

'Assemblability' is a measure of how easy or difficult it is to assemble a product. Tolerances affect the assemblability of a product, which in turn affects the cost of the product because of the scrap cost, and wasted time and energy. It is important to predict the probability of successful assembly of the parts so that the tolerance specifications can be re-evaluated and modified if necessary to increase the probability of success and lower the production cost associated with assembly. While close tolerances reduce assembly costs, they increase the cost of manufacturing the individual parts. A balance of the two types of costs must be achieved, with the objective of minimizing their sum.

When tolerances are mentioned in terms of mechanical assemblies, one considers tolerance or assembly stack models used to determine the final assembly tolerance capability for a given set of component tolerances. Worst case or statistical equations are commonly used for optimization purposes. However, this can be confusing, as a tolerance can exist on a component characteristic, but the component itself may not necessarily be capable of being assembled to form the assembly tolerance. This may be due to one or more features of the components being assembled preventing effective and repeatable placement. If this occurs then the product suffers not only from an assembly problem, but also a tolerance problem.

CA provides assembly capability predictions which can highlight potential assembly problems. We must also have a means of estimating the failure costs associated with component assembly processes in addition to the tolerance process capability predictions. These additional failure costs are independent of whether the tolerances assigned to an assembly stack or the single characteristic are capable. By identifying components with high assembly risks and potentially high failure costs, further design effort is highlighted and performed in order to identify the associated

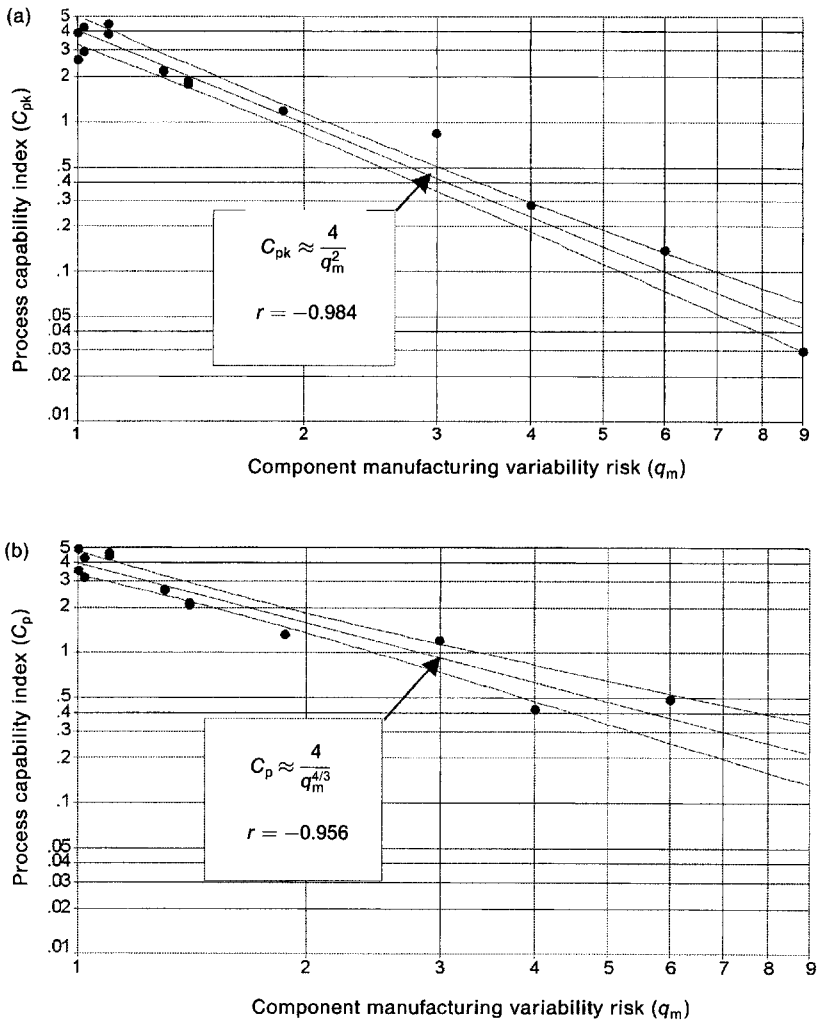


Fig. 7.2.15 Empirical relationships between (a) q_m and C_{pk} and (b) q_m and C_p (with 95% confidence limits).

tolerances, for example clearance for the optimal fit and function of the components.

7.2.3.1 Design for assembly techniques

Early work looking at designing products for mechanized assembly started over 30 years ago. Large cost savings were found to be made by careful consideration of the design of the product and its individual component parts for ease of assembly. Commercial DFA techniques are now used successfully by many companies in either workbook or software versions. The three methods which are most referred are:

- Boothroyd–Dewhurst Design for Manufacture and Assembly (DFMA)
- Computer Sciences Corporation's (CSC) DFA/Manufacturing Analysis (MA)
- Hitachi's Assembly Evaluation Method (AEM).

In fact, the use of DFA techniques is now mandatory in some companies, such as Ford and LucasVarity (now TRW). These techniques offer the opportunity for a number of benefits, including:

- Reduced part count
- Systematic component costing and process selection
- Lower component and assembly costs
- Standardized components, assembly sequence and methods across product 'families' leading to improved reproducibility
- Faster product development and reduced time to market
- Lower level of engineering changes, modifications and concessions
- Fewer parts means: improved reliability, fewer stock costs, fewer invoices from fewer suppliers and possibly fewer quality problems.

A team-based application and systematic approach are essential as there are many subjective processes

embedded, but many companies have found them to be pivotal techniques in designing cost-effective and competitive products.

Huang in 1996 gave an overview of each of the main commercial methods; however, in general, a number of design for assembly guidelines can be highlighted:

- Reduce part count and types
- Modularize the design
- Strive to eliminate adjustments (especially blind adjustments)
- Design parts for ease of handling (from bulk)
- Design parts to be self-aligning and self-locating
- Ensure adequate access and unrestricted vision
- Design parts that cannot be installed incorrectly
- Use efficient fastening or fixing techniques
- Minimize handling and reorientations
- Maximize part symmetry
- Use good detail design for assembly
- Use gravity.

DFA minimizes part count by, for example, consolidating a number of features found on two components in a single component, but it cannot indicate the process capability for the design proposed. DFA techniques can only indicate that it may be costly to assemble the component or that the cost of the component may be relatively high, but still more cost effective than the two components it replaces. Production costs should not be the only measure of performance with which to select designs. The analysis should extend to the potential variability associated with the design when in production.

There is a strong correlation between assembly efficiency and reported defect levels for a number of Motorola products evaluated using the Boothroyd–Dewhurst DFMA technique. From the results of the study, it is claimed that DFA cuts assembly defects by 80% and, therefore, has a direct influence on manufacturing quality. This may be due to the fact that more efficient design solutions have relatively fewer component parts which naturally infers fewer quality problems. However, early life failures, which are caused by latent defects, are not necessarily highlighted by DFA techniques.

Quality or robustness is seen as a natural outcome of the product when effectively using DFA techniques, and is commonly listed as a potential benefit, but it is not evaluated explicitly by any of the commercial techniques. Too often, assembly is overlooked when assessing the robustness of a design and DFA techniques do not specifically address variability within the assembly processes. Undoubtedly, the better the assemblability, the better the product quality in terms of fewer parts and simpler assembly operations. Fewer parts lead to fewer breakdowns, fewer workstations, less time to assemble and fewer overheads. Simpler assembly operations imply that the product fits together more easily, leading to shorter lead times and less rework.

From the above arguments, it has been recognized that there is a need to predict variation at the assembly level. To facilitate an assembly analysis in the first instance, it is essential to understand the structure of the proposed product design and the assembly sequence diagram is useful in this respect.

7.2.3.2 Assembly sequence diagrams

Usually, the assembly sequence of the components that make up the product is examined late in the design process when manufacturing engineers are trying to balance the assembly line. In turn, the choice of assembly sequence and the identification of potential subassemblies can affect or be affected by product testing options, market responsiveness and factory floor layout, as well as engineering issues such as ease of servicing or ease of recycling. However, there are extensive implications for the successful generation of assembly sequences during the early stages of the design process. Cycle times can be reduced and rework is decreased.

Through an assembly sequence diagram for each component in the product, the assembly variability risks highlighted by an analysis are logically mapped. An assembly sequence declaration compels the designer to focus on each stage in the assembly and therefore makes the task of identification of potential problems much easier. Only CSC's DFA/MA method uses an explicit assembly sequencing method to augment the DFA process and the approach employed here is developed, with slight modifications, from this. Figure 7.2.16 shows an assembly sequence diagram for a simple castor wheel assembly. The development of assembly sequence diagrams for more complex assemblies does become a prohibitive task, but its generation should be one of the standard engineering design tasks.

7.2.4 Component assembly variability risks analysis

In the development of the assembly variability risks analysis, expert knowledge, data found in many engineering references and information drawn from the CSC DFA/MA practitioner's manual were collated and issues related to variability converged on. Much of the knowledge for the additional assembly variability risks analysis was reviewed from the fabrication and joining data sheets called PRocess Information MAps (PRIMAs).

Product assembly is such a large area that focusing on the key operational issues, such as handling, fitting and joining, is justified at the design stage. Although much has

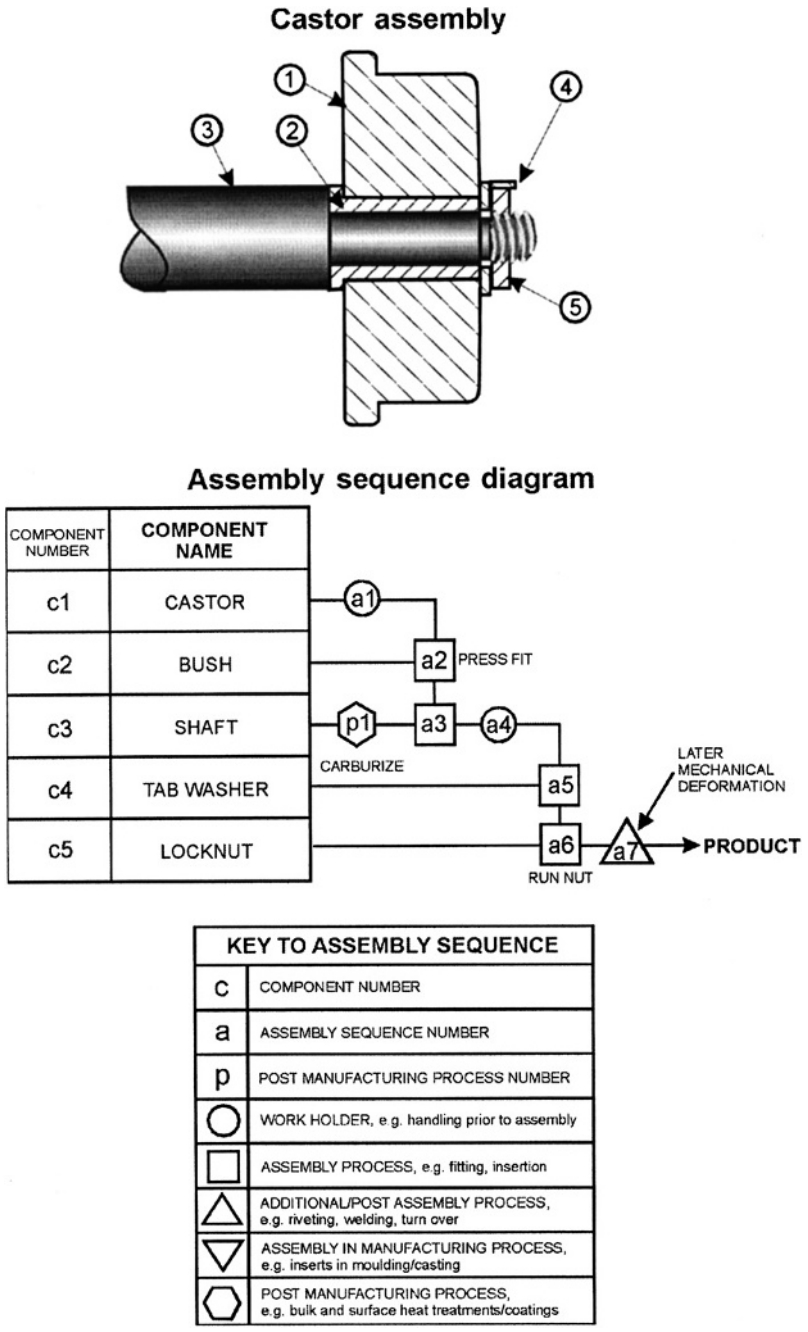


Fig. 7.2.16 Assembly sequence diagram for a castor wheel.

been written about DFA, it was not appropriate to use the DFA indices directly for these assembly operations because they are based on relative cost, not potential variability. Where the knowledge and data cannot help, and where some of the more qualitative aspects of design commence, as well as expert knowledge, some Poka Yoke ideas were used, particularly for the component fitting chart. Poka Yoke is used to prevent an error being converted into a defect. It is heavily involved in the development of any DFA process. Poka Yoke has two main steps: preventing the occurrence of a defect, and detecting a defect. Some of the assembly variability risks charts reflect these Poka Yoke philosophies.

The measure of assembly variability, q_a, derived from the analysis should be used as a relative performance indicator for each design evaluated. The design with the least potential variability problems or least failure cost should be chosen for further development. The indices should not be taken as absolutes as assembly variability is difficult to measure and validate.

The component assembly variability risk, q_a, as determined by CA, attempts to better understand the effects of the assembly situation on variability by quantifying the risks that various operations inherently exhibit. A component's assembly situation typically involves the following process issues:

- Handling characteristics, h_p
- Fitting (placing and insertion) characteristics, f_p
- Additional assembly considerations (welding, soldering, etc.), a_p
- Whether the process is performed manually or automatically.

At the current stage of development, the risk index associated with component assembly variability is the product of the component's handling and fitting risks:

$$q_a = h_p \times f_p \tag{7.2.9}$$

The additional assembly process risk, a_p, is considered in isolation after the handling and fitting operations, questioning the assembly situation of a component after initial placing, and which requires further processing, for example welding or adhesive bonding. Therefore, because additional assembly processes are performed after initial part placement, q_a simply defaults to a_p when it is considered in the assembly sequence:

$$q_a = a_p \tag{7.2.10}$$

The additional assembly processes covered in the analysis are listed below:

- Miscellaneous operations (testing, reorientation, heating, cooling)
- Later mechanical deformation (riveting)
- Adhesive bonding
- Brazing and soldering
- Resistance welding
- Fusion welding.

A systematic assessment is made of every handling and fitting process involved in the product construction (manual or automated), including the effects of alignment, placement and fastening, through the assembly sequence diagram for the design. A number of charts are used for each aspect of the assembly analysis. Using the handling process risk chart (see Figure 7.2.17), the fitting process risk (Figure 7.2.18) and the additional process risk charts, the assembly situation for most components can be analysed, accruing penalties if the design has increased potential for variability. The measure of variability returned is the component assembly variability risk, q_a. Guidance notes accompany each chart for the designer.

The assembly indices are bounded in the same way as the component manufacturing variability risk indices, q_m. Starting with the notion that an ideal component assembly situation exists, this risk index is given a value of unity. The indices are then scaled within the same conditions described for t_p (used in the manufacturing variability risks analysis). For example, $1.7 \leq h_p \leq 2.2$ indicates that the handling process is within special control. High values of q_a indicate high-risk design to process solutions. The estimated values of q_a tend to be within the range 1–5, with only those components where the risk is exceptionally high coming out beyond 5. Where detailed knowledge and experience of a particular assembly situation is available, this expertise should be used to augment the analysis. Because the assembly variability knowledge collated is a mixture of quantitative and qualitative data, validation of the indices used in the charts is more difficult than the manufacturing variability risks analysis. Experts from the assembly industry were therefore used in their final assessment and allocation. However, despite the fact that the validation of the assembly capability measures are difficult to determine empirically, they are found to be of great benefit when comparing and evaluating a number of design principles.

7.2.5 The effects of non-conformance

7.2.5.1 Design acceptability

FMEA can be used to provide a quantitative measure of the risk for a design. Because it can be applied hierarchically from system through subassembly and component levels down to individual dimensions and characteristics, it follows the progress of the design into detail. FMEA also lists potential failure modes and rates their severity (S), occurrence (O) and detectability (D). It therefore provides a possible means for linking potential variability risks with consequent design acceptability and associated costs. Note that the ratings of occurrence and detectability are equated to probability levels.

To determine the costs of failing to design into a product the levels of process capability proper to the severity of consequential failure, it has been assumed that a fault is undetected and always results in a failure. Also, the estimate is not sensitive to changes in the product cost, volume and rework costs. The cost of customer dissatisfaction, over and above the accounted cost of rework or a returned product, is very difficult to assess. The model which follows is based on research with Lucas Automotive electronics and utilizes their specific ratings for FMEA occurrence (O) and severity (S) (see Figures 7.2.19 and 7.2.20 respectively). The analysis of failure consequence is a specialist area within each industry, as is the definition of the probability of occurrence. The alignment of the ratings chosen here

h_p COMPONENT SENSITIVITY / HANDLING TECHNOLOGY					DELICATE (1)	SENSITIVE TO TEMPERATURE CHANGE	SENSITIVE TO CONTAMINATION		NOT APPLICABLE (3)
							Chemical (2)	Mechanical	
MANUAL HANDLING & GRIPPING				Not thin/ small (4)	1.3	1.2	1.5	1.1	1.0
				Thin/small	1.7	1.3	2.0	1.2	1.0
MAGAZINE PROCESSES				No gripping	1.2	1.1	1.2	1.1	1.0
	Gripping	Mechanical clamping		Not thin	1.4	1.1	1.2	1.3	1.0
				Thin	1.9	1.3	1.2	1.4	1.0
		Vacuum/ Magnetic contact			1.2	1.3	1.2	1.3	1.0
AUTOMATED PROCESSES	Parts feeder	No gripping			1.5	1.1	1.2	1.5	1.0
		Gripping	Mechanical clamping	Not thin	1.9	1.2	1.3	1.7	1.0
				Thin	2.2	1.3	1.3	1.8	1.0
			Vacuum/ magnetic contact		1.6	1.3	1.3	1.8	1.0
	Conveyor (5)	No gripping			1.2	1.1	1.2	1.2	1.0
		Gripping	Mechanical clamping	Not thin	1.5	1.2	1.5	1.7	1.0
				Thin	2.2	1.3	1.5	1.8	1.0
			Vacuum/ magnetic contact		1.3	1.3	1.5	1.8	1.0

Notes

(1) Also includes high finish.

(2) Chemical contamination to operator machine or product.

(3) If component is not sensitive to the characteristics listed, $h_p = 1.0$.

(4) A component is said to be thin if one of its major dimensions is ≤ 0.25 mm thick.

(5) Includes mechanical and air assisted mechanisms.

Fig. 7.2.17 Handling process risk, h_p.

with other existing FMEA ratings may be considered when using CA with other business practices.

'Failure' in the context here means that product performance does not meet requirements and is related back in the design FMEA to some component/characteristic being out of specified limits – a fault. The probability of occurrence of failure (O) caused by a fault can be expressed as

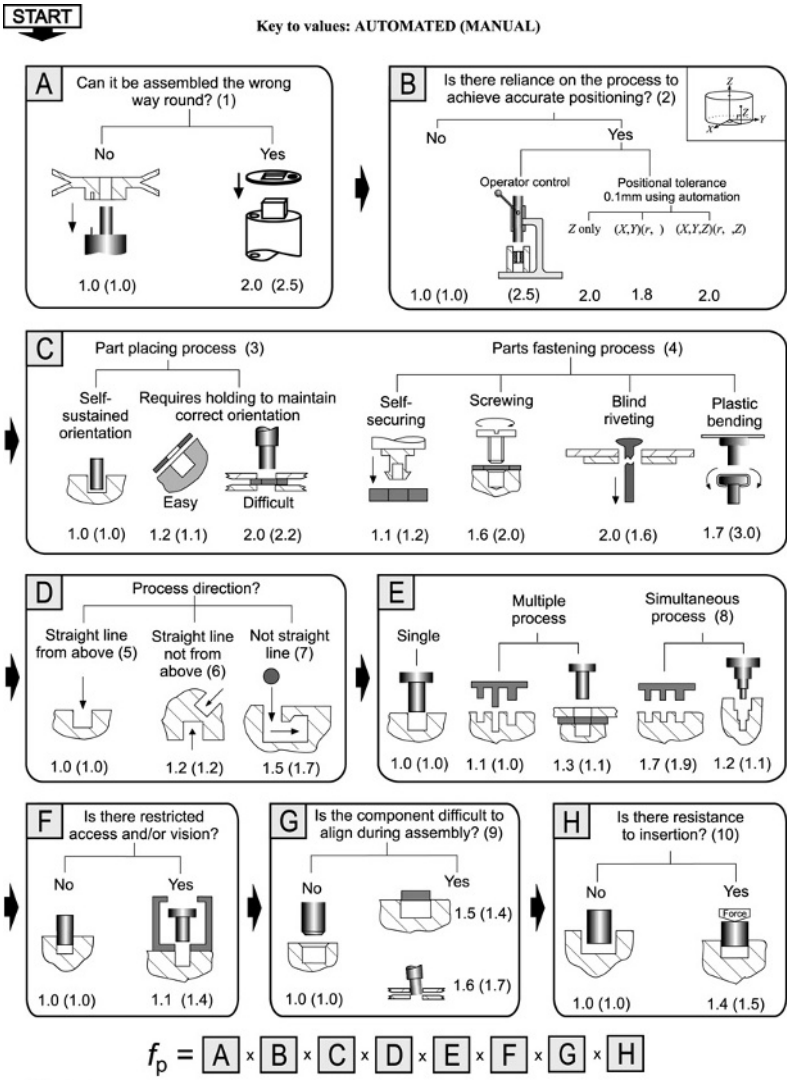
$$O = P_f \times d' \times P_{of} \tag{7.2.11}$$

where P_f is the probability of a fault in the component/characteristics; d' the probability that the fault will escape detection; P_{of} the conditional probability that, given there is a fault, the failure will actually occur.

Equation (7.2.11) recognizes that for a failure to occur, there must be fault, and that other events may need to combine with the fault to bring about a failure. The equation states that the probabilities of each of these factors occurring must be multiplied together to calculate the probability of failure.

In the special case when $d' = 1$ and $P_{of} = 1$, that is, when if a fault occurs it will not be detected and that failure is certain, occurrence equates directly to the probability of a fault. The probability of a fault in turn depends on the capability of the process used for the component/characteristic.

A standard for the minimum acceptable process capability index for any component/characteristic is normally set at $C_{pk} = 1.33$, and this standard will be used later to align costs of failure estimates. If the characteristics follow a normal distribution, $C_{pk} = 1.33$ corresponds to a fault probability of



$$f_p = A \times B \times C \times D \times E \times F \times G \times H$$

Notes

① This refers to components which can be assembled the wrong way round to the detriment of the Product function, but may not be obvious to the Product assembly process.

② Reliance on the assembly process to achieve accurate positioning of one component relative to another with no stop encountered may cause variability.

③ The part is unstable after insertion or during subsequent assembly operations and will require gripping or holding down before being secured.

④ Part is a fastener or will act as a fastener or is provided with a built-in fastening element which is secured following insertion. Screwing operations include torque settings.

⑤ Part can be inserted into a part assembly or work carrier by single axis straight line motion from vertically above.

⑥ Insertion of the part into the assembly build necessitates straight line access motion other than from the vertically above (e.g. from the side or below).

⑦ Part cannot be inserted by simple single axis straight line motion. For example, the part axis may be obstructed requiring a change in the direction of motion.

⑧ Two or more parts which are simultaneously being placed, fastened or inserted.

⑨ Part initial insertion is helped by well designed chamfers or other features which ensure required alignment and positioning.

⑩ Considerable resistance may be encountered during part insertion due to small clearances, snagging, or insertion against a force such as a spring. Damage to components due to resistance to insertion is the key issue, e.g. pick-up, distortion and out of tolerance effects.

Fig. 7.2.18 Fitting process risk, f_p.

Occurrence (O)

Rating	Guide	ppm
1	Remote possibility of failure occurring	0.1
2	Low possibility of occurrence	0.5
3		2
4	Moderate possibility of occurrence	10
5		50
6	Significant number of failures possible	200
7	High possibility of occurrence	1,000
8		5,000
9	Very high possibility of occurrence	20,000
10	Almost certain that many failures will occur	100,000

Fig. 7.2.19 Typical FMEA occurrence ratings (*O*).

Severity (*S*)

Rating	Guide to effect on user
1 *	No effect or minimal effect on customer
2 *	Minor annoyance to customer
3 *	Annoyance to customer but no loss of major function
4 *	Possible return to manufacturer
5 *	Definite return to manufacturer
6	Failure leading to violation of statutory requirement
7	Failure leading to injury or a more safety critical related problem with secondary back-up
8	Safety problem – degradation of function with possible severe injury
9	Complete failure with probable severe injury and/or loss of life
10	Catastrophic failure with high probability of loss of life

* Note: These failures are non-safety critical

Fig. 7.2.20 Typical FMEA Severity Ratings (*S*).

$$P_f = 30 \text{ ppm} = 30 \times 10^{-6} \tag{7.2.12}$$

While 30 ppm may be acceptable as a maximum probability of ocurrence for a failure of low severity, it is not acceptable as severity increases. An example tables of FMEA Severity Ratings was shown in Figure 7.2.20. In the 'definite return to manufacturer' (a warranty return) or 'violation of statutory requirement' region ($S = 5$ or $S = 6$), the designer would seek ways to enhance the process capability or else utilize some inspection or test process. Reducing d' will reduce occurrence, as indicated by Equation (7.2.11), but inspection or test is of limited efficiency.

If severity increases to, say, 'complete failure with probable severe injury and/or loss of life' ($S = 9$), the designer should reduce occurrence below the level afforded by two independent characteristics protecting against the fault. If each is designed to, then this implies a failure rate of

$$P_f = \left(30 \times 10^{-6}\right)^2 = 9 \times 10^{-10} \tag{7.2.13}$$

Again, this standard aligns with the costs of failure analysis below.

Figure 7.2.21 is a graph of occurrence against severity showing a boundary of the acceptable design based on these criteria for the case when $d' = 1$ and $P_{of} = 1$. The graph is scaled in terms of FMEA ratings for occurrence, equivalent probabilities and parts per million (ppm).

In general, just as reducing d' reduces the probability of failure occurrence, so reducing the conditional probability of failure P_{of} moves the component/characteristic towards the acceptable design zone. However, this must be applied with considerable caution. The derivation of the acceptable maximum for failure occurrence at $S = 9$ and above is essentially an application of the conditional probability: the conditional probability of failure at one characteristic is the probability of failure at the other protecting characteristic. In this case, it would normally be more appropriate to design two characteristics than to rely on one very capable one. At such low probability levels, the chance of random unforeseen phenomena throwing the process out of control is significant.

On the other hand, consider the case of a 'secondary back-up' (external to the system) for a component/characteristic of severity $S = 7$. If the designer assumes the back-up is designed for $C_{pk} = 1.33$, then the analysis would reduce that for the internal back-up $S = 9$ case above. The acceptable design limits in Figure 7.2.21 imply that it is prudent to allow a safety margin for external components not under the designer's control and scale up the conditional probability of failure.

In summary, for a component/characteristic it is possible to define an area of acceptable design in a graph of

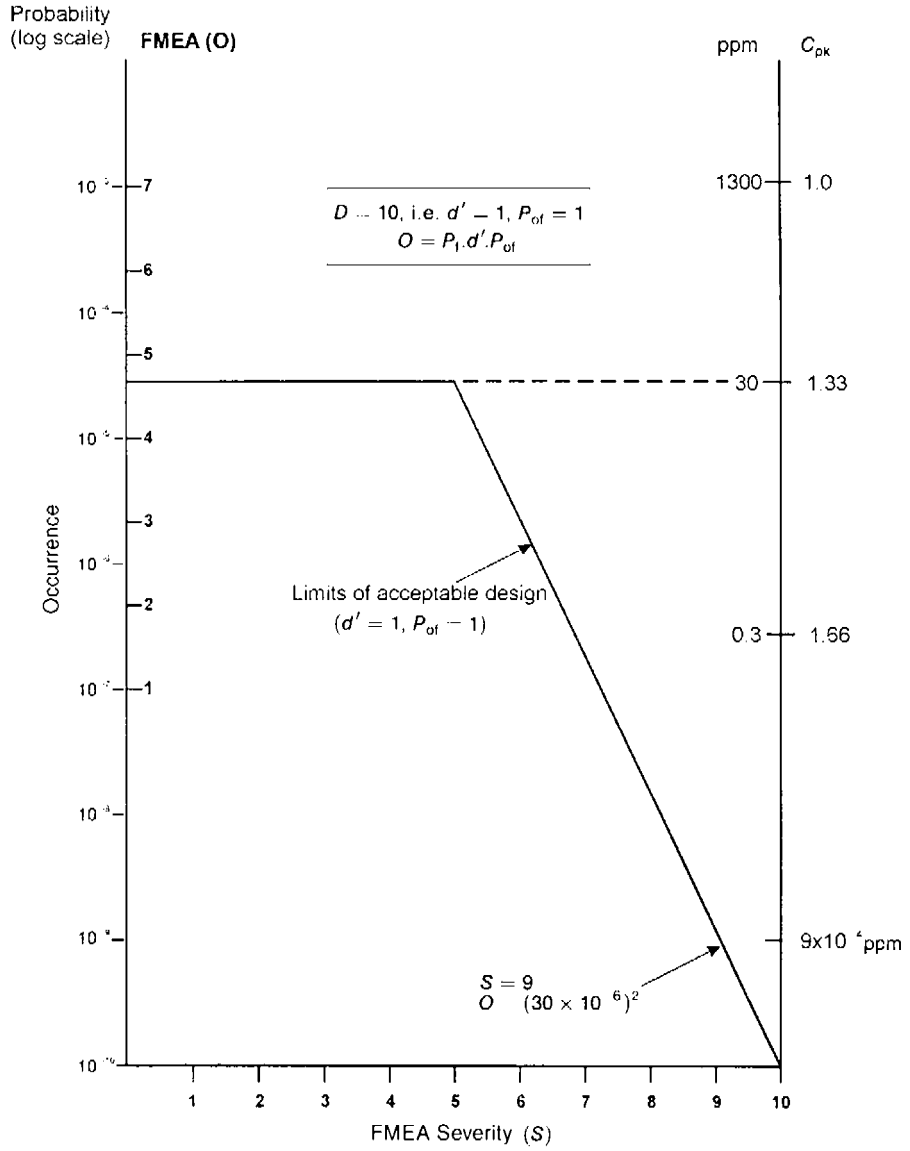


Fig. 7.2.21 The limits of acceptable design.

occurrence versus severity. The acceptability of the design can be enhanced somewhat by the addition of inspection and test operations. The requirements of process capability may be relaxed to a degree, as the conditional probability of failure reduces, but this should be subject to a generous safety margin. It will now be demonstrated that the boundaries of acceptable design proposed above can be expressed in terms of the costs of failure.

7.2.5.2 Map of quality costs

From the quality-cost arguments made in Section 7.1.2, it is possible to plot points in the graph of occurrence versus severity and construct lines of equal failure cost (% isocosts). Figure 7.2.22 shows this graph, called a Conformability Map. Because of uncertainty in the estimates, only a broad band has been defined.

The boundary of acceptable design for a component or assembly characteristic in the zone $S \geq 6$ corresponds fairly closely to a failure cost line equivalent to 0.01% of the unit cost. The region of unacceptable design is bounded by the intersection of the horizontal line of $C_{pk} = 1$ and the 1% isocost. A process is not considered capable unless $C_{pk} \geq 1$ and a failure cost of less than 1% is thought to be acceptable.

Components/characteristics in the unacceptable design zone are virtually certain to cause expensive failures unless redesigned to an occurrence level acceptable for their failure severity rating (minimum $C_{pk} = 1.33$). In the intermediate zone, if acceptable design conformability cannot be achieved, then special control action will be

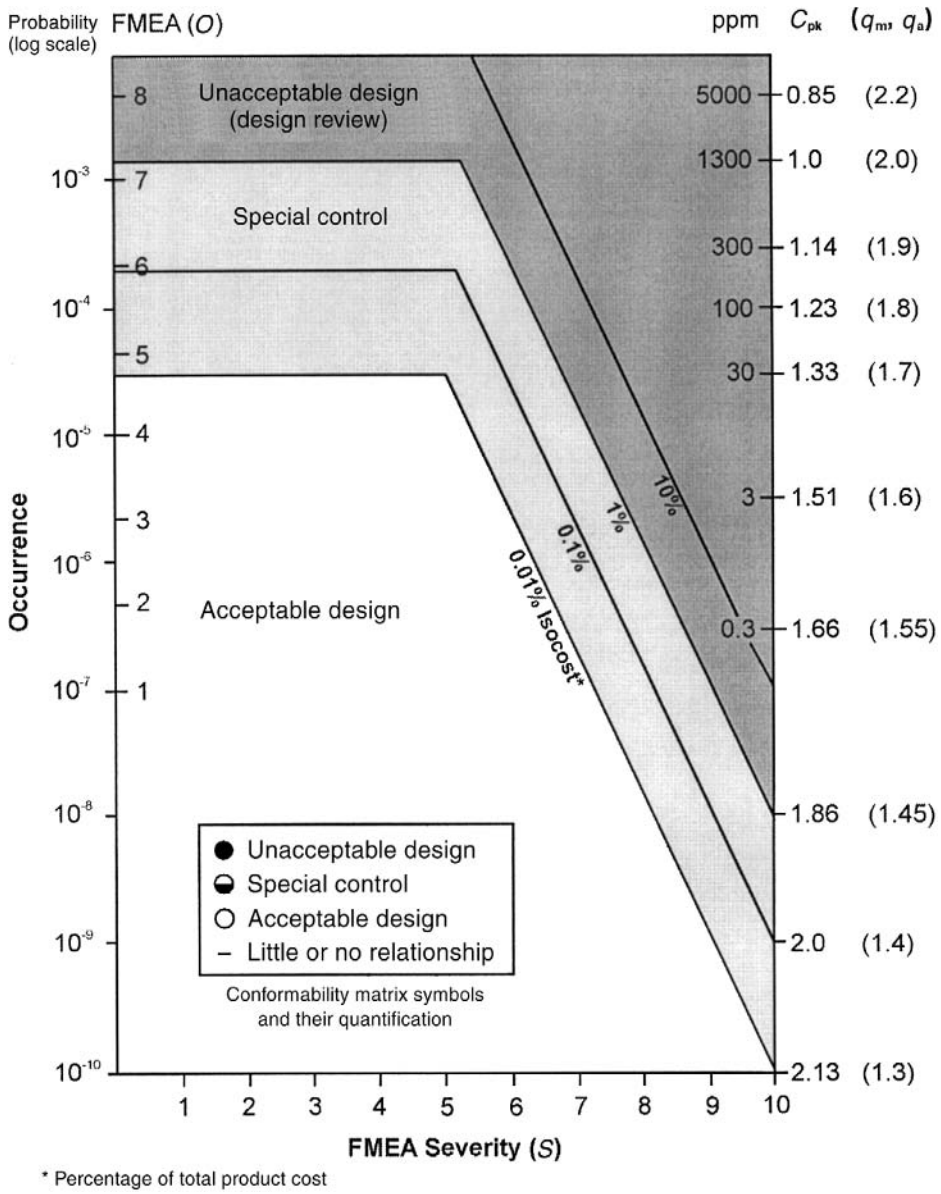


Fig. 7.2.22 The Conformability Map.

required. If special action is needed then the component/characteristic is critical. However, it is the designer's responsibility to ensure that every effort is made to improve the design to eliminate the need for special control action.

The 0.01% line on Figure 7.2.22, implies that even in a well-designed product, there is an incurred failure cost. Just 100-dimensional characteristics on the limit of acceptable design are likely to incur a 1% cost of failure. The quality cost rises dramatically where design to process capability is inadequate. Just 10 product characteristics on the 1% isocost line would give likely failure costs of 10%. (The 1% and 10% isocost lines on the Conformability Map are extrapolated from the conditions for the 0.01% and 1% isocost lines.) Clearly, the designer has a significant role in reducing the high cost of failure reported by many manufacturing organizations.

Variation is typically measured using process capability indices, as well as ppm failure, but it would be advantageous to scale the variability risks predicted by CA on the Conformability Map too. Scales showing C_{pk} and approximate variability risk measures, q_m and q_a, from CA equivalents have been added to the Conformability Map for use in the CA methodology. An empirical relationship exists between C_{pk} and q_m that allows us to do this, as discussed earlier. Because the assembly variability risk, q_a, is also scaled in the same way as q_m, it is assumed that the level of variability in assembly can also be associated to a notional process capability index. The

Conformability Map, through the inclusion of q_m, q_a and FMEA Severity Rating (S) for a particular design, can be used to give the total risk potential in terms of the isocost at the design stage. The accumulated isocosts can then be translated into potential failure costs. In this way, a total cost of failure for a design can be evaluated.

The use of the Conformability Map in determining the potential costs of non-conformance or failure should be associated with the application of evaluating and comparing different design schemes in practice. The estimated failure costs act as a measure of performance by which to make a justified selection of a particular design scheme. Assumption of costs as absolute values is not recommended due to the difficulty in obtaining accurate data for the quality-cost model (see later for guidance on the use and calculation of potential costs of failure for a design). It has been cited that ±5% is a good enough accuracy for the prediction of quality costs at the business assessment level, but the achievement of this at the design stage is a very difficult task.

The Conformability Map enables appropriate C_{pk} values to be selected and through the link with the component variability risks, q_m and q_a, it is possible to determine if a product design has characteristics that are unacceptable, and if so what the cost consequences are likely to be. The two modes of application are highlighted on Figure 7.2.23. Mode A shows that the quality loss associated with a characteristic at $C_{pk} = 1$ and FMEA

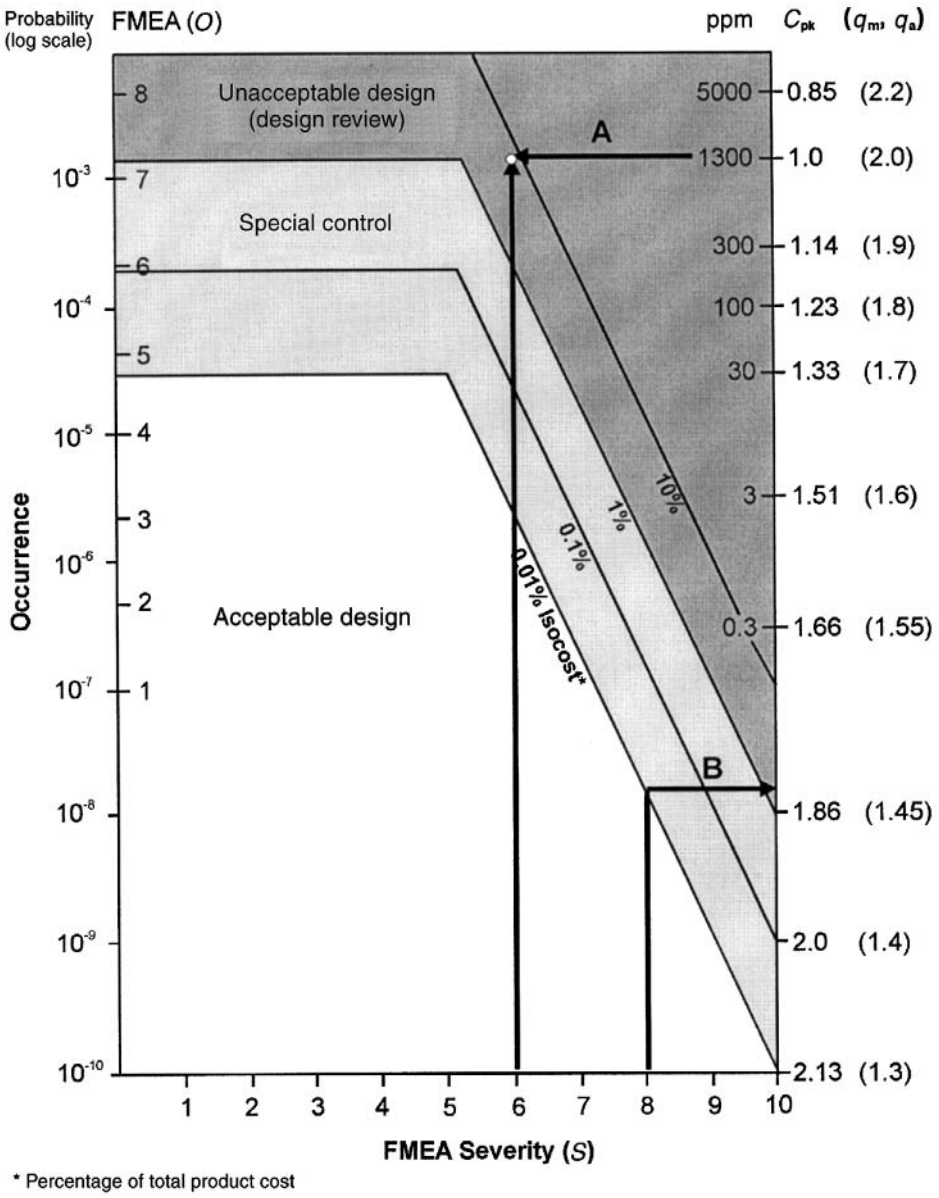


Fig. 7.2.23 Conformability Map applications.

Severity (S) = 6 could potentially be 8% of the total product cost. Mode B indicates that a target $C_{pk} = 1.8$ is required for an acceptable design characteristic having an FMEA Severity Rating (S) = 8.

7.2.6 Objectives, application and guidance for an analysis

A short review of the CA process is given before proceeding with the applications of the technique to several industrial case studies. The three key stages of CA are shown in Figure 7.2.24 within the simplified process of assessing a design scheme.

- *Component manufacturing variability risks analysis* – As mentioned previously, the first of the three key stages in CA is the component manufacturing variability risks analysis. When detailing a design, certain characteristics can be considered important. For example, tolerances on a dimension, a surface roughness requirement or some intricate geometry on a component. These important characteristics need to be controlled and understood.

The 'quality' inherent in manufacturing a component is essentially dictated by the designer, and can be accounted for by the level of knowledge in the following design/manufacture interface issues:

- Process precision and tolerance capability, t_p
- Material to process compatibility, m_p
- Component geometry to process limitations, g_p
- Surface roughness and detail capability, s_p
- Surface engineering process suitability, k_p.

From consideration of these issues, risk indices can be formulated using charts to reflect the limits of variability set by the design which can be achieved by the application of best practice in manufacturing. The analysis ultimately returns in a component manufacturing variability risk, q_m. The underlying notion of q_m is that an ideal design exists with regard to

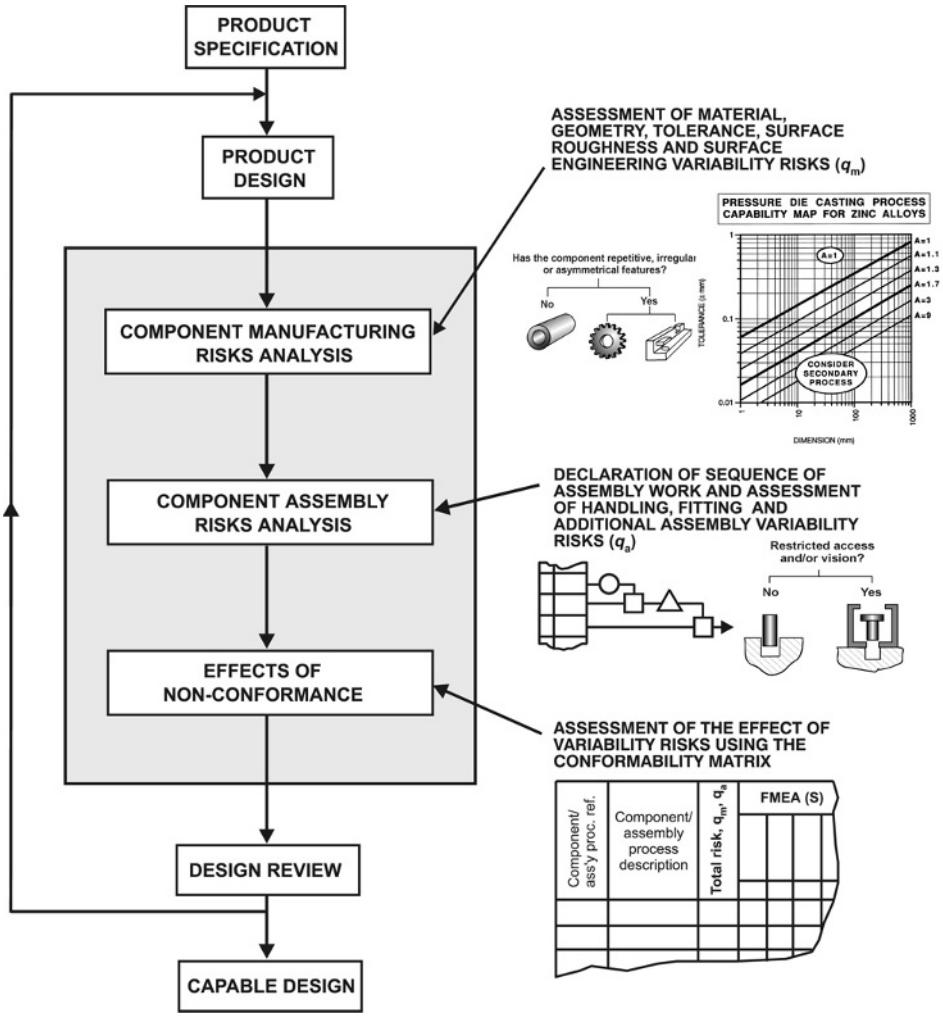


Fig. 7.2.24 CA methodology.

tolerance where the risk index is unity, indicating that variability is in control. Risk indices greater than unity exhibit a greater potential for variability during manufacture. Central to the determination of q_m is the use of the process capability maps showing the relationship between the achievable tolerance and the characteristic dimension for a number of manufacturing processes and material combinations.

- *Component assembly variability risks analysis* – Too often, assembly is overlooked when assessing the robustness of a design. DFA techniques offer the opportunity for part count reduction through a structured analysis of the assembly sequence, but they do not specifically address variability within assembly processes. A component's assembly situation involves the following process issues:
  - Handling characteristics, h_p
  - Fitting (placing and insertion) characteristics, f_p
  - Additional assembly considerations (welding, soldering, etc.), a_p
  - Whether the process is performed manually or automatically.

  Again, the notion is that an ideal component assembly design exists. Using expert knowledge, charts for a handling process risk, fitting process risk and additional processes question the assembly situation of the component, accruing penalties if the design has increased potential for variability to return the Component Assembly Variability Risk, q_a.
- *Effects of non-conformance* – The link between the component variability risks, q_m and q_a, and FMEA severity (S) is made through the Conformability Map. Research into the effects of non-conformance and associated costs of failure has found that an area of acceptable design can be defined for a component characteristic on a graph of occurrence versus severity. It is possible to plot points on this graph and construct lines of equal failure cost. These isocosts in the non-safety critical region (FMEA Severity Rating ≤ 5) come from a sample of businesses and assume levels of cost at internal failure, returns from customer inspection or test and warranty returns. The isocosts in the safety critical region (FMEA Severity Rating > 5) are based on allowances for failure investigations, legal actions and product recall, but do not include elements for loss of current or future business which can be considerable. The costs in the safety critical area have been more difficult to assess and have a greater margin of error. In essence, as failures get more severe, they cost more, so the only approach available to a business is to reduce the probability of occurrence.

7.2.6.1 Objectives

CA is primarily a team-based product design technique that, through simple structured analysis, gives the information required by designers to achieve the following:

- Determine the potential process capability associated with the component and/or assembly characteristics.
- Assess the acceptability of a design characteristic against the likely failure severity.
- Estimate the costs of failure for the product.
- Specify appropriate C_{pk} targets for in-house manufacturing or outside supply.

7.2.6.2 Application modes

The case studies that follow have mainly come from 'live' product development projects in industry. Whilst not all case studies require the methodology to predict an absolute capability, a common way of applying CA is by evaluating and comparing a number of design schemes and selecting the one with the most acceptable performance measure, either estimated C_{pk}, assembly risk or failure cost. In some cases, commercial confidence precludes the inclusion of detailed drawings of the components used in the analyses. CA has been used in industry in a number of different ways. Some of these are discussed below:

- *Variability prediction* – A key objective of the analysis is predicting, in the early stages of the product development process, the likely levels of out of tolerance variation when in production.
- *Tolerance stack analysis* – Tolerances on components that are assembled together to achieve an overall design tolerance across an assembly can be individually analysed, their potential variability predicted and their combined effect on the overall conformance determined. The analysis can be used to optimize the design through the explorations of alternative tolerances, processes and materials with the goal of minimizing the costs of non-conformance.
- *Evaluating and comparing designs* – CA may be used to evaluate alternative concepts or schemes that are generated to meet the specification requirements. It highlights the problem areas of the design, and calculates the failure cost of each, which acts as a measure of quality for the design. Since the vast majority of cost is built into a component in the early stages of the design process, it is advantageous to appraise the design as honestly and as soon as possible to justify choice in selecting a particular design scheme.

- *Requirements definition* – Where designers require tighter tolerances than normal, they must find out how this can be achieved, which secondary processes/process developments are needed and what special control action is necessary to give the required level of capability. This must be validated in some way. The variability risks analysis described previously is useful in this connection. It provides systematic questioning of a design regarding the important factors that drive variability. It estimates quantities that can be related to potential C_{pk} values, and identifies those areas where redesign effort is best focused. The results serve as a good basis for component supplier dialogue, communicating the requirements that lead to a process capable design.
- *Generation mode* – The process capability knowledge used in CA to facilitate an analysis can be used in the early stages of product design to generate process capable dimensional tolerances and surface roughness values when the material and geometry effects are minimal. Similarly, the assembly charts can be used to generate an improved assembly by avoiding features known to exhibit high risk.

7.2.6.3 Analysis procedure

The procedure is shown in flow chart form in Figure 7.2.25. In order to obtain the best results, the following points should be clearly understood before starting to evaluate a component's manufacturing or assembly situation:

- Read through this chapter carefully before attempting a design analysis.
- It is important that the product development team be familiar with the main capabilities and characteristics of the manufacturing and assembly methods selected, as well as the materials considered, in order to obtain the full benefit from the methodology. Technical and economic knowledge for some common manufacturing processes can also be found in the 1997 work of Swift and Booker (1997).
- The analysis requires the declaration of a sequence of assembly work, so familiarity with this supporting method is essential.
- When analysing a component or assembly process, complete all columns of the Variability Risks Results Table (see later for an example) and write additional notes and comments in the results table whenever possible. The table is a convenient means of recording the analysis for individual component manufacturing and assembly variability risks (q_m, q_a). It is recommended that the results table provided is used every time the analysis is applied to minimize possible errors.

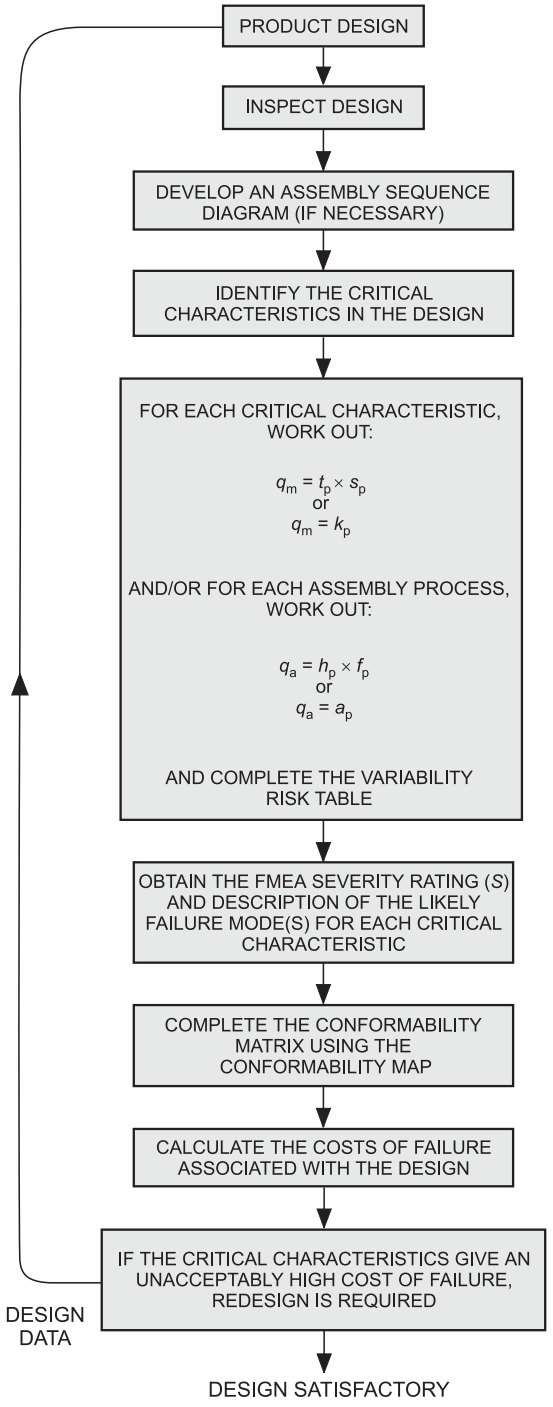


Fig. 7.2.25 CA procedure flow chart.

- The Conformability Matrix (see later for an example) primarily drives assessment of the variability effects. The Conformability Matrix requires the declaration of FMEA Severity Ratings and descriptions of the likely failure mode(s). It is helpful in this respect to have the results from a design FMEA for the product.

7.2.6.4 Example – component manufacturing variability risks analysis

Figure 7.2.26 shows the design details of a bracket, called a cover support leg. A critical characteristic is the distance from the hole centre to the opposite edge, dimension '*A*' on the bracket, which must be 50 ± 0.5 mm to effectively operate in an automated assembly machine. The assembly machine is used to produce the final product, of which the support leg is a key component. Failure to achieve this specification feeder jams, fastener insertation and securing.

Through an analysis using CA, high material and geometry variability risks were highlighted which could reduce the final capability of the characteristic. The component manufacturing variability risk, q_m, for the characteristic is calculated to be 9, this being determined from the adjusted tolerance using the process capability map for bending, as shown. This gives a predicted process capability index, $C_{pk} = 0.05$. In fact, the component was already in production when it was realized that the dimensions on most of the components were out of tolerance. A process capability analysis of the manufacturing process was conducted, and the results are shown in Figure 7.2.27. the calculated process was conducted, and the results are shown in Figure 7.2.27. The calculated process capability index was found to be $C_{pk} = 3.03$, which relates to approximately half the components falling out of specification, assuming a normal distribution.

$$C_{pk} = \frac{|\mu - L_n|}{3\sigma} \tag{7.2.14}$$

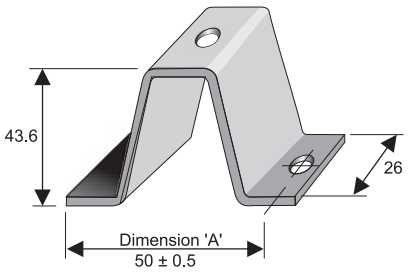
where μ is the mean of distribution; L_n the nearest tolerance limit; and σ the standard deviation Therefore:

$$C_{pk} = \frac{|49.575 - 49.5|}{3 \times 0.861} = 0.03$$

As a result of poor capability, a high level of machine downtime was experienced, and the company which manufactured the assembly machine were involved in an expensive legal dispute with their customer. Although a third party manufactured the cover support leg, liability, it was claimed, lay with the final assembly machine manufacturer. This resulted in serve commercial problems for the company, from which it never fully recovered.

An analysis using CA at an early design stage would have highlighted the risks associated with using a spring steel and long unsupported sections in the design, which were the main tolerance reducing variabilities. With references to the bending process capability map, the initial tolerance set at ±0.5 mm was just within acceptable limits at a nominal dimension of 50 mm. However,

Component name	Cover support leg
Drawing number	N/A
Material	0.6mm spring steel
Manufacturing process	Sheet metal bending
Characteristic description	Hole centre to edge
Characteristic dimension '*A*'	50 (mm)
Design tolerance	±0.5 (mm)
Surface roughness	- (m Ra)



Calculation of q_m

$m_p = 1.7 \times 1.5 = 2.55$

$g_p = 1 \times 1 \times 1 \times 1.7 \times 1 \times 1.1 = 1.87$

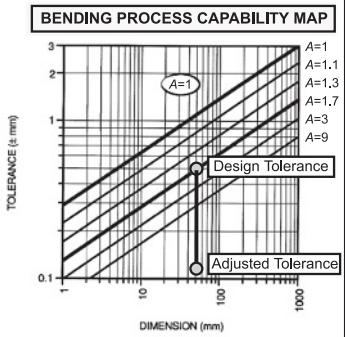
$$\text{Adjusted tolerance} = \frac{\text{Design tolerance}}{m_p \times g_p}$$

$$= \frac{\pm 0.5}{2.55 \times 1.87} = \pm 0.105$$

$t_p = 9 \times 1 = 9$

$s_p = 1 \times 1 = 1$

$q_m = t_p \times s_p = 9 \times 1 = 9$



Manufacturing variability risk, $q_m = 9$

Fig. 7.2.26 Component manufacturing variability risks analysis of a cover support leg.

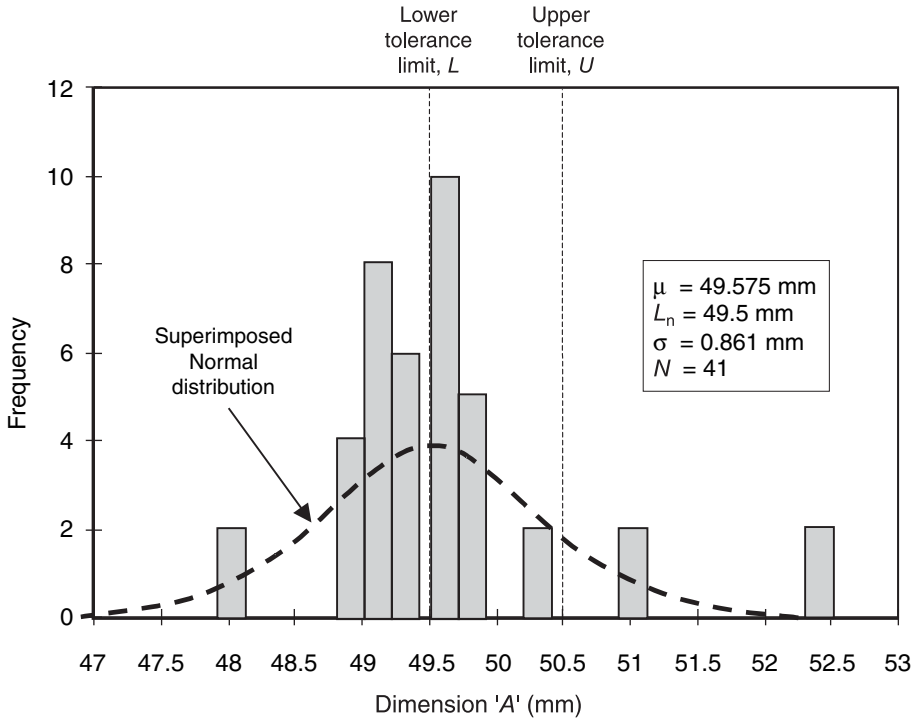


Fig. 7.2.27 Statistical process data for dimension '*A*' of the cover support leg.

the variability risks m_p and g_p decreased the probability of obtaining it substantially.

7.2.6.4.1 Completing a variability risks table

A variability risk table (as shown in Figure 7.2.28 for the cover support leg analysis above) is a more efficient and traceable way of presenting the results of the first part of the analysis. It catalogues all the important design information, such as the tolerance placed on the characteristics, the characteristic dimension itself, surface roughness value, and then allows the practitioner to input the results determined from the variability risks analysis, both manufacturing and assembly.

The documentation of the risks for each component and assembly operation follows the determination of the assembly sequence diagram, if appropriate, when the product consists of more than a single component. Every critical component characteristic or assembly stage is analysed through the use of the variability risks table. An assembly risk analysis will be performed during several of the case studies presented later.

7.2.6.5 Example – component assembly variability risks analysis

Similar to the calculation of q_m, it is possible to identify the variability risks associated with the assembly of components, q_a. In the following example shown in Figure 7.2.29, a cover and housing (with a captive nut) are fastened by a fixing bolt, which in turn is secured by a tab washer.

For example, for assembly operation number 4 for the fixing bolt we can determining the frisk indices h_p and f_p. For the handling process, assuming automatic assembly, the fixing bolt is supplied in a feeder and does not have characteristics which complicate handling. Therefore from the handling table in Figure 7.2.17, $h_p = 1.0$. Now assessing the fitting to process risk, f_p, from Figure 7.2.18:

A (cannot be assembled the wrong way)	$A = 1.0$
B (no positioning reliance to process)	$B = 1.0$
C (automatic screwing)	$C = 1.6$
D (straight line assembly from above)	$D = 1.2$
E (single process – one bolt in hole)	$E = 1.1$
F (no restricted access or vision)	$F = 1.0$
G (no alignment problems)	$G = 1.0$
H (no resistance to insertion)	$H = 1.0$

For the bolt fitting operation:

$$f_p = A \times B \times C \times D \times E \times F \times G \times H = 2.11$$

The component assembly variability risk is given by

$$q_a = h_p \times f_p$$

$$q_a = 1.0 \times 2.11 = 2.11$$

PRODUCT NAME: Cover Support Leg
PRODUCT CODE/ID: CSL1
PRODUCT QUANTITY: 50,000

SHEET No: 1 of 1
ANALYSIS REF: CSL - CA 1
ANALYSIS DATE: 1/1/99
ENGINEER: A.N.Other

Component/assembly process reference	Component/ assembly process description	Material/ process	$q_m - t_p.s_p$ (or $q_m - k_p$): Characteristic dimension (mm)	Design tolerance (mm)	Surface roughness (μm Ra)	Material process risk, m_p	Geometry process risk, g_p	Adjusted tolerance: Design tolerance / $m_p.g_p$	Tolerance process risk, t_p	Surface roughness process risk, s_p	Surface engineering process risk, k_p	$q_a - h_p.f_p$ (or $q_a - a_p$): Handling process risk, h_p	Indices for fitting process risk, f_p: A	B	C	D	E	F	G	H	Total fitting process risk, f_p	Additional process risk, a_p	Total risk (q_m or q_a)	Comments
c1	Cover Support Leg Dimension A	Sheet Spring Steel/Bending	120	±0.05	-	2.25	1.67	0.02	3	-													9	UNACCEPTABLE VALUE

Fig. 7.2.28 Variability risks table for the cover support leg.

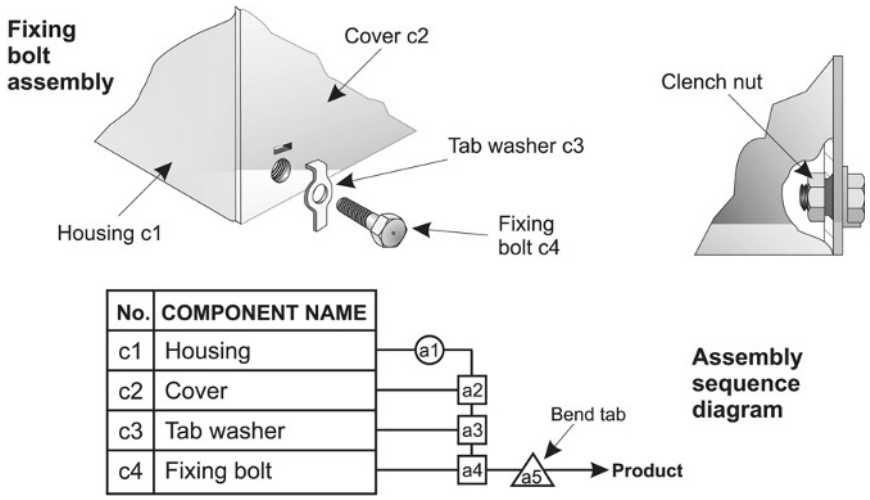


Fig. 7.2.29 Fixing bolt assembly and sequence of assembly.

Therefore,

$$q_a4 = 2.11$$

However, once the cover, tab washer and bolt are in place an additional process is carried out on the washer to bend the tab. Thus from the additional assembly process table for later mechanical deformation, the risk index, a_p, is

A (automated process)	$A = 1.0$
B (medium skill level)	$B = 1.3$
C (poor access)	$C = 1.6$
D (pressure used for deformation)	$D = 1.2$
E (no heat applied)	$E = 1.0$
F (one stage operation)	$F = 1.0$

For bending the tab washer operation, the Additional Assembly Variability Risk is given by

$$a_p = A \times B \times C \times D \times E \times F$$
$$a_p = 2.5$$

Therefore,

$$q_a 5 = 2.5$$

A completed variability risks result table for the fixing bolt assembly is shown in Figure 7.2.30, highlighting the assembly variability risks only. The risk indices can then be entered on the assembly sequence diagram as shown in Figure 7.2.31. It is evident that insertion of the fixing bolt and bending of the tab washer operations will potentially be problematic on assembly, scoring assembly risk values greater than 2.

The previous sections have illustrated the use of the various process risk charts and tables to obtain variability indices associated with the manufacture and assembly of products. It is important to be systematic with the application of the methodology and the recording of any product analyses, especially as products often contain many parts. It is also important to first declare a sequence of assembly for the individual components before proceeding with an analysis.

7.2.6.6 Completing the Conformability Matrix

The final part of the analysis is based around the completion of a Conformability Matrix relating variability risk indices for component manufacturing/assembly processes to potential failure modes, their severity and the costs of failure. The final results of an analysis are best displayed in the Conformability Matrix to provide a traceable record of the costs of failure and how these costs are related to the conformance problems through the design decisions made. The symbols, shown in Figure 7.2.32, are placed in the nodes of the Conformability Matrix, and represent the levels of design acceptability obtained with reference to the Conformability Map given in Figure 7.2.22.

The link with FMEA brings into play the additional dimension of potential variability into the assessment of the failure modes and the effects on the customer. The Conformability Matrix also highlights those 'bought-in' components and/or assemblies that have been analysed and found to have conformance problems and require further communication with the supplier. This will ultimately improve the supplier development process by highlighting problems up front.

Once the variability risks, q_m and q_a, have been calculated, the link with the particular failure mode(s) from an FMEA for each critical characteristic is made. However, determining this link, if not already evident, can be the most subjective part of the analysis and should ideally be a team-based activity. There may be many component characteristics and failure modes in a product and the matrix must be used to methodically work through this part of the analysis. Past failure data on similar products may be useful in this respect, highlighting those areas of the product that are most affected by variation. Variation in fit, performance or service life is of particular interest since controlling these kinds of variation is most closely allied with quality and reliability.

For each q_m and q_a risk value and the Severity Rating (S), a level of design acceptability is determined from where these values intersect on the Conformability Map. The symbols, relating to the levels of design acceptability, are then placed in the nodes of the Conformability Matrix for each variability risk which the failure mode is directly dependent on for the failure to occur. Once the level of design acceptability has been determined, it can then be written on the Conformability Matrix in the 'Comments' section. C_{pk} values predicted or comments for suppliers can be added too, although predicted C_{pk} values can also be written in the variability risks results table.

7.2.6.6.1 Example – determining the failure costs for product design

We will now consider calculating the potential costs of failure in more detail for the cover support leg shown earlier. The process for calculating the failure costs for a component is as follows:

- Determine the value of q_m or q_a
- Obtain an FMEA Severity Rating (S)
- Estimate the number of components to be produced (N)
- Estimate the component cost (Pc).

For example, the characteristic dimension '*A*' on the cover support leg was critical to the success of the automated assembly process, the potential failure mode being a major disruption to the production line. An FMEA Severity Rating (S) = 8 is allocated. The component cost, Pc = £5.93 and the number planned to be produced *per annum*, N = 50,000.

The characteristic was analysed using CA and q_m was found to be 9. The values of $q_m = 9$ and $S = 8$ are found to intersect on the Conformability Map above the 10% isocost line. (If they had intersected between two isocost lines, the final isocost value is found by interpolation.) If there is more than one critical characteristic on the

PRODUCT NAME: Fixing Bolt Assembly
PRODUCT CODE/ID: FBA1
PRODUCT QUANTITY: 100,000

SHEET No.: 1 of 1
ANALYSIS REF.: FBA - CA 1
ANALYSIS DATE: 1/1/99
ENGINEER: A.N.Other

Component/assembly process reference	Component/ assembly process description	Material/ process	$q_m = t_p.s_p$ (or $q_m = k_p$): Characteristic dimension (mm)	Design tolerance (mm)	Surface roughness (m Ra)	Material process risk, m_p	Geometry process risk, g_p	Adjusted tolerance = Design tolerance / $m_p.g_p$	Tolerance process risk, t_p	Surface roughness process risk, s_p	Surface eng. process risk, k_p	$q_a = h_p.f_p$ (or $q_a = a_p$): Handling process risk, h_p	Indices for fitting process risk, $f_.$ A	B	C	D	E	F	G	H	Total fitting process risk, f_p	Additional process risk, a_p	**Total risk (q_m or q_a)**	Comments
a1	c1 → WH											1									1		1	ACCEPTABLE DESIGN
a2	c2 → a1											1	1	1	1.2	1	1	1	1	1	1.2		1.2	ACCEPTABLE DESIGN
a3	c3 → a2											1.2	1	1	1	1.2	1	1	1	1	1.2		1.44	ACCEPTABLE DESIGN
a4	c4 → a3											1	1	1	1.6	1.2	1.1	1	1	1	2.11		2.11	SPECIAL CONTROL
a5	Bend Washer Tab																					2.5	2.5	UNACCEPTABLE DESIGN

Fig. 7.2.30 Variability risks table for the fixing bolt assembly.

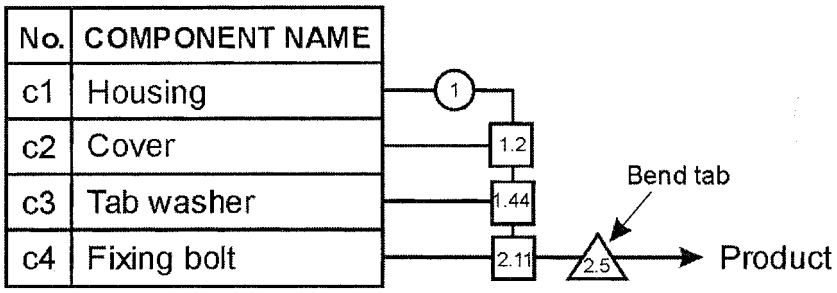


Fig. 7.2.31 Revised sequence of assembly for the fixing bolt.

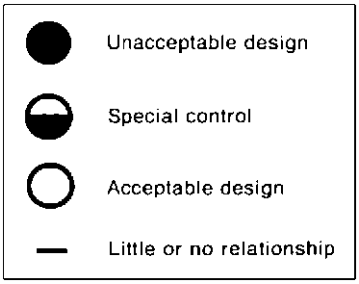


Fig. 7.2.32 Conformability Matrix symbols and their quantification.

component, then the isocosts are added to give a total isocost to be used in Equation (7.2.15). The total failure cost is determined from

$$\text{Total failure cost} = \frac{\text{isocost}(\%) \times N \times Pc}{100}$$

$$\text{Total failure cost} = \frac{10 \times 50,000 \times 5.93}{100} = £29,650 \quad (7.2.15)$$

This figure is of course an estimate of lost profit and may even be conservative, but it clearly shows that the designer has a significant role in reducing the high costs of failure reported by many manufacturing companies. The results are repeated in the Conformability Matrix in Figure 7.2.33.

7.2.7 Case studies

7.2.7.1 Electronic power assisted steering hub design

Under this heading, a flexible hub design for an automotive steering unit is analysed. The application of CA resulted from the requirement to explain to a customer how dimensional characteristics on the product, identified as safety critical, could be produced capably. A key component in this respect is the hub. The component is made by injection moulding, the material being unfilled polybutylene terephthalate (PBT) plastic. The moulding process was selected for its ability to integrate a number of functional elements into a single piece and reduce assembly costs. The design outline of the hub is shown in Figure 7.2.34(a) together with an optical plate that is mounted on it.

Figure 7.2.34(b) shows a line from the design FMEA related to the plastic moulded hub. It gives the component function, the potential failure mode, the potential effects and potential causes of failure. In addition, the columns of current controls, occurrence (O), severity (S), detectability (D) ratings and associated risk priority number (RPN) have been completed. A high severity rating was given ($S = 8$) since faulty positional readings due to tolerance variation could cause loss of car control and driver injury. The detectability was rated at $D = 8$ because complex inspection processes would be required. The possibility of occurrence was provisionally estimated at $O = 4$, giving an RPN equal to 256. The design FMEA specifies compatible dimensioning as the current control to avoid failure, and it was this aspect of the design that needed to be explored further. A quotation from a supplier had been received for volume production of the component.

Some detail on the analysis of the hub design is given below. The hub performed several functions in the controller and therefore carried several critical characteristics. The positional tolerance of the recesses to accommodate a system of location pegs needed to be close. Faces on the hub for mounting the optical plate required precise positioning to provide the necessary spacing between two optical grids (one mounted on the hub and the other carried on a torsion shaft). The depth of the moulded recesses needed to be controlled as they were part of a tolerance chain. It is important to note that the recesses and faces were in different planes and the depth was across a mould parting line.

The positional tolerance on a 10 mm dimension was ± 0.1 mm (tol_1), providing an angular position of 0.6°. Additionally, the widths of the recesses needed to be held to 4 ± 0.08 mm (tol_2), the dimensional tolerance on the faces to 1 ± 0.07 mm (tol_3) and the depth to 10 ± 0.12 mm (tol_4). Also, the thin sections of the hub gave

PRODUCT NAME: Cover Support Leg
PRODUCT CODE/ID: CSL1
PRODUCT QUANTITY: 50,000

SHEET No.: 1 of 1
ANALYSIS REF.: CSL - CA 1
ANALYSIS DATE: 1/1/99
ENGINEER: A.N.Other

Component/assembly process reference	Component/ assembly process description	Total risk (q_m or q_n)	Failure Mode Description and FMEA Severity Rating (S)											Comments (including action for suppliers)
			Assembly Line Disrupted											
			8											
c1	Cover Support Leg Dimension A	8	●											UNACCEPTABLE DESIGN
Total Failure Mode Isocost %			10											TOTAL FAILURE COST
Total Failure Mode Cost			£30K*											£30K

* Number of units = 50 000
Product cost (Pc) = £5.93
Total failure mode isocost (%) = 10

$$\text{Therefore, total failure mode cost} = \frac{10 \times 50\,000 \times 5.93}{100} = £29\,650$$

Fig. 7.2.33 Conformability Matrix example.

two geometrical concerns as these vanes were on the limits of plastic flow and distortion was likely on cooling.

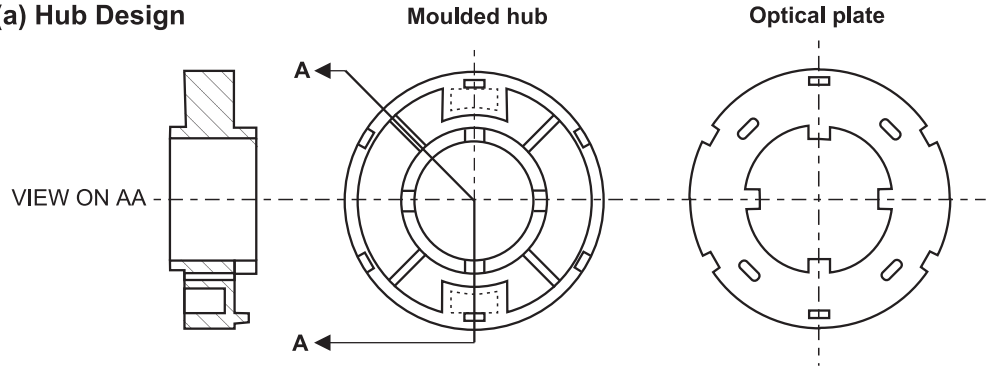
The results of the analyses carried out by the business on the hub are given in the variability risks table shown in Figure 7.2.34(c). The four critical characteristics described above were examined. The positional tolerance (tol_1) set across the thin vanes resulted in high geometry to process risk (g_p) and gave $q_m = 9$. This equates to an out of control C_{pk}, as does the q_m value for the recess depth across the mould parting line (tol_4) which came to 8. The q_m scores for the characteristics tol_2 and tol_3 suggest initially that the process will be in control giving estimated C_{pk} values of 1.33 and 1.75, respectively.

Following the completion of the variability risks table, a Conformability Matrix was produced. This was used to relate the failure modes and their severity coming out of the design FMEA to the results of the component manufacturing variability risk analysis. The portion of the matrix concerned with the moulded hub can be found in Figure 7.2.34(d) and was completed using the Conformability Map.

It is evident that the two characteristics described earlier as being out of control, tol_1 and tol_4, give costs of failure greater than 10%. Also, the characteristics tol_2 and tol_3 which may have been regarded as having acceptable C_{pk} values are shown to have costs of failure of greater than 10% and 0.2%, respectively, due to the high severity rating (S) = 8 for the potential failure modes in question. Note that two additional failure modes are also illustrated.

The analysis indicated that the conformance problems associated with the hub design had a cost of failure of more than 30%. This would represent at the annual production quantity required and target selling price, a loss to the business of several million pounds. As a result of the study the business had further detailed discussions with their suppliers and not surprisingly it turned out that the supplier would only be prepared to stand by its original quotation provided the tolerances on the hub, discussed above, were opened up considerably (more than 50%). Subsequently, this result supported the adoption of another more capable design scheme.

(a) Hub Design



(b) Design FMEA

Product function	Potential failure mode	Potential effects of failure	Potential causes of failure	Existing conditions: Current controls	Occurrence (*O*)	Severity (*S*)	Detectability (*D*)	RPN = *O*×*S*×*D*
LOCATE AND SUPPORT OPTICAL PLATE	MISALIGNMENT OF OPTICAL PLATE	SPURIOUS POSITIONAL SIGNALS	FACES OUT OF TOLERANCE	COMPATIBLE DIMENSION OF MATING PARTS	4	8	8	256

(c) Variability risks

Component/assembly process reference	Component/assembly process description	Material process	$q_m = t_p.s_p$ (or $q_m = k_p$): Characteristic dimension (mm)	Design tolerance (mm)	Surface roughness (μm Ra)	Material process risk, m_p	Geometry process risk, g_p	Adjusted tolerance = Design tolerance / (m_p·g_p)	Tolerance process risk, t_p	Surface roughness process risk, s_p	Surface eng. process risk, k_p	$q_a = h_p.f_p$ (or $q_a = a_p$): Handling process risk, h_p	Indices for fitting process risk, f_p: A	B	C	D	F	G	H	Total fitting process risk, f_p	Additional process risk, a_p	Total risk (q_m or q_a)	Comments
c1	HUB - tol_1	INJ. MOULDED UNFILLED PBT	10	±0.1	1.6	1	3.18	±0.031	9	1													9	UNACCEPTABLE DESIGN
	- tol2		4	±0.08	1.6	1	1.1	±0.073	1.7	1													1.7	$C_{pk} \approx 1.33$
	- tol_3		1	±0.07	1.6	1	1.1	±0.064	1.5	1													1.5	$C_{pk} \approx 1.75$
	- tol_4		10	±0.12	1.6	1	1.87	±0.064	8	1													8	UNACCEPTABLE DESIGN

NOT APPLICABLE TO THIS ANALYSIS

(d) Conformability matrix

Component/assembly process reference	Component/assembly process description	Total risk (q_m or q_a)	Failure Mode Description and FMEA Severity Rating (*S*): MISALIGNMENT OF OPTICAL PLATES (8)	RELATIVE MOVEMENT OF PLATES (8)	CONTACT OF PLATES (8)	Comments (including action for suppliers)
c1	HUB - tol_1	9	●	—	—	UNACCEPTABLE DESIGN
	- tol2	1.7	—	●	—	UNACCEPTABLE DESIGN
	- tol_3	1.5	—	—	◒	SPECIAL CONTROL
	- tol_4	8	—	—	●	UNACCEPTABLE DESIGN
Total Failure Mode Isocost (%)			10	10	10.05	TOTAL FAILURE COST
Total Failure Mode Cost						30.05%

Fig. 7.2.34 Hub analysis results.

7.2.7.2 Solenoid security cover

This case study concerns the initial design and redesign of a security cover assembly for a solenoid. The analysis only focuses on those critical aspects of the assembly of the product that must be addressed to meet the requirement that the electronics inside the unit are sealed from the outside environment. An FMEA Severity Rating (S) for the assembly was determined as $S = 5$, a warranty return if failure is experienced.

7.2.7.2.1 Cover assembly initial design

The initial design of the cover assembly as shown in Figure 7.2.35 uses an O-ring to seal the electronics against any contamination. Concerns were raised about three main aspects of the assembly, these being:

- The compression of the O-ring may work against the needs of an adhesive cure on final assembly with an end unit.
- There is a risk that the O-ring will not maintain its proper orientation in the cover recess during subsequent assembly processes, and therefore may not be correctly positioned on final assembly. Restricted vision of the inside of the cover is the key problem here.
- The wire cable may present problems using either manual or automated assembly.

The analysis in Figure 7.2.35 shows that there is a high risk of non-conformance for the insertion of the frame into the cover, the process relying on the position of the O-ring being maintained (operation a8). The situation is complicated by the restriction of vision during O-ring placement, and this is reflected in the analysis. Using the Conformability Map, it is possible to calculate the potential failure costs for this design scheme in meeting the sealing integrity requirement, as documented in the Conformability Matrix. The final failure cost is calculated to be £805,000. This potential failure cost for this single failure mode is far too high, representing over 10% of the total product cost. A more reasonable target value would be less than 1%. An alternative design scheme should be developed, focusing on reducing the risks of the final assembly operation to reduce the potential for non-conformance as highlighted by the analysis.

7.2.7.2.2 Cover assembly redesign

Unfortunately, the design of the wire could not be changed to a more simple arrangement, for example using a spade connector integrated with a recess for the O-ring. The wire is part of the customer's requirements and will inevitably present problems using either manual or automatic assembly operations. Looking to the O-ring, a better design would be to eliminate it altogether and integrate the seal with the wire as shown in Figure 7.2.36. The wire is then positively located with the seal in the cover hole. Clearly, the risks associated with the cover assembly have been reduced following the elimination of one positionally unstable component and its integration with another. Again, with an FMEA (S) = 5, and referring to the Conformability Map, isocosts for each assembly variability risk can be evaluated and the total failure cost is calculated to be £7000.

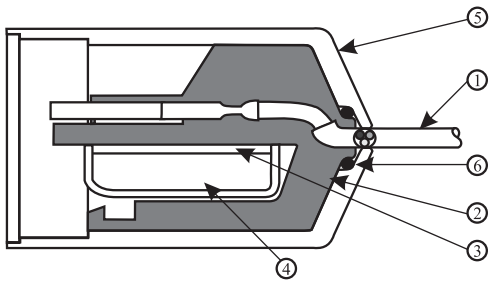
Comparing this value with the initial design's high potential cost of failure, it is evident that a major design fault in the cover has been eliminated although the assembly process must remain within special control. Subsequently, the redesign solution was chosen for further design development.

7.2.7.3 Telescopic lever assembly

Consider the telescopic lever assembly, Design A in Figure 7.2.37, which is part of a stretcher, and hence safety critical. The assembly has an FMEA Severity Rating (S) = 8, and is used in a product having a cost of £150. It is estimated that 5000 units are produced *per annum*. The assembly is subjected to bending in operation, with the maximum bending stresses occurring at a point on the main tube corresponding to the stop ring recess. In order to provide additional support, a reinforcing tube is positioned as shown. It is crucial that the tube is placed where it is, since fracture of the telescopic lever may result in injury to users and third parties. Figure 7.2.37 shows part of the results from the Conformability Matrix for Design A. At each node in the matrix, consideration has been given to the effect of the component or assembly variability risk, represented by q_m or q_a, on the failure mode in question determined from an FMEA.

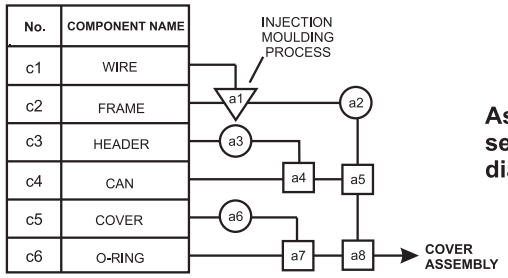
While the design is satisfactory from a design for strength point of view, the analysis highlights a number of areas where potential variability and failure severity combine to make the risks unacceptably high. For example, there are no design features which ensure the positioning of the reinforcement tube in the assembly (assembly process a3 in Figure 7.2.37). There are also several critical component tolerances which need to be controlled if its position is to be maintained in service, such as the inner diameter of the main tube, c1, the outside diameter of the reinforcement tube, c2, and the application of the adhesive, a2.

The conclusion from the analysis is that the assembly should be redesigned. This is further justified by calculating the potential costs of failure for the assembly. If this design of telescopic lever assembly fractured in service, user injury, high losses, including legal costs, could



Cover assembly

No.	COMPONENT NAME
c1	WIRE
c2	FRAME
c3	HEADER
c4	CAN
c5	COVER
c6	O-RING



Assembly sequence diagram

Component/assembly process reference	Component assembly process description	Total risk (q_m or q_a)	Failure Mode Description and FMEA Severity Rating (S)						Comments (including action for suppliers)
			SEALING INTEGRITY						
			5						
c2 a1	FRAME MOULDING	1.5	○						ACCEPTABLE DESIGN
a4	CAN TO HEADER	2.1	—						
a5	CAN/HEADER TO FRAME	2.4	—						
a7	O-RING INTO COVER	1.94	◒						SPECIAL CONTROL
a8	FRAME INTO COVER	5.1	●						UNACCEPTABLE DESIGN
Total Failure Mode Isocost %			10.51						TOTAL FAILURE COST
Total Failure Mode Cost			£805K*						£805K

*Number of units = 1 000 000

Product cost (Pc) = £7.66

Total failure mode isocost (%) = 0.01 + 0.5 + 10 = 10.51

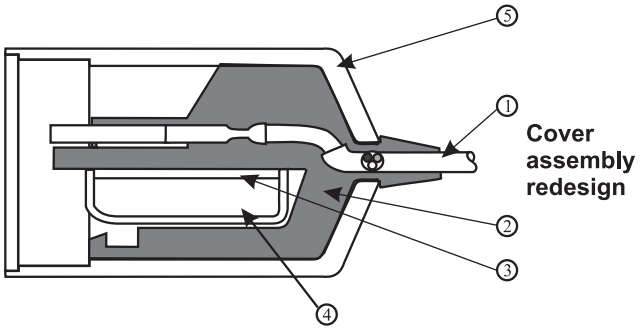
Therefore, total failure mode cost = $\frac{10.51 \times 1\,000\,000 \times 7.66}{100}$ = £805 066

Conformability Matrix

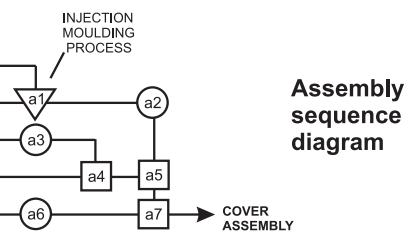
Fig. 7.2.35 Cover assembly design analysis.

be incurred. A cost of failure of £257,000 was calculated from the analysis, which is far too high representing more than 36% of annual revenue from the product. This figure was calculated by summing the isocosts for each characteristic/assembly process whose variability risks potentially contribute to each failure mode, and then multiplying the total failure mode isocost (%) by the product cost and number of items produced. The calculation of the costs for the second failure mode type (reinforcing tube moves out of position) on Design A is shown in detail in Figure 7.2.37.

Failure of this design in service did in fact result in user injury. High losses of the order of those calculated above for the particular failure mode, including legal costs, were incurred. A number of alternative designs are possible, and one which does not involve the above problems



No.	COMPONENT NAME
c1	WIRE
c2	FRAME
c3	HEADER
c4	CAN
c5	COVER



Assembly sequence diagram

Component/assembly process reference		Component assembly process description	Total risk (q_m or q_a)	Failure Mode Description and FMEA Severity Rating (S)						Comments (including action for suppliers)
				SEALING INTEGRITY						
				5						
c2	a1	FRAME MOULDING	1.5	○						ACCEPTABLE DESIGN
	a4	CAN TO HEADER	2.1	—						
	a5	CAN/HEADER TO FRAME	2.4	—						
	a7	FRAME INTO COVER	1.82	◒						SPECIAL CONTROL
Total Failure Mode Isocost %				0.09						TOTAL FAILURE COST
Total Failure Mode Cost				£6,894						£6,894

Conformability Matrix

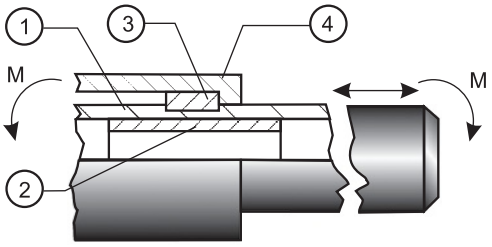
Fig. 7.2.36 Cover assembly redesign analysis.

is included with its Conformability Matrix in Figure 7.2.38. The costs of failure for this Design B were subsequently reduced to a negligible amount.

The above example demonstrates the use of CA in supporting the identification of manufacturing and assembly problems before production commences, but more importantly safety problems can be systematically identified and the potential for harm eliminated before the product is in use by the customer. Given the huge losses that are associated with safety critical products when they fail, considerations of this type must be on the agenda of all manufacturing companies, especially when the product has a high degree of interaction with the user.

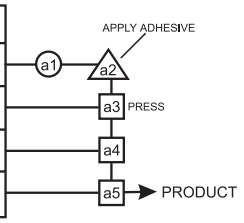
7.2.7.4 Solenoid end assembly

The following case study determines the manufacturing and assembly variability risks for a solenoid end assembly design, shown in Figure 7.2.39, and projects the potential costs of failure associated with the capability of an assembly tolerance stack. The solenoid is to operate as a fuel cut-off device in a vehicle, operated when a signal is received from the ignition. The signal allows current to



DESIGN A

No.	COMPONENT NAME
c1	MAIN TUBE
c2	REINFORCING TUBE
c3	STOP RING (SPLIT)
c4	OUTER TUBE



ASSEMBLY SEQUENCE DIAGRAM

Component/assembly process reference	Component/ assembly process description	Total risk (q_m or q_a)	Failure Mode Description and FMEA Severity Rating (S)						Comments (including action for suppliers)
			REINFORCING TUBE FITTED IN THE WRONG PLACE	REINFORCING TUBE MOVES OUT OF POSITION	DEPTH GEOMETRY OF STOP RING RECESS	STOP RING MOVES OUT OF POSITION			
			8	8	8	8			
c1	MAIN TUBE I.D.	1.6	–	●	–	–			UNACCEPTABLE DESIGN
c1'	MAIN TUBE RECESS	1.6	–	–	●	–			UNACCEPTABLE DESIGN
a1	c1 → WH	1	○	–	–	–			ACCEPTABLE DESIGN
a2	APPLY GLUE TO TUBE BORE	2.4	–	●	–	–			UNACCEPTABLE DESIGN
c2	REINFORCING TUBE	1.5	–	◒	–	–			SPECIAL CONTROL
a3	c2 → a2	4.5	●	–	–	–			UNACCEPTABLE DESIGN
c3	STOP RING (SPLIT)	1.3	–	–	–	○			ACCEPTABLE DESIGN
a4	c3 → a3	2	–	–	–	●			UNACCEPTABLE DESIGN
c4	OUTER TUBE	1	–	–	–	–			
a5	c4 → a4	1	–	–	–	–			
Total Failure Mode Isocost (%)			10.01	12.05	2	10.01			TOTAL FAILURE COST
Total Failure Mode Cost			£75.8K	£90.4K*	£15K	£75.8K			£257K

*Sample calculation of failure costs:

Number of units = 5000

Product cost (Pc) = £150

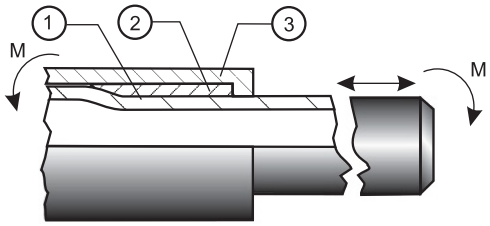
Total failure mode isocost (%) = 2 + 10 + 0.05 = 12.05

Therefore, total failure cost = $\frac{12.05 \times 5000 \times 150}{100}$ = £90 375

CONFORMABILITY MATRIX FOR DESIGN A

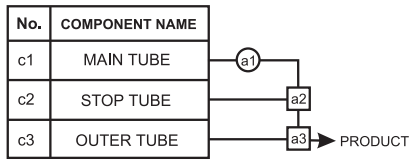
Fig. 7.2.37 Telescopic lever assembly analysis for Design A.

flow to the inductor coil which then withdraws the plunger seal and allows the fuel to flow. The solenoid assembly is to be screwed into an engine block at the fuel port in a counterbored hole. An important requirement is that the plunger displacement from the engine block face through the solenoid tolerance stack to the plunger seal face must be within a tolerance of ±0.2 mm. If this requirement is not met, fuel low restriction could occur, this being the main failure mode. The product will be in the warranty return category as it has little effect on user safety if it fails in service, which relates to an FMEA Severity Rating (S) = 5. The product cost is £7.66 and it is estimated that one million units will be manufactured in total.



DESIGN B

No.	COMPONENT NAME
c1	MAIN TUBE
c2	STOP TUBE
c3	OUTER TUBE



ASSEMBLY SEQUENCE DIAGRAM

Component/assembly process reference	Component/assembly process description	Total risk (q_m or q_a)	Failure Mode Description and FMEA Severity Rating (S): STOP TUBE NOT FITTED						Comments (including action for suppliers)
			8						
c1	MAIN TUBE	1.5	–						
a1	c1 → WH	1	◯						ACCEPTABLE DESIGN
c2	STOP TUBE	1.3	–						
a2	c2 → a1	1.3	◯						
c3	OUTER TUBE	1	–						ACCEPTABLE DESIGN
a3	c3 → a2	1	–						
Total Failure Mode Isocost (%)			0.02						TOTAL FAILURE COST
Total Failure Mode Cost			£150						£150

CONFORMABILITY MATRIX FOR DESIGN B

Fig. 7.2.38 Telescopic lever assembly analysis for Design B.

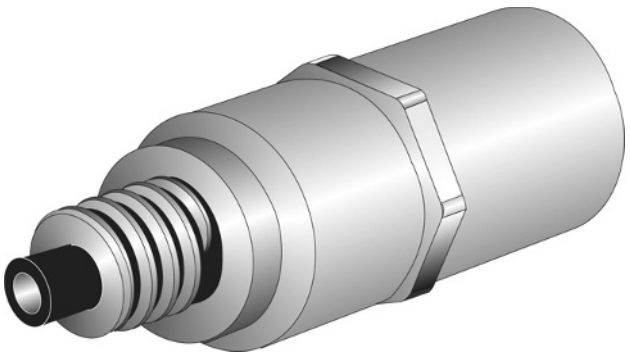


Fig. 7.2.39 Solenoid end assembly.

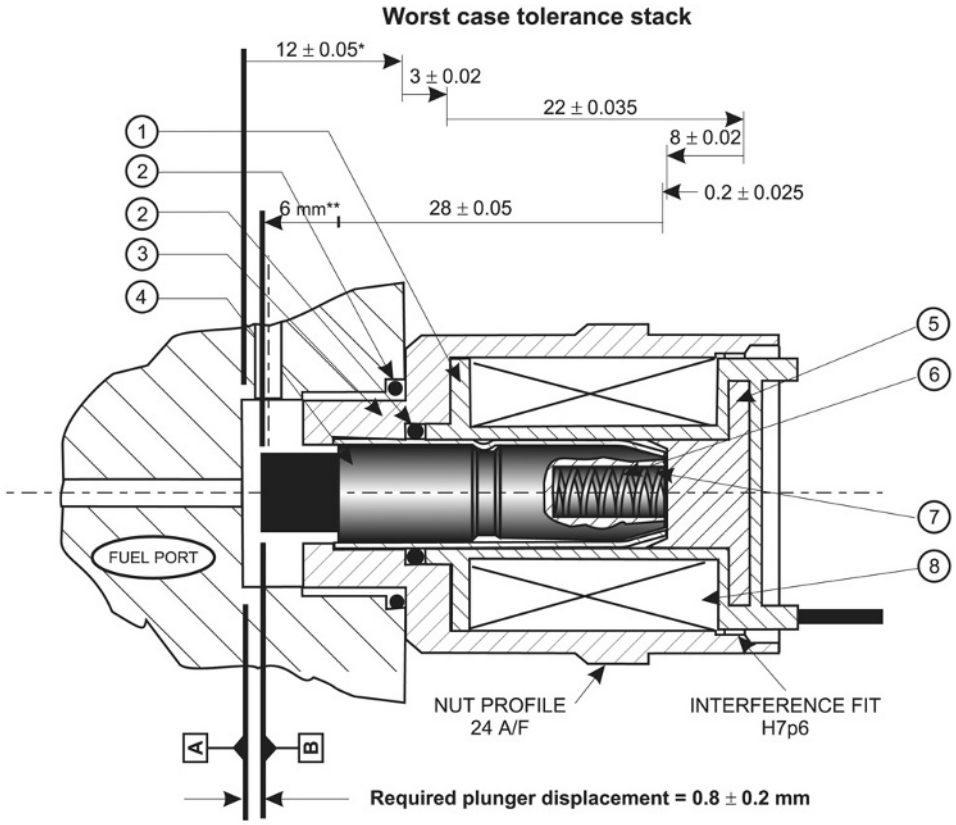
7.2.7.4.1 Solenoid end assembly initial design

The initial design is analysed using CA at a component level for their combined ability to achieve the important customer requirement, this being the tolerance of ±0.2 mm for the plunger displacement. Only those characteristics involved in the tolerance stack are analysed. The 'worst case' tolerance stack model is used as directed by the customer. This model assumes that each component tolerance is at its maximum or minimum limit and that the sum of these equals the assembly tolerance, given by Equation (7.2.16):

$$\sum_{i=1}^{n} t_i \leq t_a \tag{7.2.16}$$

where t_i is the bilateral tolerance for ith component characteristic; t_a the bilateral tolerance for assembly stack.

Figure 7.2.40 shows the initial detailed design including the tolerances required on each component in the stack to achieve the ±0.2 mm assembly tolerance, t_a (not included is the dimensional tolerance on the fuel port block of 12 ± 0.05 mm which is set by the supplier). Also shown is a table describing the process used to manufacture each component and an assembly sequence diagram is given in Figure 7.2.41. Referring back to Figure 7.2.40, the tolerance stack starts at face A on the fuel port and accumulates through the individual components to face B on the plunger seal.



PART No.	DESCRIPTION	MATERIAL	PROCESS
1	BOBBIN	PBT(30% FILL)	INJECTION MOULDED WITH ITEM 5
2	O-RING	RUBBER	–
3	BODY	COLD FORMING STEEL	IMPACT EXTRUSION
4	PLUNGER	STEEL/SILICONE RUBBER	M/C (MOULDED SILICONE RUBBER END SEAL)
5	MAGNETIC POLE	COLD FORMING STEEL	IMPACT EXTRUSION
6	SPRING	STEEL	–
7	TUBE	BRASS	ROLLED/DEEP DRAWN
8	COIL	–	–

* Tolerance fixed by supplier
** The 6 mm end seal dimension will be used in the analysis, not 28 mm

Fig. 7.2.40 Solenoid end assembly initial design.

The body is impact extruded from a cold forming steel. The characteristic dimension to be analysed in the tolerance stack is the base thickness of 3 mm (on a Ø20 mm bore) and this dimension has been assigned a tolerance of ±0.02 mm.

Following the tolerance stack through the end assembly, the bobbin dimension of 22 mm from the outside face to the back face of the magnetic pole is analysed next. This characteristic dimension does not include the tolerance on the impact extruded pole. The pole is to be moulded into the bobbin and the pole face is considered to be part of a mould-related dimension. The bobbin is injection moulded using 30% filled PBT. The tolerance assigned to the bobbin dimension is ±0.035 mm.

The pole has a characteristic dimension of 8 mm from the rear of the bobbin to the recess face and has a tolerance assigned to it of ±0.02 mm. From the pole recess face, the tube base tolerance is the last component to make up the tolerance stack. The brass tube has been given a dimensional tolerance on its base of 0.2 ± 0.025 mm. Note, the dimensional tolerance on the plunger is

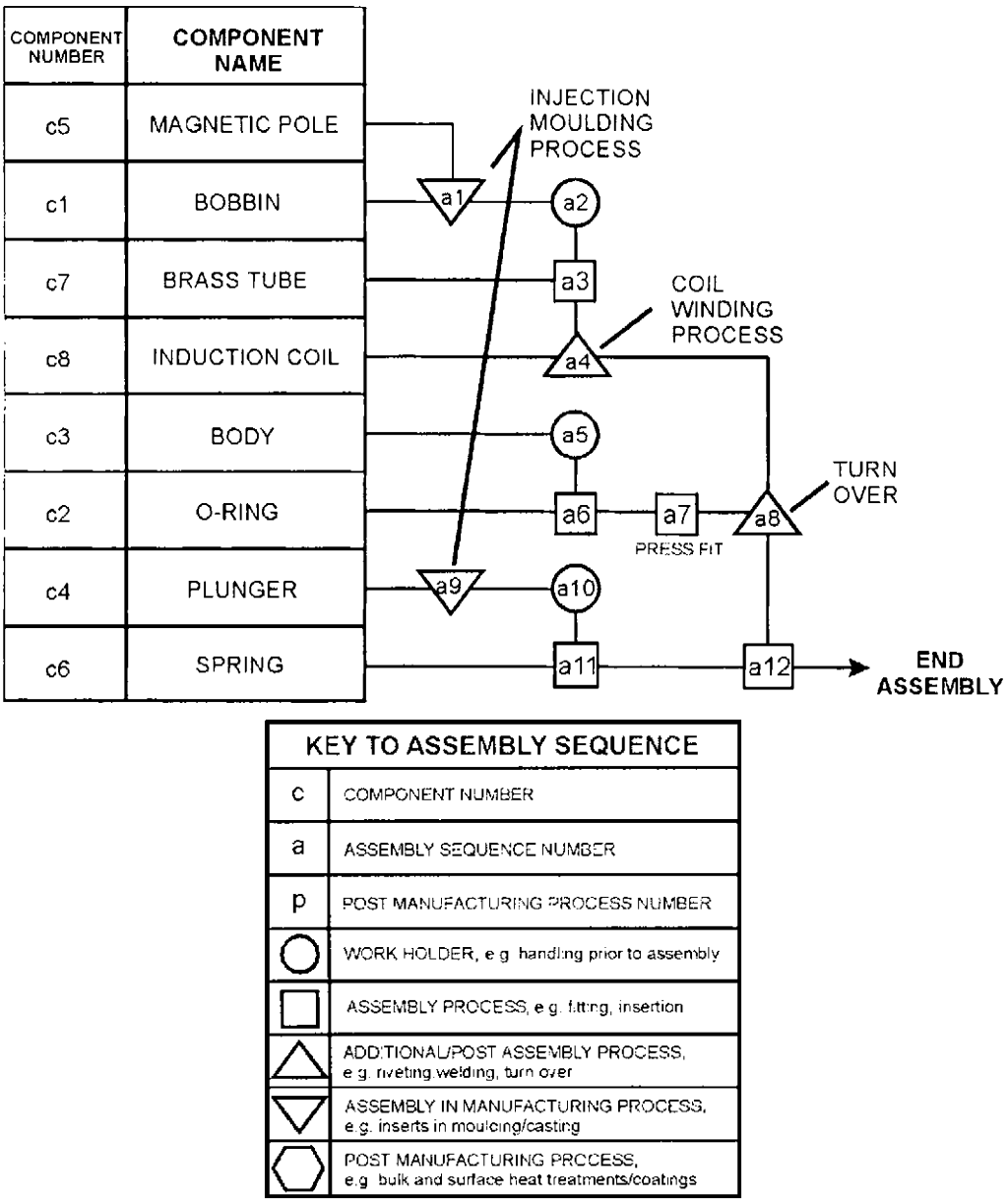


Fig. 7.2.41 Assembly sequence diagram for the solenoid end assembly design.

28 ± 0.05 mm, but the analysis will concentrate on the silicone rubber seal length of 6 mm because this is moulded onto the plunger and again is a mould-related dimension.

Figure 7.2.42 shows the variability risks analysis based on the tolerances assigned to meet the ±0.2 mm tolerance for the assembly. Given that an FMEA Severity Rating (S) = 5 has been determined, which relates to a 'definite return to manufacturer', both impact extruded components are in the unacceptable design region, as well as the bobbin and plunger end seal as shown on the Conformability Matrix in Figure 7.2.43. The tolerance for the brass tube base thickness has no risk and is an acceptable design.

The associated cost of failure for the solenoid end assembly is calculated to be over £3 million for a product cost of £7.66 and production volume of one million units. This figure is for the tolerance stack failure mode alone as this is most important to the customer. Although the assembly variability risks are analysed, they are not taken into account in the final costs of failure. In conclusion, the process capabilities of several characteristics in this tolerance stack are inadequate and will not meet the customer's requirements consistently.

				$q_m = t_p.s_p$ (or $q_m = k_p$)									$q_a = h_p.f_p$ (or $q_a = a_p$)											
Component/assembly process reference		Component/ assembly process descrption	Material/ process	Characteristic dimension (mm)	Design tolerance (mm)	Surface roughness (μm Ra)	Material process risk, m_p	Geometry process risk, g_p	Adjusted tolerance / Design tolerance = $m_p.g_p$	Tolerance process risk, t_p	Surface roughness process risk, sp	Surface eng. Process risk, k_p	Handling process risk, h_p	Indices for fitting process risk, f_p								Total fitting process risk, f_p	Additional process risk, a_p	**Total risk (q_m or q_a)**	Comments
														A	B	C	D	E	F	G	H				
c5		MAGNETIC POLE	IMPACT EXTRUDED COLD FORMING STEEL	8	±0.02	0.8	1.2	1.7	±0.01	9	1													9	UNACCEPTABLE DESIGN
c1	a1	BOBBIN	PBT (30% FILL) INJECTION MOULD WITH INSERT c5	22	±0.035	–	1	1	±0.035	9	1													9	UNACCEPTABLE DESIGN
		a1 → WH											1									1		1	ACCEPTABLE DESIGN
c7		TUBE	ROLLED/ DEEP DRAWN BRASS	0.2	±0.025	1.6	1	1	±0.025	1	1													1	ACCEPTABLE DESIGN
	a2	c7 → a2	TUBE INSERTED INTO BOBBIN I/D										1	1	1	1.1	1	1	1	1	1.4	1.54		1.54	ACCEPTABLE DESIGN
	a4		COIL WINDING PROCESS																						
c3		BODY	IMPACT EXTRUDED COLD FORMING STEEL	3 (Ø20)	±0.02	0.8	1.2	1.1	±0.015	9	1													9	UNACCEPTABLE DESIGN
	a5	c3 → WH											1									1		1	ACCEPTABLE DESIGN
c2		O-RING	RUBBER																						
	a6	c2 → a5	O-RING INSERTED INTO BODY										1.5	1	1	1	1	1	1.1	1	1	1.1		1.65	ACCEPTABLE DESIGN
	a7	a3 → a5	COIL ASS'Y FITTED INTO BODY ASS'Y										1	1	1	1	1	1.2	1.1	1	1.4	1.85		1.85	SPECIAL CONTROL
	a8		TURN OVER																				1.3	1.3	ACCEPTABLE DESIGN
c4	a9	PLUNGER	RUBBER END SEAL MOULDED ONTO M/C PLUNGER	6 (28)	±0.05	–	1	1	±0.05	4	1													4	UNACCEPTABLE DESIGN
	a10	a9 → WH											1									1		1	ACCEPTABLE DESIGN
c6		SPRING	STEEL																						
	a11	c6 → a10	SPRING INSERTED INTO PLUNGER										1.5	1	1	1	1	1	1	1	1	1		1.5	ACCEPTABLE DESIGN
	a12	A11 → a8	PLUNGER ASSEMBLY INSERTED										1	1	1	1.1	1.2	1	1	1	1.4	1.85		1.85	SPECIAL CONTROL

Fig. 7.2.42 Variability risks analysis for the solenoid end assembly initial design.

7.2.7.4.2 Solenoid end assembly redesign

This is similar to the initial design, but involves turning as a secondary process on the body to improve a key tolerance capability as indicated in Figure 7.2.44. The body is still impact extruded, but the face which mates with the fuel port block is machined, together with a shoulder on the inside diameter. The front face of the pole, now fully machined and assuming no component cost increase, is assembled up to the machined shoulder on the body. Only the tolerances on the pole length, tube and plunger end seal remain in the stack, reducing the number of components to five.

The variability risks table for the redesign is shown in Figure 7.2.45 and the Conformability Matrix in Figure 7.2.46. Clearly, machining the critical faces on the impact extruded components has reduced the risks associated with conforming to the ±0.2 mm tolerance for the plunger displacement.

The associated potential cost of failure has reduced significantly to a little over £3000. However, there is an

Component/assembly process reference	Component/ assembly process description	Total risk (q_m or q_a)	Failure Mode Description and FMEA Severity Rating (*S*)						Comments (including action for suppliers)
			FUEL FLOW RESTRICTION						
			5						
c1	BOBBIN LENGTH	9	●						UNACCEPTABLE DESIGN
c3	BODY THICKNESS	9	●						UNACCEPTABLE DESIGN
c4	PLUNGER END SEAL LENGTH	4	●						UNACCEPTABLE DESIGN
c5	MAGNETIC POLE LENGTH	9	●						UNACCEPTABLE DESIGN
c7	TUBE	1	○						ACCEPTABLE DESIGN
Total Failure Mode Isocost (%)			40.01						TOTAL FAILURE COST
Total Failure Mode Cost			£3.07M*						£3.07M

* Number of units = 1 000 000

Product cost (*Pc*) = £7.66

Total failure mode isocost (%) = 10 + 10 + 0.01 + 10 + 10 = 40.01

Therefore, total failure mode cost = $\frac{40.01 \times 1\,000\,000 \times 7.66}{100}$ = £3 064 766

Fig. 7.2.43 Conformability Matrix for the solenoid end assembly initial design.

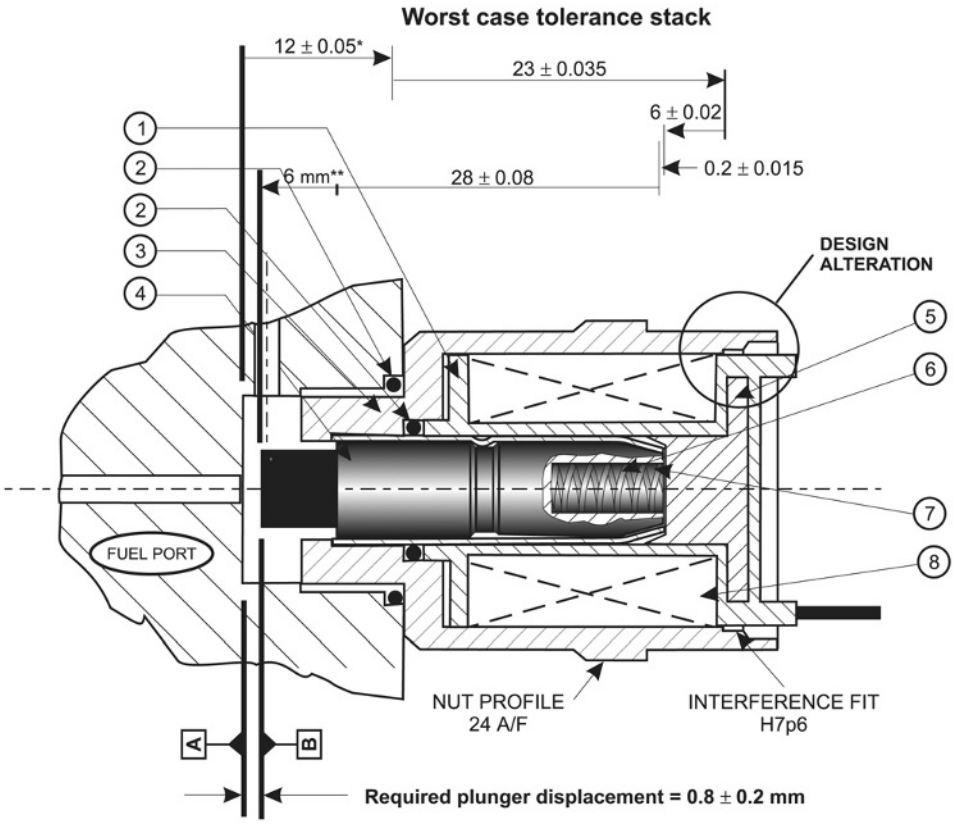
additional cost associated with the extra machining process which adds to the overall product cost. Since it is likely a secondary process involving machining, this will take place on the body thread anyway; the case for turning these critical faces may be further justified. Although machining these faces will raise the cost of the component slightly, this must be secondary to satisfying the overriding customer requirement of meeting the plunger displacement tolerance.

As highlighted by the difference in the potential failure costs, the redesign scheme must be chosen for further design development. Of course, other design schemes could also be explored, but the initial design shown here is of inherently poor quality and therefore must be rejected.

7.2.8 Summary

Decisions made during the design stage of the product development process account for a large proportion of the problems that incur failure costs in production and service. It is possible to relate these failure costs back to the original design intent where variability, and the lack of understanding of variability, is a key failure costs driver. The correct choice of tolerance on a dimensional characteristic can be crucial for the correct functioning of the product in service and tolerance selection can have a large contribution to the overall costs of the product, both production and quality loss.

Process capability indices are not generally specified by designers and subsequently the impact of design decisions made on the production department cannot be fully understood because tolerances alone do not contain enough information. Variability in component manufacture has proved difficult to predict in the early stages of the design process and there are many influencing factors that the designer may not necessarily be able to anticipate. The material and geometrical configuration of the design, and the compatibility with the manufacturing process are the main variability drivers. Although design rules and general manufacturing capability information are available, they are rarely presented in a useful or practical form, especially when innovative



PART No.	DESCRIPTION	MATERIAL	PROCESS
1	BOBBIN	PBT(30% FILL)	INJECTION MOULDED WITH ITEM 5
2	O-RING	RUBBER	–
3	BODY	COLD FORMING STEEL	IMPACT EXTRUDE THEN M/C
4	PLUNGER	STEEL/SILICONE RUBBER	M/C (MOULDED SILICONE RUBBER END SEAL)
5	MAGNETIC POLE	COLD FORMING STEEL	MACHINED
6	SPRING	STEEL	–
7	TUBE	BRASS	ROLLED/DEEP DRAWN
8	COIL	–	–

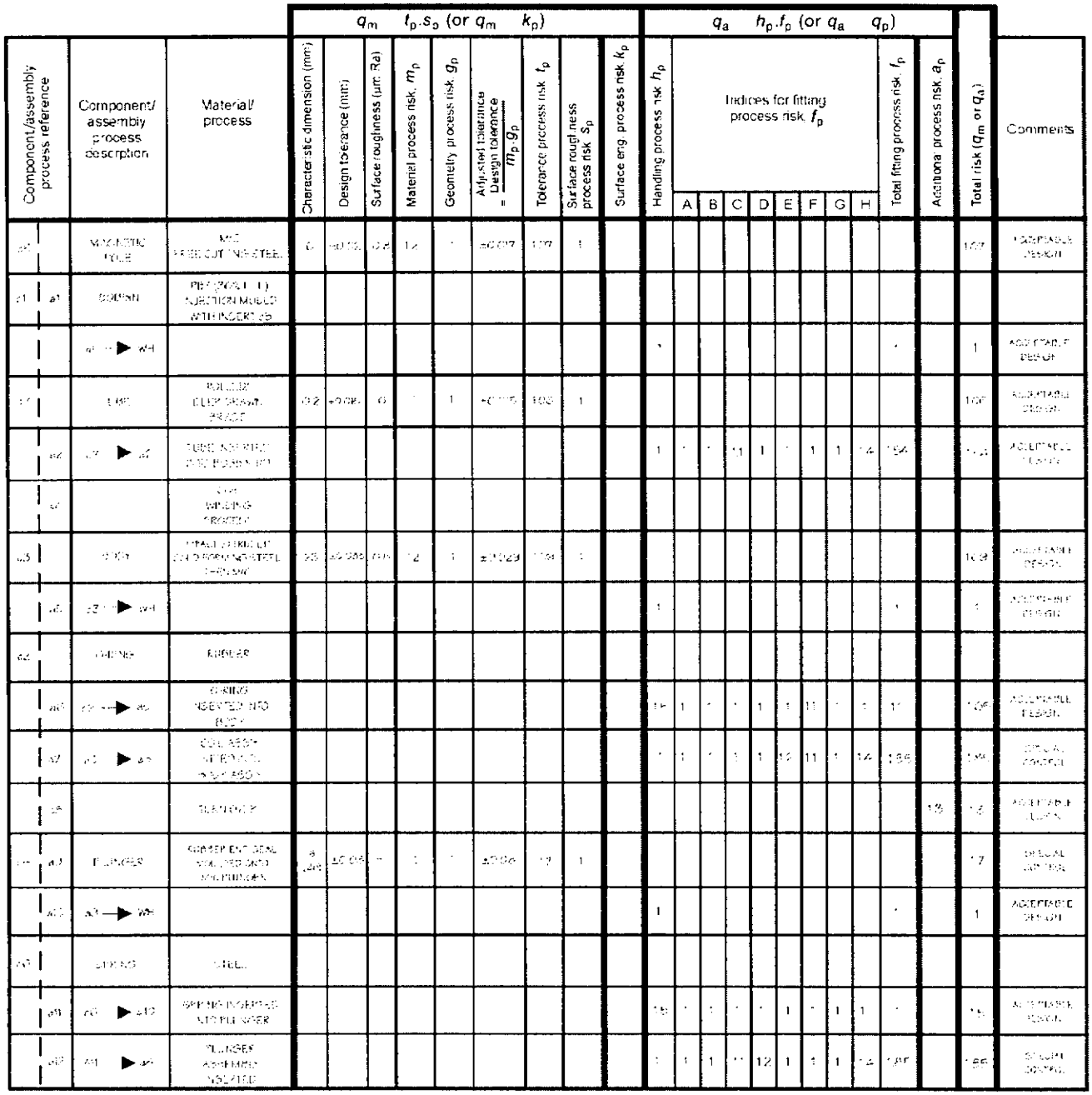
* Tolerance fixed by supplier
** The 6 mm end seal dimension will be used in the analysis, not 28 mm

Fig. 7.2.44 Solenoid end assembly redesign.

design is required. There is a need to set realistic tolerances and anticipate the variability associated with the design to help reduce failure costs later in the product's life-cycle.

The CA methodology is useful in this respect. It is comprised of three sections: the component manufacturing variability risks analysis, the Component assembly variability risks analysis and the determination of the Effects of Non-conformance through the Conformability Map.

The component manufacturing variability risks analysis models the important design/manufacture interface issues which reflect the likely process capability that can be achieved. Included is the assessment of tolerance, geometry, material and surface roughness variability in component manufacture. Quantitative and qualitative manufacturing knowledge is used to support various aspects of the analysis and is taken from a wide range of sources. The concept of an ideal design allows the analysis to generate risk indices, where values greater than unity have a potential for increased variation in production. A simple cost–tolerance relationship is used in the Process Capability Maps, developed for over 60 manufacturing process/material combinations. The maps



Component/assembly process reference	Component/ assembly process description	Material/ process	q_m $t_p.s_p$ (or q_m k_p): Characteristic dimension (mm)	Design tolerance (mm)	Surface roughness (μm Ra)	Material process risk, m_p	Geometry process risk, g_p	Adjusted tolerance = Design tolerance / $m_p.g_p$	Tolerance process risk t_p	Surface roughness process risk s_p	Surface eng. process risk k_p	q_a $h_p.f_p$ (or q_a q_p): Handling process risk h_p	Indices for fitting process risk, f_p: A	B	C	D	E	F	G	H	Total fitting process risk, f_p	Additional process risk, a_p	Total risk (q_m or q_a)	Comments
a0	MAGNETIC POLE	MIC FREE CUTTING STEEL	0	±0.02	0.8	12	1	±0.017	107	1													107	ACCEPTABLE DESIGN
a1 a1	BOBBIN	PBT (ZOLL 1) INJECTION MOULD WITH INSERT 3B																						
	a1 → WH											1									1		1	ACCEPTABLE DESIGN
a2	TUBE	ROLLED DEEP DRAWN BRASS	0.2	±0.08	0	1	1	±0.005	105	1													105	ACCEPTABLE DESIGN
a2	a2 → a2	TUBE INSERTED INTO BOBBIN BOTH										1	1	1	11	1	1	1	1	14	154		154	ACCEPTABLE DESIGN
a3		COIL WINDING PROCESS																						
a3	BODY	IMPACT STRUCTURE COLD FORMING STEEL THIN WALL	0.5	±0.025	0.8	12	1	±0.029	109	1													109	ACCEPTABLE DESIGN
a3	a3 → WH											1									1		1	ACCEPTABLE DESIGN
a4	O-RING	RUBBER																						
a4	a4 → a4	O-RING INSERTED INTO BODY										15	1	1	1	1	1	11	1	1	11		105	ACCEPTABLE DESIGN
a7	a7 → a8	COIL ASSY INTERNAL SUB ASSY										1	1	1	1	1	12	11	1	14	185		185	SPECIAL CONTROL
a8		GLAND OVER																				13	13	ACCEPTABLE DESIGN
a9	PLUNGER	RUBBER END SEAL MOULDED ONTO MIC PLUNGER	8 (48)	±0.05	1	1	1	±0.05	17	1													17	SPECIAL CONTROL
a10	a9 → WH											1									1		1	ACCEPTABLE DESIGN
a10	SPRING	STEEL																						
a11	a10 → a10	SPRING INSERTED INTO PLUNGER										15	1	1	1	1	1	1	1	1	1		15	ACCEPTABLE DESIGN
a12	a11 → a6	PLUNGER ASSEMBLY INSERTED										1	1	1	11	12	1	1	1	14	185		185	SPECIAL CONTROL

Fig. 7.2.45 Variability risks analysis for the solenoid end assembly redesign.

are subsequently employed to determine the process capability estimates for the component characteristics analysed. Through empirical studies, a close correlation between the process capability estimates using the component manufacturing variability risks analysis and shop-floor process capability has been observed.

Most literature tends only to focus on tolerance stack analysis when assessing the capability of assemblies. The variability of the actual assembly operations is rarely considered and does not rely solely on the tolerances accumulating throughout the assembly, but on the feasibility and inherent technical capability of the assembly operations performed, manually or automatically. Developers and practitioners of DFA techniques reason that an assembly with a high assembly efficiency is a better quality product. The natural outcome of having a high assembly efficiency leads to fewer parts in the assembly and, therefore, fewer quality problems to tackle in production. The outcome is not due to any specific analysis process in the DFA technique to address variability, and

Component/assembly process reference	Component/assembly process description	Total risk (q_m or q_a)	Failure Mode Description and FMEA Severity Rating (S)						Comments (including action for suppliers)
			FUEL FLOW RESTRICTION						
			5						
c3	BODY THICKNESS	1.09	○						ACCEPTABLE DESIGN
c4	PLUNGER END SEAL LENGTH	1.7	◒						SPECIAL CONTROL
c5	MAGNETIC POLE LENGTH	1.07	○						ACCEPTABLE DESIGN
c7	TUBE	1.06	○						ACCEPTABLE DESIGN
Total Failure Mode Isocost (%)			0.04						TOTAL FAILURE COST
Total Failure Mode Cost			£3K*						£3K

* Number of units = 1 000 000
Product cost (Pc) = £7.66
Total failure mode isocost (%) = 0.01 + 0.01 + 0.01 + 0.01 = 0.04

$$\text{Therefore, total failure mode cost} = \frac{0.04 \times 1\,000\,000 \times 7.66}{100} = £3064$$

Fig. 7.2.46 Conformability Matrix for the solenoid end assembly redesign.

there still exists a need for analysing the assembly capability of designs, rather than a production cost driven approach. A useful technique for facilitating an assembly risks analysis is the declaration of a sequence of assembly for the components. Through such a diagram, each component in the assembly and, therefore, the potential areas for assembly risk can be logically mapped through the product design.

The component assembly variability risks analysis has the purpose of better understanding the effects of a component's assembly situation on variability, by quantifying the risks that various assembly operations inherently exhibit. The analysis processes are supported by expert knowledge and presented in charts. Again, the theory is that an ideal component assembly situation exists where the assembly risk is unity. Using the charts to reflect the handling process risk, fitting process risk and the risks associated with additional assembly and joining processes, the assembly situation of the component is questioned, accruing penalties at each stage if the design has increased potential for variability.

Current quality-cost models are useful for identifying general trends in a long-term improvement programme, but are of limited use in the identification of the failure costs associated with actual design decisions. A link between the costs that can be typically expected in practice due to failure or non-conformance of the product in production or service, and the probability of fault occurrence, is made using FMEA through the Conformability Map. The underlying concept assumes that as failures become more severe, they are going to cost more when they fail. The quality-cost model embedded in the Conformability Map allows the designer to assess the level of acceptability, special control or unacceptability for non-safety critical and safety critical component characteristics in the design by determination of the process capability measures from the previous two stages of the analysis. The Conformability Map also allows failure isocosts (percentages of total product cost), and, therefore, the total failure cost to be estimated with knowledge of the likely product cost and production volume. The nature of the underlying cost models limits the accuracy of the failure cost estimates at an absolute

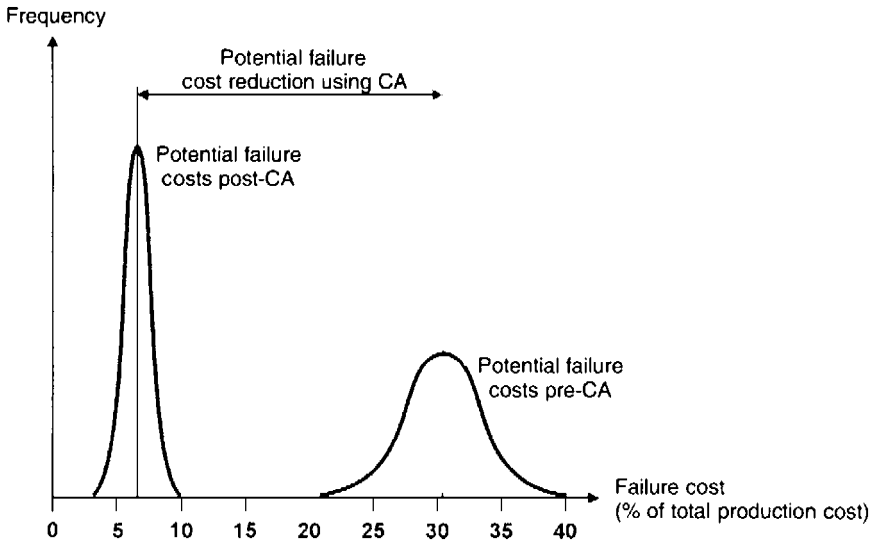


Figure 7.2.47 Influence of the team-based application of CA on several product introduction projects.

level and so they become useful in evaluating and comparing design schemes for their potential quality loss. The model can alternatively be employed in setting capability targets for characteristics to incur allowable failure costs dependent on the failure severity of the product.

Through performing an analysis using CA, many modes of application have been highlighted. This has resulted from the way that the CA design performance measures allow a non-judgemental 'language' to develop between the design team. It has also been found not to inhibit the design process, but provide a structured analysis with which to trace design decisions. The knowledge embedded within CA also allows the designer to generate process capable solutions and open up discussion with suppliers. The analysis is currently facilitated through the use of a paper-based assessment. This has many benefits, including improved team working, and provides a more unconstrained approach than if the analyses were computer based. It also allows the knowledge to be readily visible and available at anytime for the designer to scrutinize and manipulate if they chose to do so.

The potential benefits of using CA in the early stages of the product development process have been found to be:

- Early awareness of potential design problems through a systematic analysis
- Produces more process capable designs with regard to their manufacture and assembly
- Reduces internal/external failure costs
- Reduces lead times
- Focused discussions with suppliers.

Finally, the main benefit as far as competitive business performance is concerned is the potential for reduction in failure costs. Studies using CA very early in the development process of a number of projects have indicated that the potential failure costs were all reduced through an analysis. This is shown in Figure 7.2.47, where this potential failure cost reduction is shown as the difference between *pre-CA* and *post-CA* application by the teams analysing the product designs.

This page is intentionally left blank

Section Eight

Product safety and legal liability

This page is intentionally left blank

Chapter 8.1

8.1

Project management

Donald J. Smith

8.1.1 Setting objectives and specifications

Realistic reliability, maintainability and safety (RAMS) objectives need to be set with due regard to the customer's design and operating requirements and cost constraints. In the case of contract development or plant engineering, these are likely to be outlined in a tender document or a requirements specification. Some discussion and joint study with the customer may be required to establish economic reliability values which sensibly meet his or her requirements and are achievable within the proposed technology at the costs allowed for. Over-specifying the requirement may delay the project when tests eventually show that objectives cannot be met and it is realized that budgets will be exceeded. It is not uncommon for safety targets to be overstated due to ill-informed targetting methods.

When specifying an MTBF it is a common mistake to state a confidence level; in fact the MTBF requirement stands alone. The addition of a confidence level implies a statistical demonstration and supposes that the MTBF would be established by a single demonstration at the stated confidence. On the contrary, a design objective is a target and must be stated without statistical limitations.

Vague statements such as 'high reliability' and 'the highest quality' should be avoided at all costs. They are totally subjective and cannot be measured. Therefore they cannot be demonstrated or proved.

Consideration of the equipment type and the use to which it is put will influence the parameters chosen.

A major contribution to the problems associated with reliability and quality comes from the lack of (or inadequacy of) the engineering design specification. It should specify the engineering requirements in full, including reliability and MTTR parameters. These factors should include:

1. Functional description: speeds, functions (including safety-related functions), human interfaces and operating periods.
2. Environment: temperature, humidity, etc.
3. Design life: related to wearout and replacement policy.
4. Physical parameters: size and weight restrictions, power supply limits.
5. Standards: BS, US MIL, Def Con, etc., standards for materials, components and tests.
6. Finishes: appearance and materials.
7. Ergonomics: human limitations and safety considerations.
8. Reliability, availability and maintainability: module reliability and MTTR objectives. Equipments R and M related to module levels.
9. Manufacturing quantity: projected manufacturing levels – First off, Batch, Flow.
10. Maintenance philosophy: type and frequency of preventive maintenance. Repair level, method of diagnosis, method of second-line repair.

8.1.2 Planning, feasibility and allocation

The design and assurance activities described in this book simply will not take place unless there is real management understanding and commitment to a reliability and maintainability programme with specific resources allocated. Responsibilities have to be placed on individuals for each of the activities and a reliability programme

Reliability, Maintainability and Risk; ISBN: 9780750666947
Copyright © 2005 Elsevier Ltd. All rights of reproduction, in any form, reserved.

manager has to be appointed with sufficient authority and the absence of conflicting priorities (that is, programme dates) to control the RAM objectives. Milestones, with dates, will be required against which progress can be measured as, for example:

Completion of feasibility study (including RAM calculations).
Reliability objectives for modules and bought-out items allocated.
Test specification prepared and agreed.
Prototype tests completed.
Modifications arising from tests completed.
Demonstrations of reliability and maintainability.
Design review dates.

The purpose of a feasibility study is to establish if the performance specification can be met within the constraints of cost, technology, time and so on. This involves a brief reliability prediction, based perhaps on a block diagram approach, in order to decide if the design proposal has a reasonable chance of being engineered to meet the requirements. Allocation of objectives is important if the objectives are not to be met by a mixture of over- and under-design.

It is useful to remember that there are three levels of RAM measurement:

Prediction: A modelling exercise which relies on the validity of historical failure rates to the design in question. This provides the lowest level of confidence.

Statistical demonstration test: This provides sample failure information (perhaps even zero failures in a given amount of time). It is usually in a test rather than field environment. Whilst providing more confidence than paper *Prediction* it is still subject to statistical risk and the limitations of a test environment.

Field data: Except in the case of very high reliability systems (e.g. submerged cable and repeater), realistic numbers of failures are obtained and can be used in a reliability growth programme as well as for comparison with the original targets.

8.1.3 Programme activities

The extent of the reliability and maintainability activities in a project will depend on:

The severity of the requirement.
The complexity of the product.
Time and cost constraints.
Safety considerations.
The number of items to be produced.

A Safety and Reliability Plan must be produced for each project or development. Without this there is nothing against which to audit progress and, therefore, no formal measure of progress towards the targets. Figure 8.1.1 shows a simple RAMS design cycle which provides a model against which to view the activities. These include:

- *Feasibility study* – an initial 'prediction' to ascertain if the targets are realistic or impossible.
- *Setting objectives* – discussed above with allocation and feasibility.
- *Contract requirements* – the formal agreement on the RAM targets, warranty, acceptance criteria, etc.
- *Design reviews* – these are intended to provide an evaluation of the design at defined milestones. The design review team should include a variety of skills and be chaired by a person independent of the design team. The following checklist is a guide to the factors which might be considered:
  1. Electrical factors involving critical features, component standards, circuit trade-offs, etc.
  2. Software reliability including configuration control, flowcharts, user documentation, etc.
  3. Mechanical features such as materials and finish, industrial design, ergonomics, equipment practice and so on.
  4. Quality and reliability covering environmental testing, RAM predictions and demonstrations, FMECA, test equipment and procedures, trade-offs, etc.
  5. Maintenance philosophy including repair policy, MTTR prediction, maintenance resource forecasts, customer training and manuals.
  6. Purchased items involving lead times, multiple sourcing, supplier evaluation and make/buy decisions.
  7. Manufacturing and installation covering tolerances, burn-in, packaging and transport, costs, etc.
  8. Other items include patents, value engineering, safety, documentation standards and product liability.
- *RAMS predictions* – this focuses attention on the critical failure areas, highlights failures which are difficult to diagnose and provides a measure of the design reliability against the objectives. FMEA, FTA and other modelling exercises are used, in the design reviews, to measure conformance to the RAMS targets.
- *Design trade-offs* – these may be between R and M and may involve sacrificing one for the other as, for example, between the reliability of the wrapped joint and the easy replaceability of a connector. Major trade-offs will involve the design review whereas others will be made by the designer.
- *Prototype tests* – these cover marginal, functional, parametric, environmental and reliability tests. It is

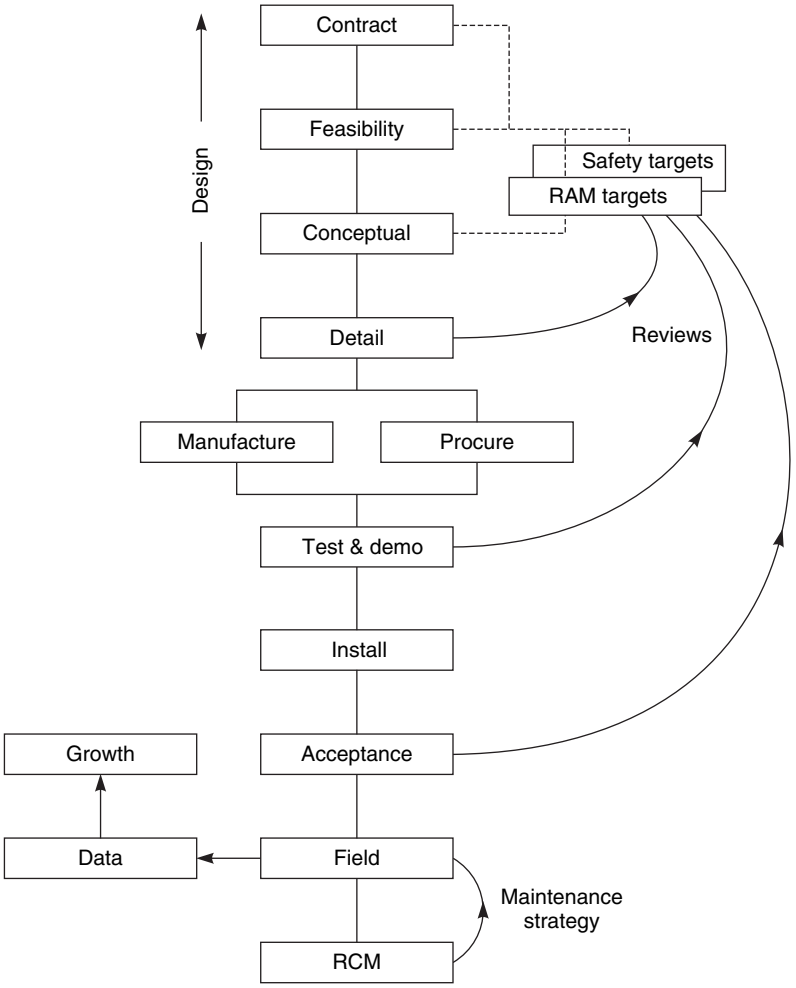


Fig. 18.1 RAMS cycle.

the first opportunity to observe reliability in practice and to make some comparison against the predictions.

- *Parts selection and approval* – involves field tests or seeking field information from other users. The continued availability of each part is important and may influence the choice of supplier.
- *Demonstrations* – since these involve statistical sampling, test plans have to be calculated at an early stage so that the risks can be evaluated.
- *Spares provisioning* – this affects reliability and maintainability and has to be calculated during design.
- *Data collection and failure analysis* – failure data, with the associated stress information, is essential to reliability growth programmes and also to future predictions. A formal failure-reporting scheme should be set up at an early stage so that tests on the earliest prototype modules contribute towards the analysis.
- *Reliability growth* – establishing report and analysis to confirm that field reliability growth meets targets.
- *Training* – design engineers should be trained to a level where they can work with the R and M specialist. Customer training of maintenance staff is another aspect which may arise.

8.1.4 Responsibilities

RAM is an integral part of the design process. In many cases mere lip service is given to it and this leads to little more than high-level predictions being carried out too late in the design. These have no effect whatever in bringing the design nearer to the targets. Reliability and maintainability are engineering parameters and the responsibility for their achievement is therefore primarily with the design team. Quality assurance techniques play a vital role in achieving the goals but cannot be used to 'test in' reliability to a design which has its own inherent level. Three distinct responsibilities therefore emerge

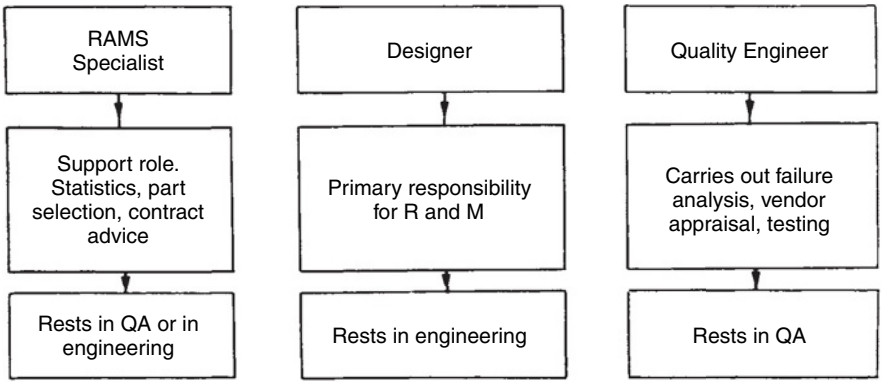


Fig. 18.2

which are complementary but do not replace each other. See Figure 8.1.2.

8.1.5 Functional safety capability

The advent of IEC 61508 (functional safety) has led to the activities associated with achieving functional safety being identified as an area of specific responsibility. Although largely achieved as a result of working to an ISO 9000 based management system, they nevertheless require an additional procedure(s) to provide the detail. The main items can be summarized as follows:

Contract/project review: All contracts, prior to acceptance by a company, need to be examined to ascertain if they involve safety-related requirements. These requirements may be stated directly by the client or may be implicit by reference to some standard. Clients may not always use appropriate terms to refer to safety-related applications or safety–integrity requirements. Therefore, a competent person will need to decide if a contract is safety related or not. Similarly, all major modifications, new product designs or major projects have to be assessed in this way.

Control of modifications: Change proposals need to be declared safety related or not. The change proposal document must therefore have a space dedicated to describing the impact of the change, which must be declared safety related or not by a competent person.

Project management (i.e. control of specific tasks): Every project requires a 'safety authority' who is independent of the actual design tasks. Again, it has to be someone who is deemed competent to plan, instigate and review all the functional safety assessment activities associated with the project.

Competence: The references to competence (in this context) imply that all the safety-related tasks (e.g. carrying out assessments, doing audits, acting as project manager) have been defined and that a 'register' is maintained identifying those who are able to carry out each task. This needs to be reviewed on a regular basis and used to select individuals for each task. It will also identify training needs (or the need for consultancy assistance) where all the skills are not currently available in-house.

Remedial action: The defect recording documentation (used in test, commissioning and in the field) will also need to cater, in the same way as the change proposal, for identifying the safety-related implications of incidents and failures. Responsibilities for carrying out remedial action will also need to be defined.

Functional safety validation: This consists of a positive demonstration that all the safety-related requirements have been achieved. It is usually satisfied by the satisfactory completion of the final acceptance test together with a cross-reference to each of the requirements. A matrix is often generated to fulfil this need. In addition, a documented close out of all the outstanding remedial actions from reviews, tests, defects and changes can be appended to the matrix.

This topic is addressed in detail in *Functional Safety: A Straightforward Guide to IEC 61508*, 2nd edition, 2004, Smith D.J. and Simpson K.G.L., Butterworth-Heinemann ISBN 07506 6269 7.

8.1.6 Standards and guidance documents

There are a number of standards which might be called for. The more important are as follows:

- BS 5760: Reliability of systems, equipment and components – this is in a number of parts. Part 1 is Guide to Reliability Programme Management and outlines the reliability activities. Other parts deal with prediction, data, practices and so on.

- UK Ministry of Defence Standard 00–40 reliability and maintainability: this is divided into eight parts. Parts 1 and 2 are concerned with project requirements and the remainder with requirement documents, training, procurement and so on.
- US Military Standard 785A Reliability Program for Systems and Equipment Development and Production: specifies programme plans, reviews, predictions and so on.
- US Military Standard 470 Maintainability Programme Requirements: A document, from 1966, which covers the programme plan and specifies activities for design criteria, design review, trade-offs, data collection, predictions and status reporting.

This page is intentionally left blank

Chapter 8.2

8.2

Contract clauses and their pitfalls

Donald J. Smith

8.2.1 Essential areas

Since the late 1950s in the United States, reliability and maintainability requirements have appeared in both military and civil engineering contracts. These contracts often carry penalties for failure to meet these objectives. For 30 years in the UK, suppliers of military and commercial electronic and telecommunication equipment have also found that clauses specifying reliability and maintainability were being included in invitations to tender and in the subsequent contracts. Suppliers of highly reliable and maintainable equipment are often well able to satisfy such conditions with little or no additional design or manufacturing effort, but incur difficulty and expense since a formal demonstration of these parameters may not have been previously attempted. Furthermore, a failure-reporting procedure may not exist and therefore historical data as to a product's reliability or repair time may be unobtainable.

The inclusion of system-effectiveness parameters in a contract involves both the suppliers of good and poor equipment in additional activities. System effectiveness clauses in contracts range from a few words – specifying Availability, failure rate or MTBF of all or part of the system – to many pages containing details of design and test procedures, methods of collecting failure data, methods of demonstrating reliability and repair time, limitations on component sources, limits to size and cost of test equipment, and so on. Two types of pitfall arise from such contractual conditions:

1. Those due to the omission of essential conditions or definitions.
2. Those due to inadequately worded conditions which present ambiguities, concealed risks, eventualities unforeseen by both parties, etc.

The following headings are essential if reliability or maintainability is to be specified.

8.2.1.1 Definitions

If a mean time to repair or down time is specified, then the meaning of repair time must be defined in detail. Mean time to repair is often referred to as intended mean down time.

Failure itself must also be thoroughly defined at system and module levels. It may be necessary to define more than one type of failure (for example, total system failure or degradation failure) or failures for different operating modes (for example, in flight or on ground) in order to describe all the requirements. MTBFs might then be ascribed to the different failure types. MTBFs and failure rates often require clarification as to the meaning of 'failure' and 'time'. The latter may refer to operating time, revenue time, clock time, etc. Types of failure which do not count for the purpose of proving the reliability (for example, maintenance induced or environment outside limits) have also to be defined.

For process-related equipment it is usual to specify Availability. Unless, however, some failure modes are defined, the figures can be of little value. For example, in a safety system, failure may consist of spurious alarm or of failure to respond. Combining the two failure rates produces a misleading figure and the two modes must be evaluated separately. Figure 8.2.1 reminds us of the Bathtub Curve with early, random and wearout failures. Reliability parameters usually refer to random failures unless stated to the contrary, it being assumed that burn-in failures are removed by screening and wearout failures eliminated by preventive replacement.

It should be remembered that this is a statistical picture of the situation and that, in practice, it is rarely possible to

Reliability, Maintainability and Risk; ISBN: 9780750666947
Copyright © 2005 Elsevier Ltd. All rights of reproduction, in any form, reserved.

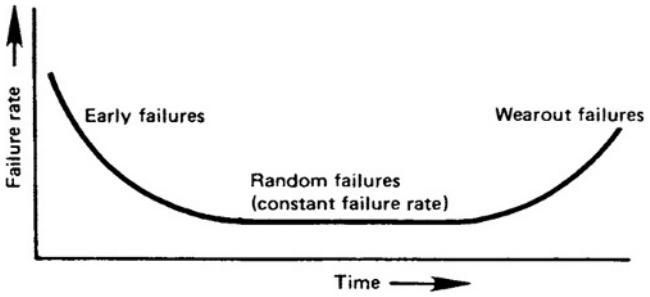


Fig. 8.2.1

ascribe a particular failure to any of these categories. It is therefore vital that, if reliability is being demonstrated by a test or in the field, these early and wearout failures are eliminated, as far as possible, by the measures already described. The specification should make clear which types of failure are being observed in a test.

Parameters should not be used without due regard to their meaning and applicability. Failure rate, for example, has little meaning except when describing random failures. Remember that in systems involving redundancy, constant failure rate may not apply except in some special cases. Availability, MTBF or reliability should then be specified in preference.

Reliability and maintainability are often combined by specifying the useful parameter, Availability. This can be defined in more than one way and should therefore be specifically defined. The usual form is the Steady State Availability, which is MTBF/(MTBF + MDT), where MDT is the Mean Down Time.

8.2.1.2 Environment

A common mistake is to fail to specify the environmental conditions under which the product is to work. The specification is often confined to temperature range and maximum humidity, and this is not always adequate. Even these two parameters can create problems, as with temperature cycling under high-humidity conditions. Other stress parameters include pressure, vibration and shock, chemical and bacteriological attack, power supply variations and interference, radiation, human factors and many others. The combination or the cycling of any of these parameters can have significant results.

Where equipment is used as standby units or held as spares, the environmental conditions will be different to those experienced by operating units. It is often assumed that because a unit is not powered, or in store, it will not fail. In fact the environment may be more conducive to failure under these circumstances. Self-generated heat and mechanical self-cleaning wiping actions are often important ingredients for reliability. If equipment is to be transported while the supplier is liable for failure, then the environmental conditions must be evaluated. On the other hand, over-specifying environmental conditions is a temptation which leads to over-design and higher costs. Environmental testing is expensive, particularly if large items of equipment are involved and if vibration tests are called for. These costs should be quantified by obtaining quotations from a number of test houses before any commitment is made to demonstrate equipment under environmental conditions.

Maintainability can also be influenced by environment. Conditions relating to safety, comfort, health and ergonomic efficiency will influence repair times since the use of protective clothing, remote-handling devices, safety precautions, etc. increases the active elements of repair time by slowing down the technician.

8.2.1.3 Maintenance support

The provision of spares, test equipment, personnel, transport and the maintenance of both spares and test equipment is a responsibility which may be divided between supplier and customer or fall entirely on either. These responsibilities must be described in the contract and the supplier must be conscious of the risks involved in the customer not meeting his or her side of the bargain.

If the supplier is responsible for training the customer's maintenance staff then levels of skill and training have to be laid down.

Maintenance philosophy, usually under customer control, plays a part in determining reliability. Periodic inspection of a non-attended system during which failed redundant units are changed yields a different MTBF to the case of immediate repair of failed units, irrespective of whether they result in system failure. The maintenance philosophy must therefore be defined.

A contract may specify an MTTR supported by a statement such as 'identification of faulty modules will be automatic and will be achieved by automatic test means. No additional test equipment will be required for diagnosis.' This type of requirement involves considerable additional design effort in order to permit all necessary diagnostic signals to be made accessible and for measurements to be made. Additional hardware will be required either in the form of BITE or an 'intelligent' portable terminal with diagnostic capability. If such a requirement is overlooked when costing and planning the design the subsequent engineering delay and cost is likely to be considerable.

8.2.1.4 Demonstration and prediction

The supplier might be called upon to give a statistical demonstration of either reliability or repair time. In the

case of maintainability a number of corrective or preventive maintenance actions will be carried out and a given MTTR (or better) will have to be achieved for some proportion of the attempts. In this situation it is essential to define the tools and equipment to be used, the maintenance instructions, test environment and technician level. The method of task selection, the spares and the level of repair to be carried out also require stating. The probability of failing the test should be evaluated since some standard tests carry high supplier's risks. When reliability is being demonstrated then a given number of hours will be accumulated and a number of failures stated, above which the test is failed. Again, statistical risks apply and the supplier needs to calculate the probability of failing the test with good equipment and the customer that of passing inadequate goods.

Essential parameters to define here are environmental conditions, allowable failures (for example, maintenance induced), operating mode, preventive maintenance, burn-in, testing costs. It is often not possible to construct a reliability demonstration which combines sensible risks (<15%) for both parties with a reasonable length of test. Under these circumstances the acceptance of reliability may have to be on the basis of accumulated operating hours on previously installed similar systems.

An alternative to statistical or historical demonstrations of repair time and reliability is a guarantee period wherein all or part of the failure costs, and sometimes redesign costs, are borne by the supplier. In these cases great care must be taken to calculate the likely costs. It must be remembered that if 100 items of equipment meet their stated MTBF under random failure conditions, then after operating for a period equal to one MTBF, 63 of them, on average, will have failed.

From the point of view of producer's risk, a warranty period is a form of reliability demonstration since, having calculated the expected number of failures during the warranty, there is a probability that more will occur. Many profit margins have been absorbed by the unbudgeted penalty maintenance arising from this fact.

A prediction is often called for as a type of demonstration. It is desirable that the data source is agreed between the two parties or else the Numbers Game will ensue as various failure rates are 'negotiated' by each party seeking to turn the prediction to his or her favour.

8.2.1.5 Liability

The exact nature of the supplier's liability must be spelt out, including the maximum penalty which can be incurred. If some qualifying or guarantee period is involved it is necessary to define when this commences and when the supplier is free of liability. The borders between delivery, installation, commissioning and operation are often blurred and therefore the beginning of the guarantee period will be unclear.

It is wise to establish a mutually acceptable means of arbitration in case the interpretation of later events becomes the subject of a dispute. If part of the liability for failure or repair is to fall on some other contractor, care must be taken in defining each party's area. The interface between equipments guaranteed by different suppliers may be physically easy to define but there exists the possibility of failures induced in one item of equipment owing to failure or degraded performance in another. This point should be considered where more than one supplier is involved.

8.2.2 Other areas

The following items are often covered in a detailed invitation to tender.

8.2.2.1 Reliability and maintainability programme

The detailed activities during design, manufacturing and installation are sometimes spelt out contractually. In a development contract this enables the customer to monitor the reliability and maintainability design activities and to measure progress against agreed milestones. Sometimes standard programme requirements are used as, for example:

US Military Standard 470, Maintainability Program Requirements. US Military Standard 785, Requirements for Reliability Program. BS 4200: Part 5 Reliability programmes for equipment.
BS 5760 Reliability of constructed and manufactured products, systems, equipment and components.

Typical activities specified are:

Prediction – data sources, mathematical models.
Testing – methods and scheduling of design, environmental and other tests.
Design review – details of participation in design reviews.
Failure mode and effect analysis – details of method and timing.
Failure reporting – failure reporting documents and reporting procedures.

8.2.2.2 Reliability and maintainability analysis

The supplier may be required to offer a detailed reliability or maintainability prediction together with an

explanation of the techniques and data used. Alternatively, a prediction may be requested using defined data and methods of calculation. Insistence on optimistic data makes it more difficult to achieve the predicted values whereas pessimistic data leads to over-design.

8.2.2.3 Storage

The equipment may be received by the customer and stored for some time before it is used under conditions different to normal operation. If there is a guarantee period then the storage conditions and durations will have to be defined. The same applies to storage and transport of spares and test equipments.

8.2.2.4 Design standards

Specific design standards are sometimes described or referenced in contracts or their associated specifications. These can cover many areas, including:

Printed-board assemblies – design and manufacture
Wiring and soldering
Nuts, bolts and threads
Finishes
Component ratings
Packaging

A problem exists that these standards are very detailed and most manufacturers have their own version. Although differences exist in the fine detail they are usually overlooked until some formal acceptance inspection takes place, by which time retrospective action is difficult, time consuming and costly.

8.2.2.5 Safety-related equipment

Clauses relating to the safety–integrity (i.e. the reliability in respect of hazardous failure modes) of equipment are now commonplace. Sometimes specific integrity requirements are made for defined failure modes. These may be expressed as qualitative 'probability of failure on demand' or 'failure rate' targets. Alternatively they may be expressed as a SIL requirement as defined in IEC 61508. Examples might be:

- The probability of failure on demand of the shutdown system in response to outlet overpressure shall not exceed 5×10^{-3}.
- The failure rate in respect of delivering unignited gas from the appliance shall not exceed 3×10^{-3} Pa.
- The safety–integrity of the emergency shutdown system with respect to failure to respond to any valid input shall meet the requirements of SIL 2 of IEC 61508.

Specific failure modes are not always defined by the customer as, for example:

- The emergency shutdown system shall meet the requirements of SIL 2 of IEC 61508.

It then becomes necessary to establish, with the customer, what the hazardous failure modes are of the application for which he is obtaining the equipment. Even less specific requirements are such as:

- The emergency shutdown system shall meet the requirements of IEC 61508.

Here, it may become necessary for the vendor to supply the expertize needed to establish hazards (i.e. HAZOP) and to target the integrity levels needed to meet credible risk criteria.

8.2.3 Pitfalls

The foregoing lists those aspects of reliability and maintainability likely to be mentioned in an invitation to tender or in a contract. There are pitfalls associated with the omission or inadequate definition of these factors and some of the more serious are outlined below.

8.2.3.1 Definitions

The most likely area of dispute is the definition of what constitutes a failure and whether or not a particular incident ranks as one or not. There are levels of failure (system, unit), types of failure (catastrophic, degradation), causes of failure (random, systematic, over-stress) and there are effects of failure (dormant, hazardous). For various combinations of these, different MTBF and MTTR objectives with different penalties may be set. It is seldom sufficient, therefore, to define failure as not performing to specification since there are so many combinations covered by that statement. Careful definition of the failure types covered by the contract is therefore important.

8.2.3.2 Repair time

Repair times could be divided into elements. Initially they can be grouped into active and passive elements and, broadly speaking, the active elements are dictated by system design and the passive by maintenance and operating arrangements. For this reason, the supplier should never guarantee any part of the repair time which is influenced by the user.

8.2.3.3 Statistical risks

A statistical maintainability test is described by a number of repair actions and an objective MTTR

which must not be exceeded on more than a given number of attempts. A reliability test involves a number of hours and a similar pass criterion of a given number of failures. In both cases producer and consumer risks apply and unless these risks are calculated they can prove to be unacceptable. Where published test plans are quoted, it is never a bad thing to recalculate the risks involved. It is not difficult to find a test which requires the supplier to achieve an MTBF 50 times the value which is to be proved in order to stand a reasonable chance of passing the test.

8.2.3.4 Quoted specifications

Sometimes a reliability or maintainability programme or test plan is specified by calling up a published standard. Definitions are also sometimes dealt with in this way. The danger with blanket definitions lies in the possibility that not all the quoted terms are suitable and that the standards will not be studied in every detail.

8.2.3.5 Environment

Environmental conditions affect both reliability and repair times. Temperature and humidity are the most usual to be specified and the problem of cycling has already been pointed out. If other factors are likely to be present in field use then they must either be specifically excluded from the range of environment for which the product is guaranteed or included, and therefore allowed for in the design and in the price. It is not desirable to specify every parameter possible, since this leads to overdesign.

8.2.3.6 Liability

When stating the supplier's liability it is important to establish its limit in terms of both cost and time. Suppliers must ensure that they know when they are finally free of liability.

8.2.3.7 In summary

The biggest pitfall of all is to assume that either party wins any advantage from ambiguity or looseness in the conditions of a contract. In practice, the hours of investigation and negotiation which ensue from a dispute far outweigh any advantage that might have been secured, to say nothing of the loss of goodwill and reputation. If every effort is made to cover all the areas discussed as clearly and simply as possible, then both parties will gain.

8.2.4 Penalties

There are various ways in which a penalty may be imposed on the basis of maintenance costs or the cost of system outage. It must be remembered that any cash penalty must be a genuine and reasonable pre-estimate of the damages thought to result. Some alternatives are briefly outlined.

8.2.4.1 Apportionment of costs during guarantee

Figure 8.2.2(a) illustrates the method where the supplier pays the total cost of corrective maintenance during the guarantee period. He or she may also be liable for the cost of redesign made necessary by systematic failures. In some cases the guarantee period recommences for those parts of the equipment affected by modifications. A disadvantage of this arrangement is that it gives the customer no great incentive to minimize maintenance costs until the guarantee has expired. If the maintenance is carried out by the customer and paid for by the supplier then the latter's control over the preventive maintenance effectiveness is minimal. The customer should never be permitted to benefit from poor maintenance, for which reason this method is not very desirable.

An improvement of this is obtained by Figure 8.2.2(b), whereby the supplier pays a proportion of the costs during the guarantee and both parties therefore have an incentive to minimize costs. In Figure 8.2.2(c) the supplier's proportion of the costs decreases over the liability period. In Figure 8.2.2(d) the customer's share of the maintenance costs remains constant and the supplier pays the excess. The arrangements in (b) and (c) both provide mutual incentives. Arrangement (d), however, provides a mixed incentive. The customer has, initially, a very high incentive to reduce maintenance costs but once the ceiling has been reached this disappears. On the other hand, (d) recognizes the fact that for a specified MTBF the customer should anticipate a given amount of repair. Above this amount the supplier pays for the difference between the achieved and contracted values.

8.2.4.2 Payment according to down time

The above arrangements involve penalties related to the cost of repair. Some contracts, however, demand a payment of some fixed percentage of the contract price during the down time. Providing that the actual sum paid is less than the cost of the repair this method is similar to Figure 8.2.2(b), although in practice it is not likely to be so generous. In any case, an arrangement of this type must be subject to an upper limit.

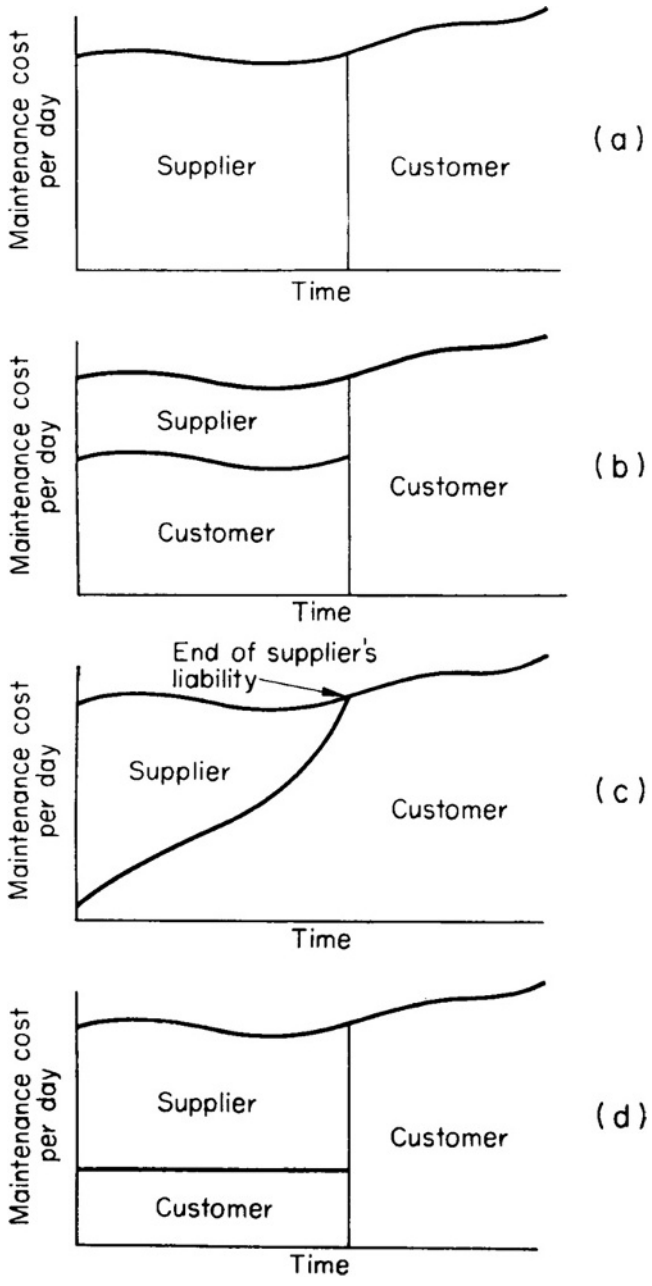


Fig. 8.2.2 Methods of planning penalties.

8.2.4.3 In summary

Except in case (a) it would not be practicable for the supplier to carry out the maintenance. Usually the customer carries out the repairs and the supplier pays according to some agreed rate. In this case the supplier must require some control over the recording of repair effort and a right to inspect the customer's maintenance records and facilities from time to time. It should be remembered that achievement of reliability and repair time objectives does not imply zero maintenance costs. If a desired MTBF of 20,000 h is achieved for each of 10 items of equipment, then in one year (8760 h) about four failures can be expected. On this basis (d) is fairer than (a). When part of a system is subcontracted to another supplier, then the prime contractor must ensure that he or she passes on an appropriate allocation of the reliability commitments in order to be protected.

8.2.5 Subcontracted reliability assessments

It is common, in the development of large systems, for either the designer or the customer to subcontract the task of carrying out Failure Mode Analysis and Reliability Predictions. It may be that the customer requires the designer to place such a contract with a firm of consultants approved by the customer. It is desirable for such work to be covered by a contract which outlines the scope of work and the general agreement between the two parties. Topics to be covered include:

Data bank sources to be used.
Traceability where non-published data are used.
Target Reliability, Availability or MTBF.
Specific duty cycles and environmental profiles.
Extent of the Failure Mode Analysis required.
Types of recommendation required in the event of the prediction indicating that the design will not meet the objectives.
Requirement for ranking of major contributors to system failure.
If the prediction indicates that the design more than meets the objective, a requirement to identify the areas of over-design.
Identification of critical single-point or Common Cause failures.
Identification of Safety Hazards.
Recommendations for maintenance (e.g. replacement strategy, periodic inspection time).
Calculations of spares-holding levels for defined probabilities of stockout.
Aspects of Human Error required in the analysis.
Arrangements for control and review of the assessment work, including reporting (e.g. Conceptual design report, Interim prediction and report, Detailed Failure Mode Analysis, Final Design Qualification report, etc.); Schedules, Costs, Invoicing.

8.2.6 Examples

(a) The following requirements might well be placed in an invitation to tender for a piece of measuring

equipment. They are by no means intended as a model contract and the reader might care to examine them from both the designer's and customer's points of view.

Loss of measurement shall include the total loss of temperature recording as well as a loss of recording accuracy exceeding 20%.

Mode 1: the loss of 2 or more consecutive measurements.
Mode 2: the loss of recording accuracy of temperature within the range (>1–20%).

Bidders shall satisfy 'XYZ' that the equipment will meet the following requirements.

MTBF (Mode 1) > 5 years
MTBF (Mode 2) > 10 years

The MTBF shall be achieved without the use of redundancy but by the use of appropriate component quality and stress levels. It shall be demonstrated by means of a failure mode analysis of the component parts. FARADIP.THREE shall be used as the failure rate data source except where alternative sources are approved by 'XYZ'.

The above specification takes no account of the infant mortality failures usually characterized by a decreasing failure rate in the early life of the equipment. The supplier shall determine a suitable burn-in period and arrange for the removal of these failures by an appropriate soak test.

No wearout failure mechanisms, characterized by an increasing failure rate, shall be evident in the life of the equipment. Any components requiring preventive replacement in order to achieve this requirement shall be highlighted to 'XYZ' for consideration and approval.

In the event of the MTBFs not being demonstrated, at 80% confidence, after 10 device years of operation have been accumulated then the supplier will carry out any necessary redesign and modification in order to achieve the MTBF objectives.

During the life of the equipment any systematic failures shall be dealt with by the supplier, who will carry out any necessary redesign and modification. A systematic failure is one which occurs three or more times for the same root cause.

(b) The failure mode 'downstream overpressure resulting from loss of control of the pressure reduction station' is safety-related. It will meet safety–integrity level SIL 2 of IEC 61508. (*Note*: this infers a failure rate, for the failure mode described, of less than 10^{-3} Pa. It also implies various qualitative requirements.

This page is intentionally left blank

Chapter 8.3

8.3

Product liability and safety legislation

Donald J. Smith

Product liability is the liability of a supplier, designer or manufacturer to the customer for injury or loss resulting from a defect in that product. There are reasons why it has recently become the focus of attention. The first is the publication in July 1985 of a directive by the European Community, and the second is the wave of actions under United States law which has resulted in spectacular awards for claims involving death or injury. By 1984, sums awarded resulting from court proceedings often reached $1 million. Changes in the United Kingdom became inevitable and the Consumer Protection Act reinforces the application of strict liability. It is necessary, therefore, to review the legal position.

8.3.1 The general situation

8.3.1.1 Contract law

This is largely governed by the Sale of Goods Act 1979, which requires that goods are of merchantable quality and are reasonably fit for the purpose intended. Privity of Contract exists between the buyer and seller which means that only the buyer has any remedy for injury or loss and then only against the seller, although the cascade effect of each party suing, in turn, the other would offset this. However, exclusion clauses are void for consumer contracts. This means that a condition excluding the seller from liability would be void in law. Note that a contract does not have to be in writing and that a sale, in this context, implies the existence of a contract.

8.3.1.2 Common law

The relevant area is that relating to the Tort of Negligence, for which a claim for damages can be made. Everyone has a duty of care to his or her neighbour, in law, and failure to exercise reasonable precautions with regard to one's skill, knowledge and the circumstances involved constitutes a breach of that care. A claim for damages for common law negligence is, therefore, open to anyone and not restricted as in Privity of Contract. On the other hand, the onus is with the plaintiff to prove negligence which requires proof:

That the product was defective.
That the defect was the cause of the injury.
That this was foreseeable and that the plaintiff failed in his or her duty of care.

8.3.1.3 Statute law

The main Acts relevant to this area are:

Sale of Goods Act 1979
- Goods must be of Merchantable Quality
- Goods must be fit for purpose

Unfair Contract Terms Act 1977
- Exclusion of personal injury liability is void
- Exclusion of damage liability only if reasonable

Consumer Protection Act 1987
- Imposes strict liability
- Replaces the Consumer Safety Act 1978

Health and Safety at Work Act 1974, section 6
- Involves the criminal law. Places a duty to construct and install items, processes and materials without health or safety risks. It applies to places of work. Responsibility involves everyone including management. The Consumer Protection Act extends section 6 of the Health and Safety at Work Act to include all areas of use. European legislation will further extend this (see Section 8.3.4.5).

Reliability, Maintainability and Risk; ISBN: 9780750666947
Copyright © 2005 Elsevier Ltd. All rights of reproduction, in any form, reserved.

8.3.1.4 In summary

The present situation involves a form of strict liability but:

Privity of Contract excludes third parties in contract claims.
The onus is to prove negligence unless the loss results from a breach of contract.
Exclusion clauses, involving death and personal injury, are void.

8.3.2 Strict liability

8.3.2.1 Concept

The concept of strict liability hinges on the idea that liability exists for no other reason than the mere existence of a defect. No breach of contract, or act of negligence, is required in order to incur responsibility and their manufacturers will be liable for compensation if products cause injury.

The various recommendations which are summarized later involve slightly different interpretations of strict liability ranging from the extreme case of everyone in the chain of distribution and design being strictly liable, to the manufacturers being liable unless they can prove that the defect did not exist when the product left them. The Consumer Protection Act makes manufacturers liable whether or not they were negligent.

8.3.2.2 Defects

A defect for the purposes of product liability, includes:

Manufacturing	Presence of impurities or foreign bodies.
	Fault or failure due to manufacturing or installation.
Design	Product not fit for the purpose stated.
	Inherent safety hazard in the design.
Documentation	Lack of necessary warnings.
	Inadequate or incorrect operating and maintenance instructions resulting in a hazard.

8.3.3 The Consumer Protection Act 1987

8.3.3.1 Background

In 1985, after 9 years of discussion, the European Community adopted a directive on product liability and member states were required to put this into effect before the end of July 1988. The Consumer Protection Bill resulted in the Consumer Protection Act 1987, which establishes strict liability as described above.

8.3.3.2 Provisions of the Act

The Act will provide that a producer (and this includes manufacturers, those who import from outside the EC and retailers of 'own brands') will be liable for damage caused wholly or partly by defective products which include goods, components and materials but exclude unprocessed agricultural produce. 'Defective' is defined as not providing such safety as people are generally entitled to expect, taking into account the manner of marketing, instructions for use, the likely uses and the time at which the product was supplied. Death, personal injury and damage (other than to the product) exceeding £275 are included.

The consumer must show that the defect caused the damage but no longer has the onus of proving negligence. Defences include:

- The state of scientific and technical knowledge at the time was such that the producer could not be expected to have discovered the defect. This is known as the 'development risks' defence.
- The defect results from the product complying with the law.
- The producer did not supply the product.
- The defect was not present when the product was supplied by the manufacturer.
- The product was not supplied in the course of business.
- The product was in fact a component part used in the manufacture of a further product and the defect was not due to this component.

In addition, the producer's liability may be reduced by the user's contributory negligence. Further, unlike the privity limitation imposed by contract law, any consumer is covered in addition to the original purchaser.

The Act sets out a general safety requirement for consumer goods and applies it to anyone who supplies goods which are not reasonably safe having regard to the circumstances pertaining. These include published safety standards, the cost of making goods safe and whether or not the goods are new.

8.3.4 Health and Safety at Work Act 1974

8.3.4.1 Scope

Section 6 of this Act applies strict liability to articles produced for use at work, although the Consumer

Protection Act extends this to all areas. It is very wide and embraces designers, manufacturers, suppliers, hirers and employers of industrial plant and equipment. We are now dealing with criminal law and failure to observe the duties laid down in the Act is punishable by fine or imprisonment. Claims for compensation are still dealt with in civil law.

8.3.4.2 Duties

The main items are:

To design and construct products without risk to health or safety.
To provide adequate information to the user for safe operation.
To carry out research to discover and eliminate risks.
To make positive tests to evaluate risks and hazards.
To carry out tests to ensure that the product is inherently safe.
To use safe methods of installation.
To use safe (proven) substances and materials.

8.3.4.3 Concessions

The main concessions are:

- It is a defence that a product has been used without regard to the relevant information supplied by the designer.
- It is a defence that the design was carried out on the basis of a written undertaking by the purchaser to take specified steps sufficient to ensure the safe use of the item.
- One's duty is restricted to matters within one's control.
- One is not required to repeat tests upon which it is reasonable to rely.

8.3.4.4 Responsibilities

Basically, everyone concerned in the design and provision of an article is responsible for it. Directors and managers are held responsible for the designs and manufactured articles of their companies and are expected to take steps to assure safety in their products. Employees are also responsible. The 'buck' cannot be passed in either direction.

8.3.4.5 European Community legislation

In 1989/1990 the EC agreed to a framework of directives involving health and safety. This legislation will eventually replace the Health and Safety at Work Act, being more prescriptive and detailed than the former. The directive mirrors the Health and Safety at Work Act by setting general duties on both employees and employers for all work activities.

In implementing this European legislation the Health and Safety Commission will attempt to avoid disrupting the framework which has been established by the Health and Safety at Work Act. The directive covers:

The overall framework
The workplace
Use of work equipment
Use of personal protective equipment
Manual handling
Display screen equipment.

8.3.4.6 Management of Health and Safety at Work Regulations 1992

These lay down broad general duties which apply to almost all Great Britain on shore and offshore activities. They are aimed at improving health and safety management and can be seen as a way of making more explicit what is called for by the H&SW Act 1974. They are designed to encourage a more systematic and better organized approach to dealing with health and safety, including the use of risk assessment.

8.3.5 Insurance and product recall

8.3.5.1 The effect of product liability trends

- An increase in the number of claims.
- Higher premiums.
- The creation of separate Product Liability Policies.
- Involvement of insurance companies in defining quality and reliability standards and procedures.
- Contracts requiring the designer to insure the customer against genuine and frivolous consumer claims.

8.3.5.2 Some critical areas

- *All risks* – this means all risks specified in the policy. Check that your requirements are met by the policy.
- *Comprehensive* – essentially means the same as the above.
- *Disclosure* – the policy holder is bound to disclose any information relevant to the risk. Failure to do so, whether asked for or not, can invalidate a claim. The test of what should be disclosed is described as 'anything the prudent insurer should know'.

- *Exclusions* – The Unfair Contract Terms Act 1977 does not apply to insurance, so read and negotiate accordingly. For example, defects related to design could be excluded and this would considerably weaken a policy from the product liability standpoint.
- Prompt notification of claims.

8.3.5.3 Areas of cover

Premiums are usually expressed as a percentage of turnover and cover is divided into three areas:

Product liability – cover against claims for personal injury or loss.
Product guarantee – cover against the expenses of warranty/repair.
Product recall – cover against the expenses of recall.

8.3.5.4 Product recall

A design defect causing a potential hazard to life, health or safety may become evident when a number of products are already in use. It may then become necessary to recall, for replacement or modification, a batch of items, some of which may be spread throughout the chain of distribution and others in use. The recall may vary in the degree of urgency depending on whether the hazard is to life, health or merely reputation. A hazard which could reasonably be thought to endanger life or to create a serious health hazard should be treated by an emergency recall procedure. Where less critical risks involving minor health and safety hazards are discovered a slightly less urgent approach may suffice. A third category, operated at the vendor's discretion, applies to defects causing little or no personal hazard and where only reputation is at risk.

If it becomes necessary to implement a recall the extent will be determined by the nature of the defect. It might involve, in the worst case, every user or perhaps only a specific batch of items. In some cases the modification may be possible in the field and in others physical return of the item will be required. In any case, a full evaluation of the hazard must be made and a report prepared.

One person, usually the Quality Manager, must be responsible for the handling of the recall and must be directly answerable to the Managing Director or Chief Executive. The first task is to prepare, if appropriate, a Hazard Notice in order to warn those likely to be exposed to the risk. Circulation may involve individual customers when traceable, field service staff, distributors, or even the news media. It will contain sufficient information to describe the nature of the hazard and the precautions to be taken. Instructions for returning the defective item can be included, preferably with a prepaid return card. Small items can be returned with the card whereas large ones, or products to be modified in the field, will be retained while arrangements are made.

Where products are despatched to known customers a comparison of returns with output records will enable a 100% check to be made on the coverage. Where products have been despatched in batches to wholesalers or retail outlets the task is not so easy and the quantity of returns can only be compared with a known output, perhaps by area. Individual users cannot be traced with 100% certainty. Where customers have completed and returned record cards after purchase the effectiveness of the recall is improved.

After the recall exercise has been completed a major investigation into the causes of the defect must be made and the results progressed through the company's Quality and Reliability Programme. Causes could include:

Insufficient test hours
Insufficient test coverage
Insufficient information sought on materials
Insufficient industrial engineering of the product prior to manufacture
Insufficient production testing
Insufficient field/user trials
Insufficient user training.

Section **Nine**

Design techniques: electronics

This page is intentionally left blank

Chapter 9.1

9.1

Circuit simulation

Mike Tooley

Computer simulation provides you with a powerful and cost-effective tool for designing, simulating, and analysing a wide variety of electronic circuits. In recent years, the computer software packages designed for this task have not only become increasingly sophisticated but also have become increasingly easy to use. Furthermore, several of the most powerful and popular packages are now available at low cost either in evaluation, 'lite' or student versions. In addition, there are several excellent freeware and shareware packages.

Whereas early electronic simulation software required that circuits were entered using a complex *netlist* that described all of the components and connections present in a circuit, most modern packages

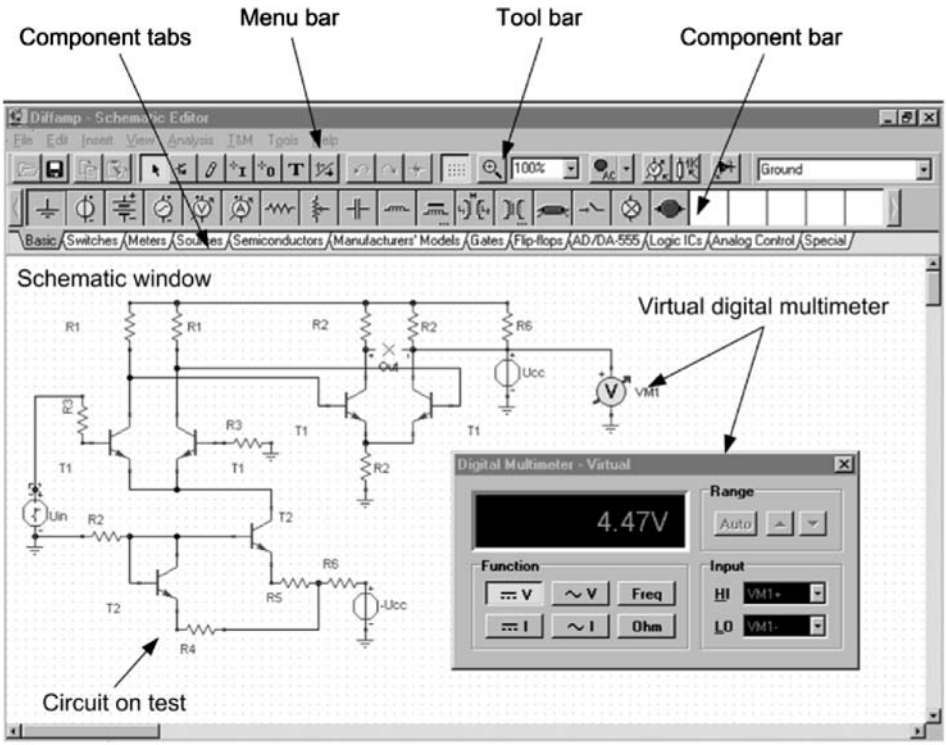


Fig. 9.1.1 Using Tina Pro to construct and test a circuit prior to detailed analysis.

Electronic Circuits; ISBN: 9780750669238
Copyright © 2006 Elsevier Ltd. All rights of reproduction, in any form, reserved.

use an on-screen graphical representation of the circuit on test. This, in turn, generates a netlist (or its equivalent) for submission to the computational engine that actually performs the circuit analysis using mathematical models and *algorithms*. In order to describe the characteristics and behaviour of components such as diodes and transistors, manufacturers often provide models in the form of a standard list of parameters.

Most programs that simulate electronic circuits use a set of algorithms that describe the behaviour of electronic components. The most commonly used algorithm was developed at the Berkeley Institute in the United States and it is known as *SPICE* (Simulation Program with Integrated Circuit Emphasis).

Results of circuit analysis can be displayed in various ways, including displays that simulate those of real test instruments (these are sometimes referred to as *virtual instruments*). A further benefit of using electronic circuit simulation software is that, when a circuit design has been finalized, it is usually possible to export a file from the design/simulation software to a PCB layout package. It may also be possible to export files for use in screen printing or CNC drilling. This greatly reduces the time that it takes to produce a finished electronic circuit.

9.1.1 Types of analysis

Various types of analyses are available within modern SPICE-based circuit simulation packages. These are discussed below.

9.1.1.1 DC analysis

DC analysis determines the DC operating point of the circuit under investigation. In this mode any wound components (e.g. inductors and transformers) are short-circuited and any capacitors that may be present are left open-circuit. In order to determine the initial conditions, a DC analysis is usually automatically performed prior to a transient analysis. It is also usually performed prior to an AC small-signal analysis in order to obtain the linearized, small-signal models for non-linear devices. Furthermore, if specified, the DC small-signal value of a transfer function (ratio of output variable to input source), input resistance, and output resistance is also computed as a part of the DC solution. The DC analysis can also be used to generate DC transfer curves in which a specified independent voltage or current source is stepped over a user-specified range and the DC output variables are stored for each sequential source value.

9.1.1.2 AC small-signal analysis

The AC small-signal analysis feature of SPICE software computes the AC output variables as a function of frequency. The program first computes the DC operating point of the circuit and determines linearized, small-signal models for all of the non-linear devices in the circuit (e.g. diodes and transistors). The resultant linear circuit is then analysed over a user-specified range of frequencies. The desired output of an AC small-signal analysis is usually a transfer function (voltage gain, transimpedance, etc.). If the circuit has only one AC input, it is convenient to set that input to unity and zero phase, so that output variables have the same value as the transfer function of the output variable with respect to the input.

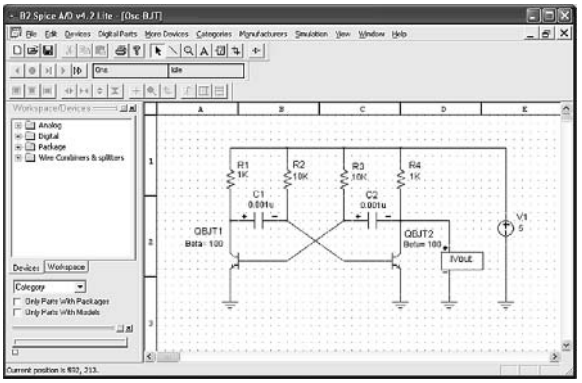


Fig. 9.1.2 An astable multivibrator circuit being simulated using B2 Spice.

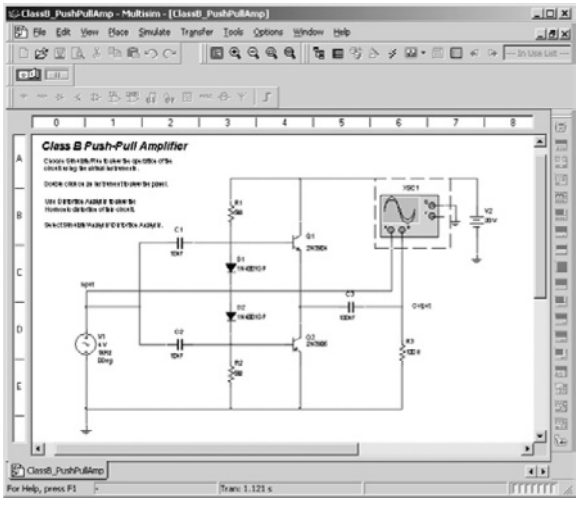


Fig. 9.1.3 A Class B push–pull amplifier circuit being simulated by Multisim.

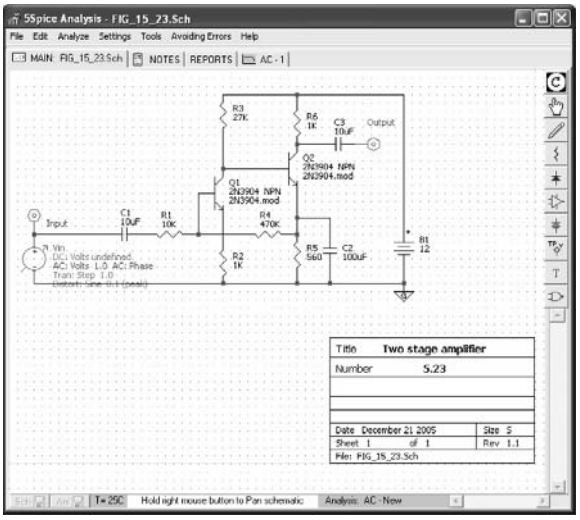


Fig. 9.1.4 High-gain amplifier being analysed using the 5Spice Analysis package.

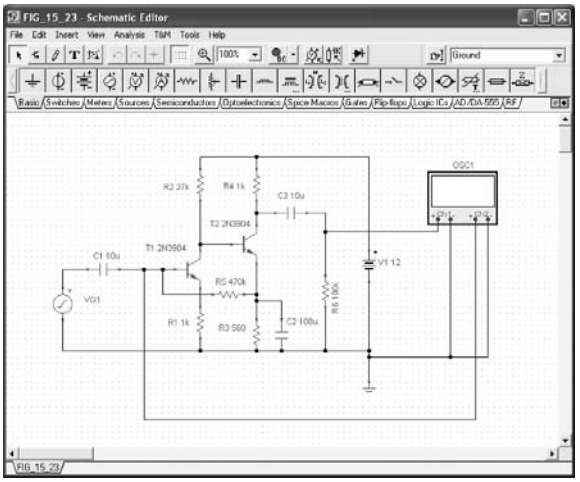


Fig. 9.1.6 High-gain amplifier being analysed using the Tina Pro package.

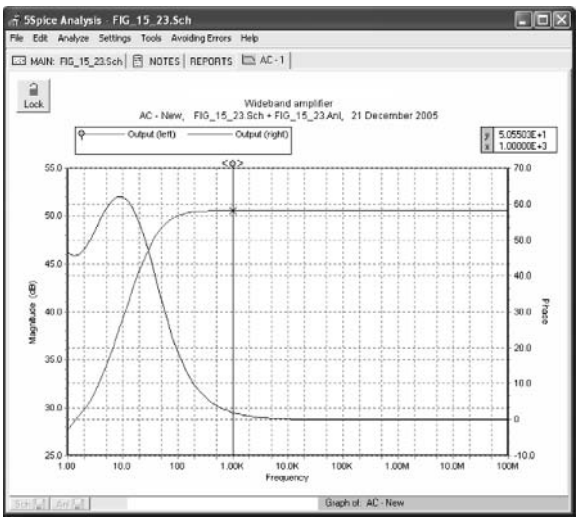


Fig. 9.1.5 Gain and phase plotted as a result of small-signal AC analysis of the circuit in Figure 9.1.4.

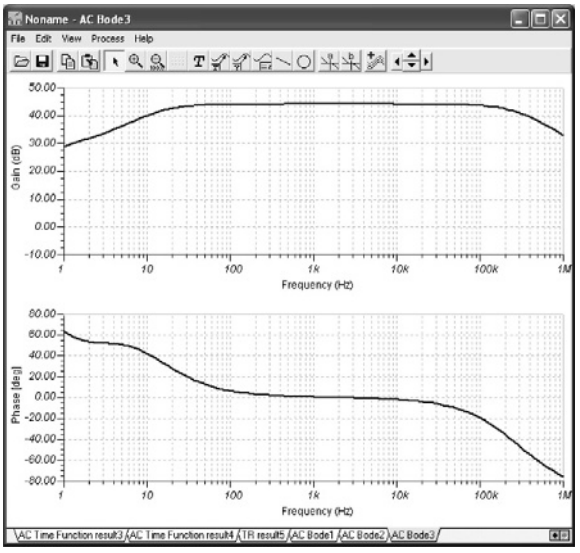


Fig. 9.1.7 Gain and phase plotted as a result of small-signal AC analysis of the circuit in Figure 9.1.6.

9.1.1.3 Transient analysis

The transient analysis feature of a SPICE package computes the transient output variables as a function of time over a user-specified time interval. The initial conditions are automatically determined by a DC analysis. All sources that are not time dependent (for example, power supplies) are set to their DC value.

9.1.1.4 Pole-zero analysis

The pole-zero analysis facility computes the poles and/or zeros in the small-signal AC transfer function. The program first computes the DC operating point and then determines the linearized, small-signal models for all the non-linear devices in the circuit. This circuit is then used to find the poles and zeros of the transfer function.

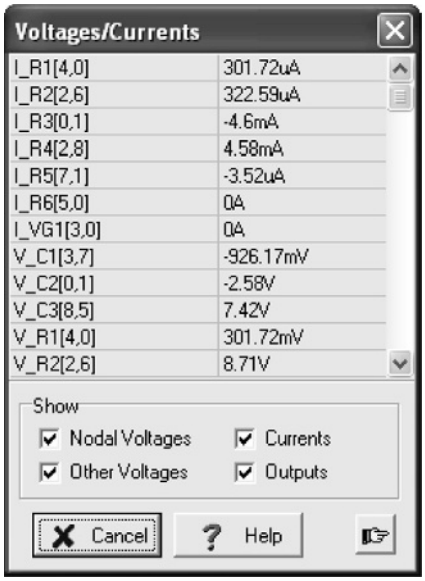


Fig. 9.1.8 Results of DC analysis of the circuit shown in Figure 9.1.6.

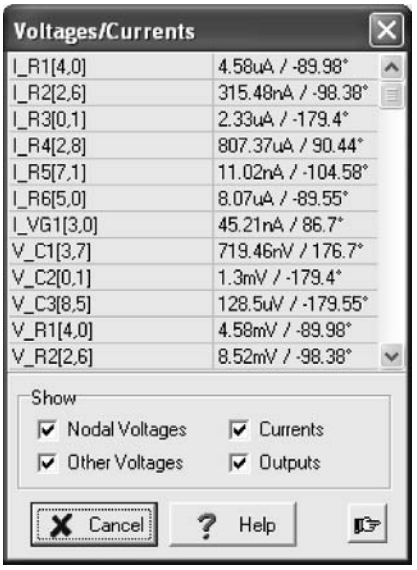


Fig. 9.1.10 Results of AC analysis of the circuit shown in Figure 9.1.6.

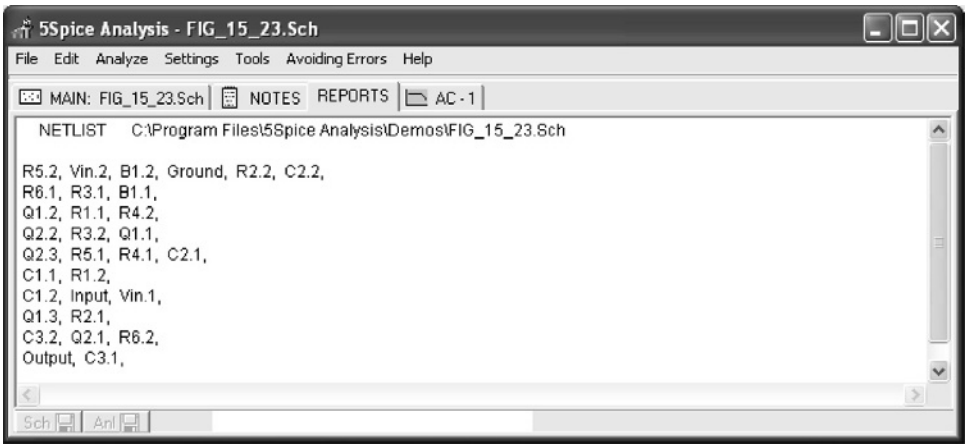


Fig. 9.1.9 Computer generated netlist for the circuit shown in Figure 9.1.4.

Two types of transfer functions are usually supported. One of these determines the voltage transfer function (i.e. output voltage divided by input voltage) and the other usually computes the output *transimpedance* (i.e. output voltage divided by input current) or *transconductance* (i.e. output current divided by input voltage). These two transfer functions cover all the cases and one can make it possible to determine the poles/zeros of functions like impedance ratio (i.e. input impedance divided by output impedance) and voltage gain. The input and output ports are specified as two pairs of nodes. Note that, for complex circuits it can take some time to carry out this analysis and the analysis may fail if there is an excessive number of poles or zeros.

9.1.1.5 Small-signal distortion analysis

The distortion analysis facility provided by SPICE-driven software packages computes steady-state

harmonic and inter-modulation products for small input signal magnitudes. If signals of a single frequency are specified as the input to the circuit, the complex values of the second and third harmonics are determined at every point in the circuit. If there are signals of two frequencies input to the circuit, the analysis finds out the complex values of the circuit variables at the sum and difference of the input frequencies, and at the difference of the smaller frequency from the second harmonic of the larger frequency.

9.1.1.6 Sensitivity analysis

Sensitivity analysis allows you to determine either the DC operating-point sensitivity or the AC small-signal sensitivity of an output variable with respect to all circuit variables, including model parameters. The software calculates the difference in an output variable (either a node voltage or a branch current) by perturbing each parameter of each device independently. Since the method is a numerical approximation, the results may demonstrate second order affects in highly sensitive parameters, or may fail to show very low but non-zero sensitivity. Further, since each variable is perturbed by a small fraction of its value, zero-valued parameters are not analysed (this has the benefit of reducing what is usually a very large amount of data).

9.1.1.7 Noise analysis

The noise analysis feature determines the amount of noise generated by the components and devices (e.g. transistors) present in the circuit that is being analysed. When provided with an input source and an output port, the analysis calculates the noise contributions of each device (and each noise generator within the device) to the output port voltage. It also calculates the input noise to the circuit, equivalent to the output noise referred to the specified input source. This is done for every frequency point in a specified range. After calculating the spectral densities, noise analysis integrates these values over the specified frequency range to arrive at the total noise voltage/current (over this frequency range).

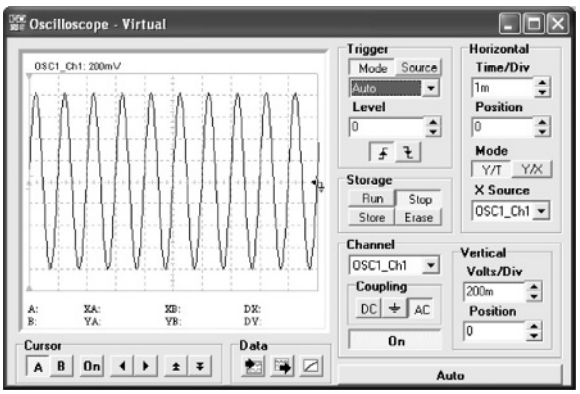


Fig. 9.1.11 Using the virtual oscilloscope in Tina Pro to display the output voltage waveform for the circuit shown in Figure 9.1.6.

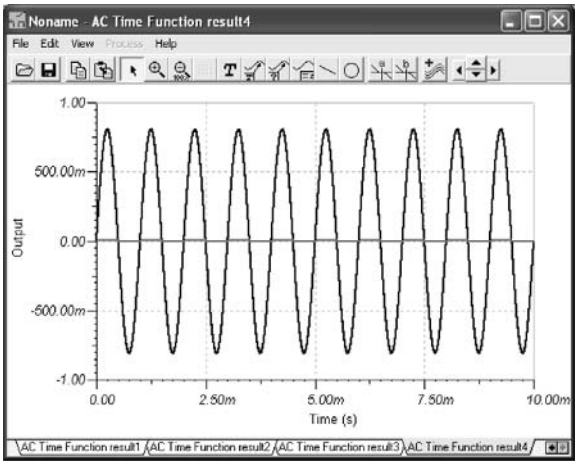


Fig. 9.1.12 Alternative waveform plotting facility provided in Tina Pro.

9.1.1.8 Thermal analysis

Many SPICE packages will allow you to determine the effects of temperature on the performance of a circuit. Most analyses are performed at normal ambient temperatures (e.g. 27 °C) but it can be advantageous to look at the effects of reduced or increased temperatures, particularly where the circuit is to be used in an environment in which there is a considerable variation in temperature.

9.1.2 Netlists and component models

The following is an example of how a netlist for a simple *differential amplifier* is constructed (note that the line numbers have been included solely for explanatory purposes):

1. SIMPLE DIFFERENTIAL PAIR
2. VCC 7 0 12
3. VEE 8 0 −12
4. VIN 1 0 AC1

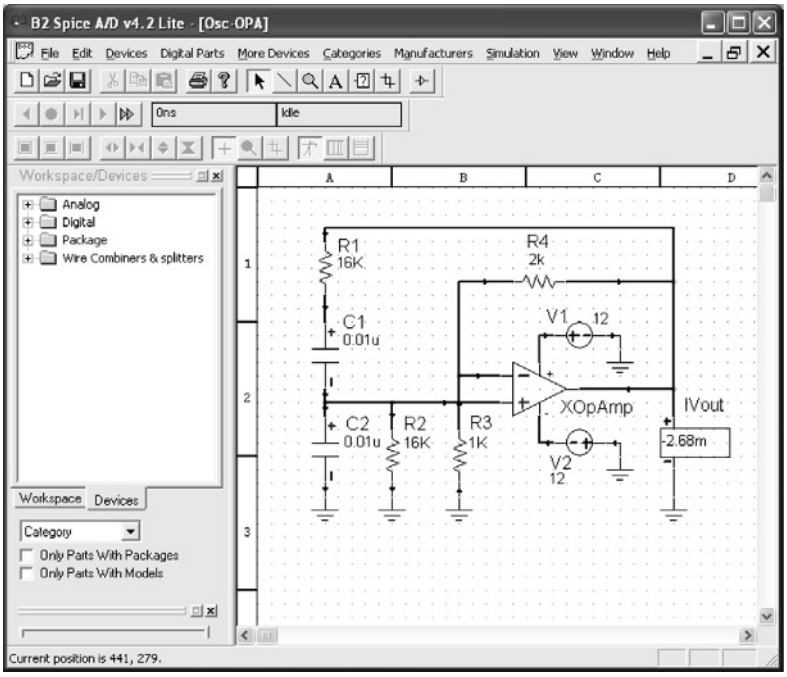


Fig. 9.1.13 Analysis of a Wien Bridge oscillator using B2 Spice.

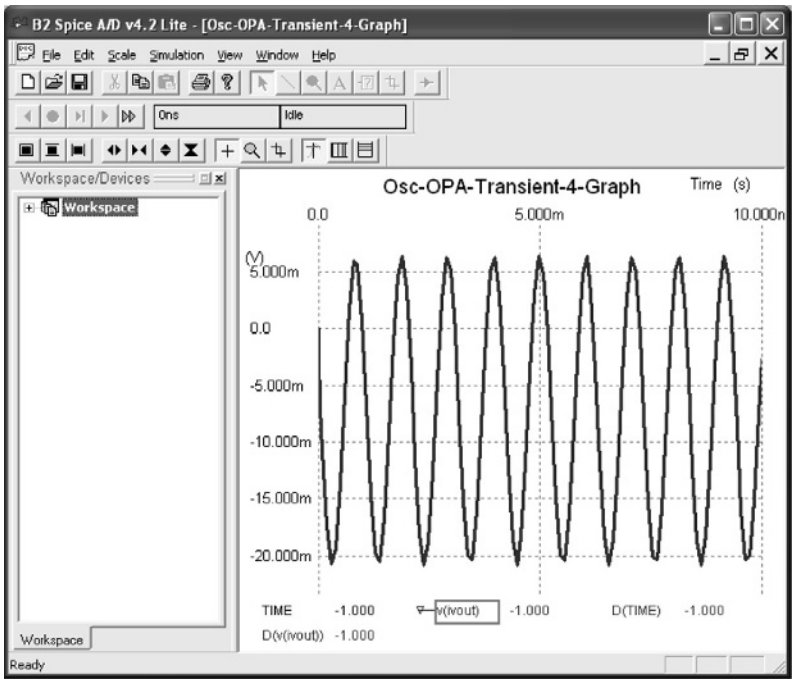


Fig. 9.1.14 Transient analysis of the circuit in Figure 9.1.13 produced the output waveform plot.

5. RS1 1 2 1K
6. RS2 6 0 1K
7. Q1 3 2 4 MOD1
8. Q2 5 6 4 MOD1
9. RC1 7 3 10K
10. RC2 7 5 10K
11. RE 4 8 10K
12. MODEL MODI NPN BF=50 VAF=50 IS=1.E-12 RB=100 CJC=.5PF TF=.6NS
13. .TF V(5) VIN
14. .AC DEC 10 1 100MEG
15. .END

Lines 2 and 3 define the supply voltages. V_{CC} is +12 V and is connected between node 7 and node 0 (signal ground). V_{EE} is −12 V and is connected between node 8 and node 0 (signal ground). Line 4 defines the input voltage which is connected between node 1 and node 0 (ground) whilst lines 5 and 6 define 1 kΩ resistors (RS1 and RS2) connected between 1 and 2, and 6 and 0.

Lines 7 and 8 are used to define the connections of two transistors (Q1 and Q2). The characteristics of these transistors (both identical) are defined by MOD1 (see line 12). Lines 9, 10 and 11 define the connections of three further resistors (RC1, RC2 and RE, respectively). Line 12 defines the transistor model. The device is NPN and has a current gain of 50. The corresponding circuit is shown in Figure 9.1.15.

Most semiconductor manufacturers provide detailed SPICE models for the devices that they produce. The following is a manufacturer's SPICE model for a 2N3904 transistor:

NPN (Is=6.734f Xti=3 Eg=l.11 Vaf=74.03 Bf=416.4 Ne=1.259 Ise=6.734 Ikf=66.78m Xtb=1.5 Br=.7371

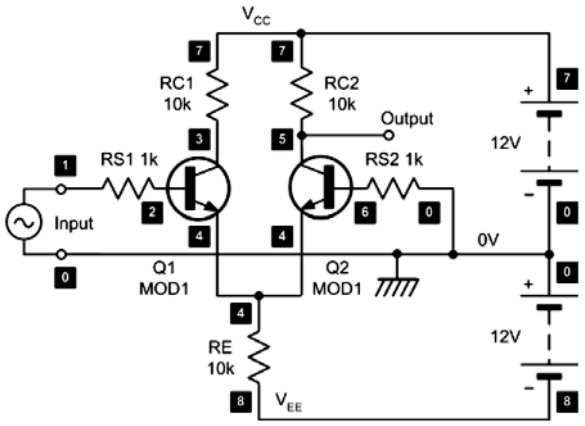


Fig. 9.1.15 Differential amplifier with the nodes marked for generating a netlist.

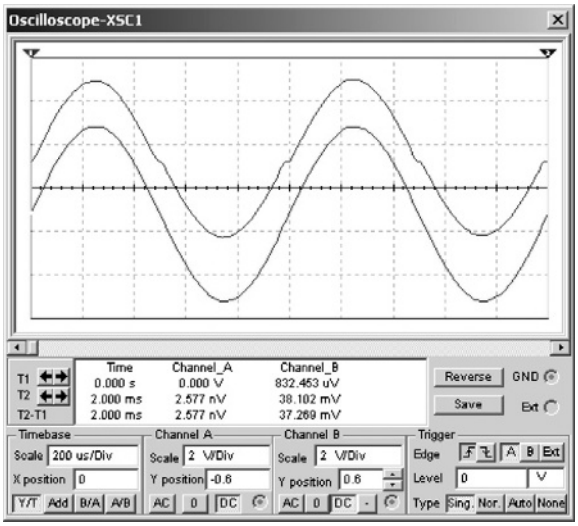


Fig. 9.1.16 Cross-over distortion evident in the output waveform from the Class B amplifier shown in Figure 9.1.3.

Nc=2Isc=0 Ikr=0 Rc=l Cjc=3.638p Mjc=.3085 Vjc=.75 Fc=.5 Cje=4.493p Mje=.2593 Vje=.75 Tr=239.5n Tf=301.2p Itf=.4 Vtf=4 Xtf=2 Rb=10)

9.1.3 Logic simulation

As well as an ability to carry out small-signal AC and transient analysis of linear circuits (see Figures 9.1.3 and 9.1.16), modern SPICE software packages usually incorporate facilities that can be used to analyse logic and also 'mixed-mode' (i.e. analog and digital) circuits. Several examples of digital logic analysis are shown in Figures 9.1.17–9.1.19.

Figure 9.1.17 shows a four-stage shift register based on J–K bistables. The result of carrying out an analysis of this circuit is shown in Figure 9.1.18.

Finally, Figure 9.1.19 shows how a simple combinational logic circuit can be rapidly 'assembled' and tested and its logical function checked.

9.1.4 Practical investigation

9.1.4.1 Objective

To investigate the use of SPICE software to analyse a DC coupled power amplifier circuit.

9.1.4.2 Components and test equipment

PC with SPICE software (e.g. Tina Pro, Multisim, or B2 Spice).

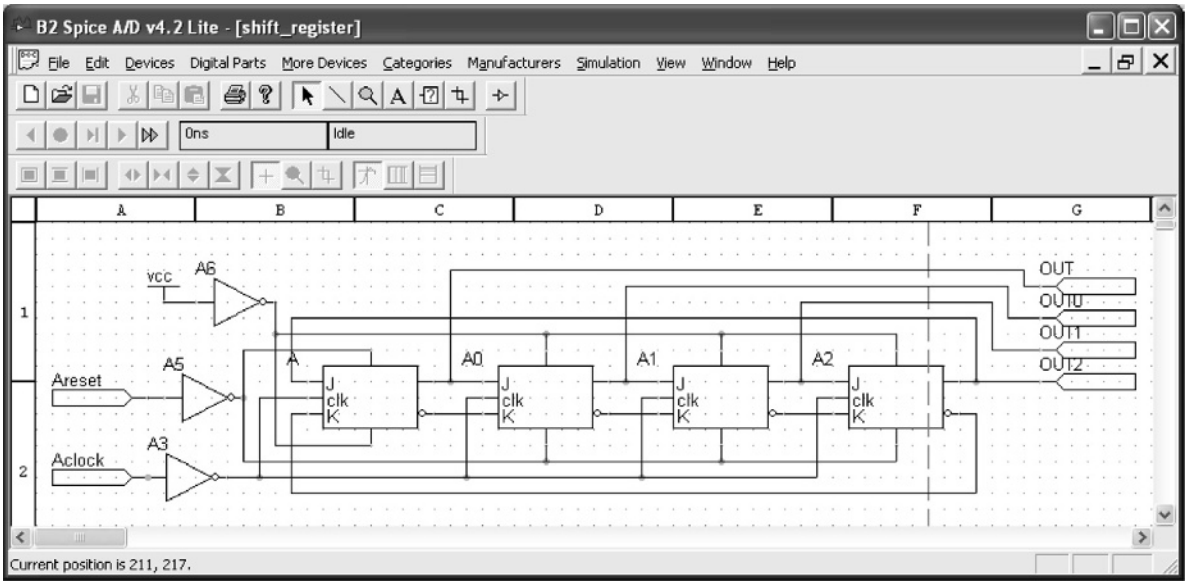


Fig. 9.1.17 Four-stage circulating shift register simulated using B2 Spice.

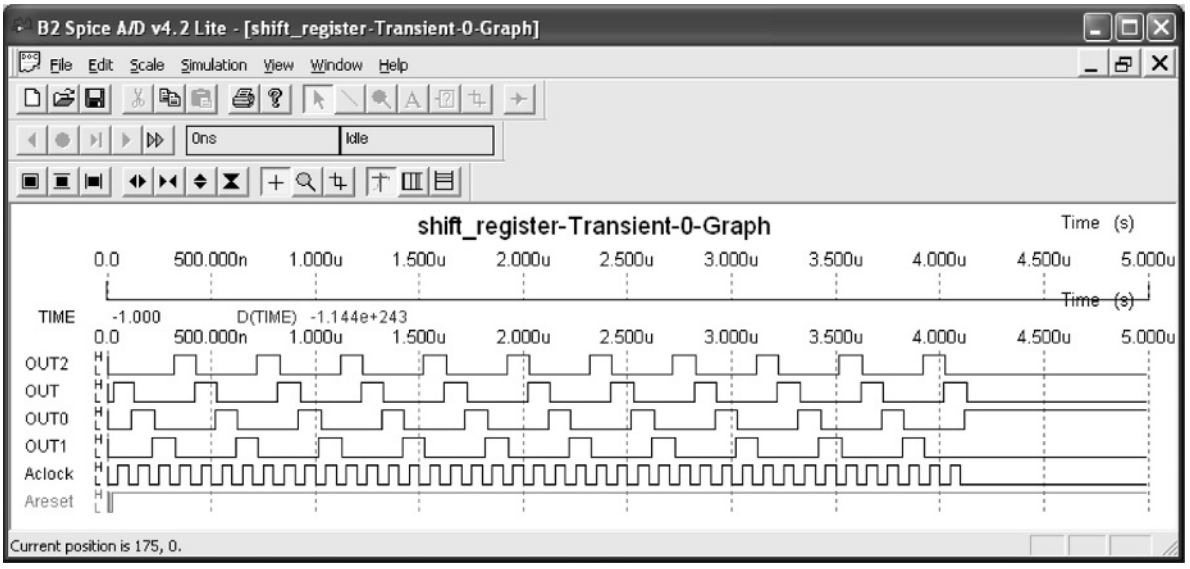


Fig. 9.1.18 Waveforms for the our stage circulating shift register in Figure 9.1.17.

9.1.4.3 Procedure

Construct the circuit shown in Figure 9.1.20 using the screen drawing facilities provided by the SPICE program. Carry out a DC analysis of the circuit and record the voltages in Table 9.1.1. Check that these are what you would expect.

Use the data from the DC analysis to determine: (a) the quiescent (no-signal) collector current in T7 and T8 and (b) the output offset voltage (i.e. the output voltage when the input voltage is zero).

Carry out an AC analysis of the amplifier and determine the voltage gain of the circuit (this should be the same as the ratio of R12 – R2). Check that the voltage gain is what you would have expected. Also determine the maximum input voltage before the output voltage becomes clipped.

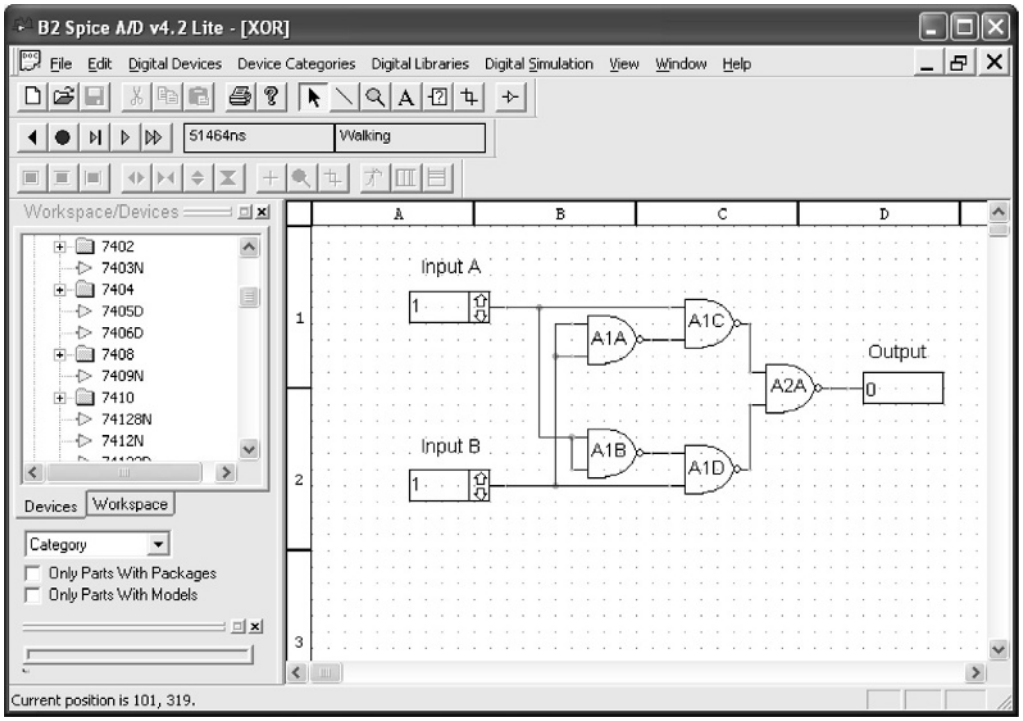


Fig. 9.1.19 Using B2 Spice to check the function of a simple combinational logic circuit.

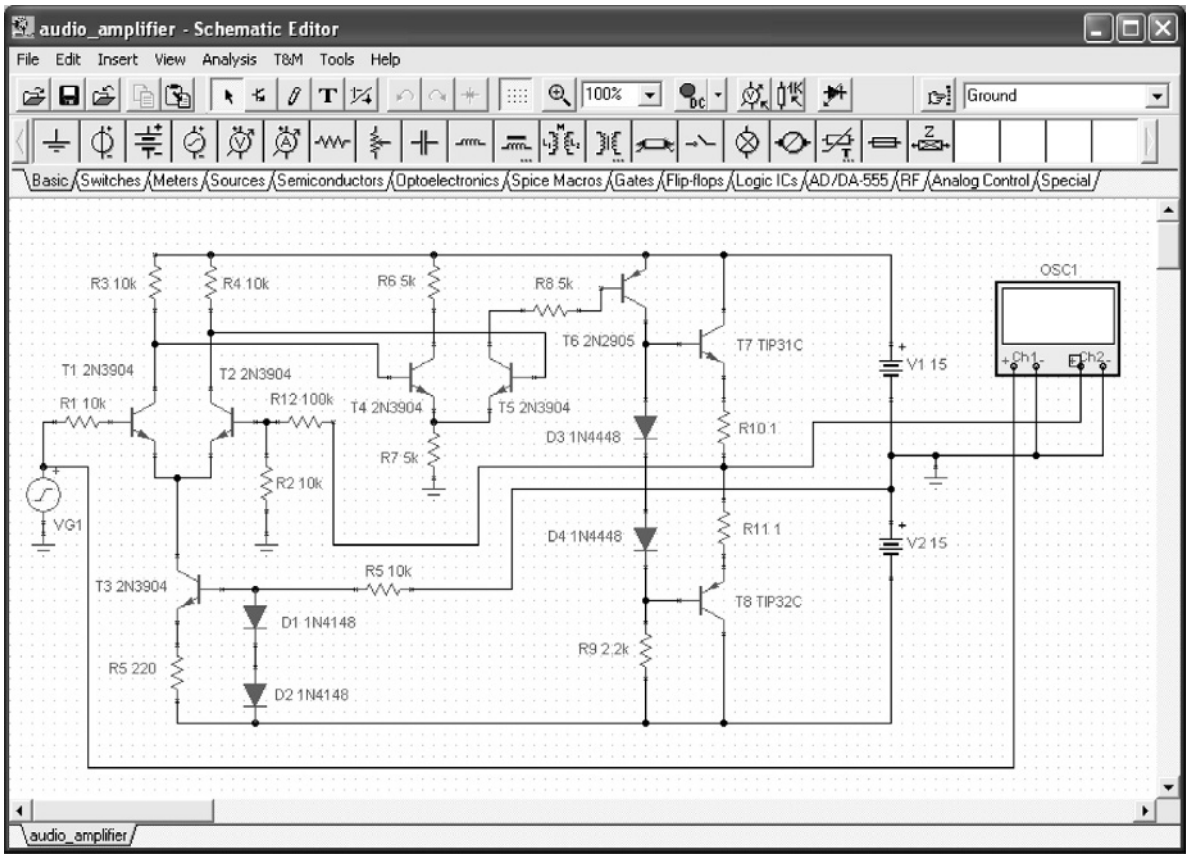


Fig. 9.1.20 DC coupled power amplifier circuit used in the Practical Investigation (Tina Pro).

Table 9.1.1 Table of results

Transistor	Collector	Base	Emitter
T1			
T2			
T3			
T4			
T5			
T6			
T7			
T8			

Chapter 9.2

9.2

The PIC microcontroller

Mike Tooley

The PIC microcontroller is a device that is normally used in stand-alone (or *embedded*) applications to perform simple logic, timing and input/output control. In recent years such devices have become increasingly sophisticated in terms of both the variety of I/O peripheral facilities provided and the ability to connect to other devices in bus and networked configurations. This latter feature makes it possible to have several microcontrollers working together, exchanging data and control information, in distributed systems that can be of virtually any size.

9.2.1 PIC architecture

PIC microcontrollers first became popular in the mid-1990s. Since then the range of devices available has increased dramatically. Nowadays a PIC device exists for almost any embedded application, from small 6-pin devices ideal for simple control applications, through to powerful high-speed devices packed with I/O features and large amounts of memory.

PICs are based on *RISC* (Reduced Instruction Set Computer) architecture and, as a consequence, they use a relatively small number of instructions. In fact, some PIC chips have as few as 33 instructions compared with some general purpose microprocessors (such as the Z80) that may have several hundred. Because it is only necessary to remember a relatively small number of commands, it is relatively easy to learn to program a PIC using its own assembly language (note, however, that it may take several instructions to achieve on a PIC what can be done in a single instruction using a fully fledged microprocessor). And, if you don't like the idea of having to learn assembly code programming you can make use of C, BASIC or Flowcode (a language based almost entirely on flowcharts).

Another feature of PIC devices is that they use *Harvard architecture*. This means that they contain completely separate memory and buses for program instructions and data memory. Program memory can be ROM, EPROM or Flash, whereas data memory must be read/write memory (commonly known as RAM).

Mainstream microprocessors (like the Z80) are categorized in terms of the number of data bits that they can manipulate (i.e. the 'width' of the data bus). They can be 8-bit, 16-bit, 32-bit and 64-bit and 128-bit devices. PIC microcontrollers, by contrast, are all based on an 8-bit data bus and, because of this, they can *only* operate on 8-bits of data at a time (despite this, you will sometimes encounter instructions that reference more data bits, i.e. 12-bits, 14-bits and 16-bits). Furthermore, although the data memory is 8-bits wide for all PIC microcontrollers, the program memory varies from 12-bits to 16-bits.

9.2.2 PIC families

In the early 1990s, PIC microcontrollers were grouped in three families, often referred to as 'Base-Line', 'Mid-Range', and 'High-End'. Many of these devices are incompatible with the latest development platforms and software tools. Despite this, it is usually possible to locate a modern device that is compatible

Electronic Circuits; ISBN: 9780750669238
Copyright © 2006 Ltd. All rights of reproduction, in any form, reserved.

with (and will generally outperform) one of these early devices.

The base-line family comprised just four devices, which all featured 12-bit instructions. The PIC16C54 and PIC16C56 were both 18-pin devices having 12 digital I/O lines. The PIC16C54 had 512 × 12-bit words of EPROM program memory, whilst the PIC16C56 had 1K × 12-bit words. Both of these devices had just 32 bytes of RAM data memory (ample for most simple control applications). The larger PIC16C55 and PIC16C57 28-pin devices had 20 digital I/O lines. The PIC16C55 had 512 × 12-bit words of EPROM program memory and 32 bytes of RAM data memory, whilst the PIC16C57 had 2K × 12-bit words of EPROM program memory and 80 bytes of RAM data memory. All of these devices also featured an 8-bit real-time clock/counter and a watchdog timer.

The mid-range family of PIC devices contained a number of devices that expanded on the capabilities of the base-line devices. Of particular note are the PIC16C71, which added A/D inputs capable of reading analog signals, and the popular PIC16C84, which featured EEPROM program and data memory. Because of the fact that the PIC16C84 could be erased electronically (without the need for the ultraviolet eraser required for EPROM erasure), it soon became a favourite with hobbyists and must have been the most popular PIC for many years.

The high-end device was the PIC17C42. This device came in a 40-pin package and featured 16-bit wide instructions. It also allowed for external memory to be interfaced and supported up to 64K × 6-bit of addressable program memory space. The device provided 33 digital I/O pins, EEPROM data memory, a serial port (USART), three 16-bit timer/counters and two high-speed pulse width modulation (PWM) outputs.

There are now more than 200 different PIC chips available and the 'Base-Line', 'Mid-Range' and 'High-End' classification has now been replaced by several different ranges of PIC families such as the PIC-10, PIC-12, PIC-16 and PIC-18. Generally speaking, the higher the family number, the more powerful the device in terms of processing capabilities, memory sizes and I/O features. Another feature of many of the devices is *Flash program memory*. These devices can be erased and reprogrammed electrically, as with the previously popular PIC16C84. These Flash memory devices are signified by a letter 'F' in their part number (e.g. PIC16F877A). Note that Flash devices are much easier to work with for one-off prototyping because erasure and reprogramming is greatly simplified.

9.2.3 Choosing a PIC device

When choosing a PIC device for a particular project it is important to select a device that is well supported, not only in terms of being a member of one of the current PIC families but also in terms of the programming and environment that you intend to use for software development. It is also important to ensure that the device incorporates all of the peripheral I/O facilities that you will need. Such facilities might include:

- Bus and communication interfaces (such as RS232/RS485, I^2C, CAN, LIN, USB, TCP/IP)
- Display peripheral interfaces (such as LED or LCD drivers)

Table 9.2.1 Some popular PIC families

PIC family	Characteristics
PIC10F	The PIC10F range features just a few 6/8-pin microcontrollers. These chips are supplied in DIP packages.
PIC12F	The PIC12F devices include the PIC12F629, PIC12F675 and PIC12F683. They are all 8-pin devices featuring six digital I/O lines and differ in the amount of program and data memory that they provide. The PIC12F683 also provides analog inputs.
PIC16F	The PIC16F family is very large and includes devices in 14-, 18-, 28- and 40-pin packages. The chips vary in terms of the amount of memory they provide and also in their I/O features. Some devices worth noting include the PIC16F630, PIC16F676, PIC16F684 and PIC16F688, which are 14-pin equivalents to the PIC12F devices mentioned above, the PIC16F84A, which is the most current equivalent to the popular PIC16C84, and the PIC16F877A, which is a powerful 40-pin device that has become extremely popular. The PIC16F877A features 8192 × 14-bit Flash program memory, 368 bytes of RAM data memory, 256 bytes of EEPROM data memory, 33 digital I/O lines, eight 10-bit A/D channels, three counter/timers and a serial port (USART)
PIC18F	The PIC18F devices form the new high-end range. One feature that they all share is that their instruction set is optimized for C compilers. Many feature large amounts of memory and special I/O features including Universal Serial Bus (USB) and Controller Area Network (CAN) interfaces. One interesting device is the 40-pin PIC18F452, which is similar in many respects and pin-compatible with the PIC16F877A. It is also worth noting that some PIC18F devices are supplied in 80-pin packages and provide more than 70 digital I/O lines. However, these are in surface-mount packages and so are difficult to experiment with!

- Capture/compare facilities
- Pulse Width Modulators (PWMs)
- Counters/timers
- Watchdog facilities
- Analog-to-digital (A/D) converters
- Analog comparators and operational amplifiers
- Brown-out detectors
- Low-voltage detectors
- Temperature sensors
- Oscillators
- Voltage references
- Digital-to-analog (D/A) converters

9.2.4 Features of the PIC16F84

The PIC16F84 is an excellent choice for most simple microcontroller projects. This device is used in many current PIC designs and it has the following features:

- 35 single-word instructions (see Table 9.2.3)
- 1K × 14-bit EEPROM (Flash) program memory
- 68 × 8-bit general purpose SRAM registers
- 15 × 8-bit special function hardware registers
- 64 × 8-bit EEPROM data memory
- 1,000,000 data memory erase/write cycles (typical)
- Data retention >40 years

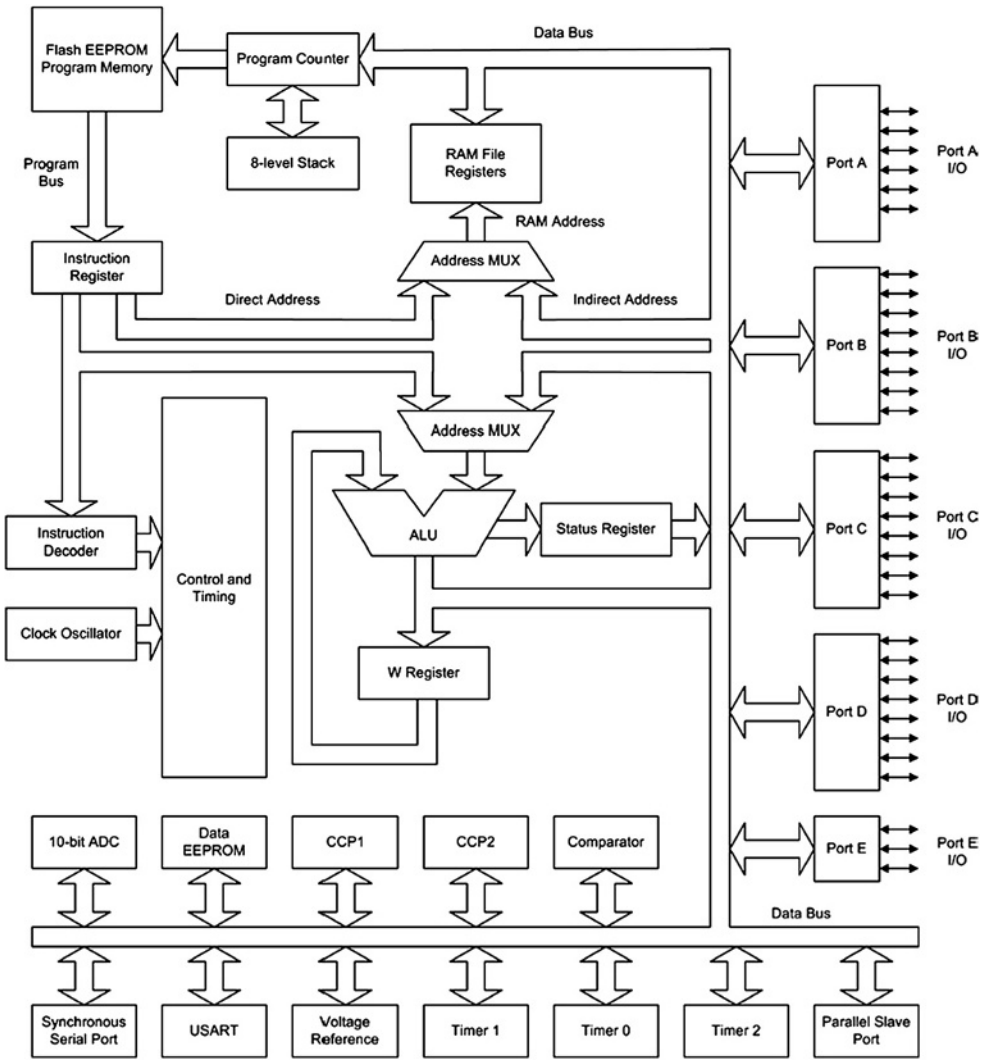


Fig. 9.2.1 Simplified architecture of a 16F877A PIC microcontroller.

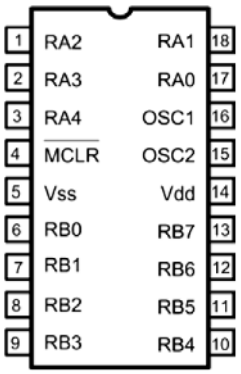


Fig. 9.2.2 Pin connections for the PIC16C84 and PIC16F84 devices.

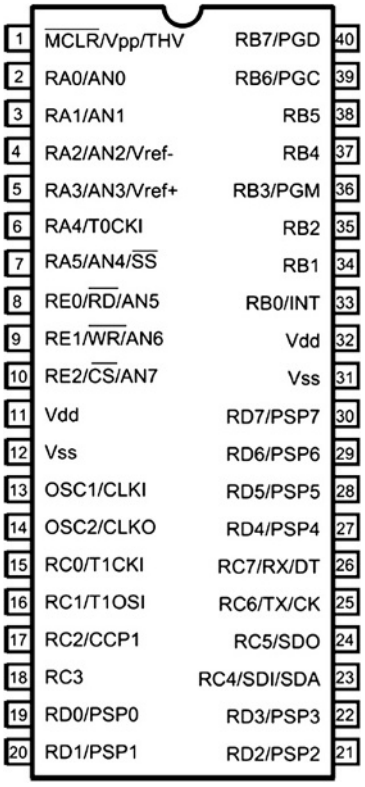


Fig. 9.2.3 Pin connections for the PIC16F877 device.

- 5 data input/output pins, Port A
- 8 data input/output pins, Port B
- 25 mA current sink max. per pin
- 20 mA current source per pin
- 8-bit timer/counter with prescaler
- Power-on reset (POR)
- Power-up timer (PWRT)
- Oscillator start-up timer (OST)
- Watchdog timer (WDT) with own on-chip RC oscillator
- Power saving 'sleep' function
- Serial in-system programming
- Selectable oscillator options:
  - RC: low-cost RC oscillator
  - XT: standard crystal resonator (100 kHz to 4 MHz)
  - HS: high-speed crystal/resonator (4–10 MHz – PIC16F84-10 only)
  - LP: power saving low-frequency crystal (32–200 kHz)
- Interrupts:
  - External, RB0/INT pin
  - TMR0 timer overflow
  - Port B RB4–RB7 interrupt on change
  - Data EEPROM write complete
- Operating voltage range: 2.0–6.0 V
- Power consumption: <2 mA at 5 V, 4 MHz
- 60 μA typical at 2 V, 32 kHz
- <1 μA typical standby at 2 V (PIC16F84)

Figure 9.2.2 shows the pin connections for the PIC16F84 whilst Figure 9.2.4 shows how the device can be used in a simple control application that uses four switched inputs (S1–S4) and three relay outputs (RL1–RL3).

9.2.5 Features of the PIC16F877

The PIC16F877 is used for many current PIC designs. This chip has the following features:

- 35 single-word instructions (see Table 9.2.3)
- 8K × 14-bit EEPROM (Flash)
- 368 bytes of general purpose SRAM registers
- 256 × 8-bit EEPROM data memory
- 1,000,000 data memory erase/write cycles (typical)
- Data retention >40 years
- 6 data input/output pins, Port A
- 8 data input/output pins, Port B
- 8 data input/output pins, Port C
- 8 data input/output pins, Port D
- 3 data input/output pins, Port E
- Two 8-bit and one 16-bit timers
- Two capture and compare PWM modules
- Universal synchronous/asynchronous receiver transmitter (USART)
- Synchronous serial port (SSP)
- Parallel slave port (PSP)
- Brown-out detection and reset facility
- 10-bit, 8-channel ADC

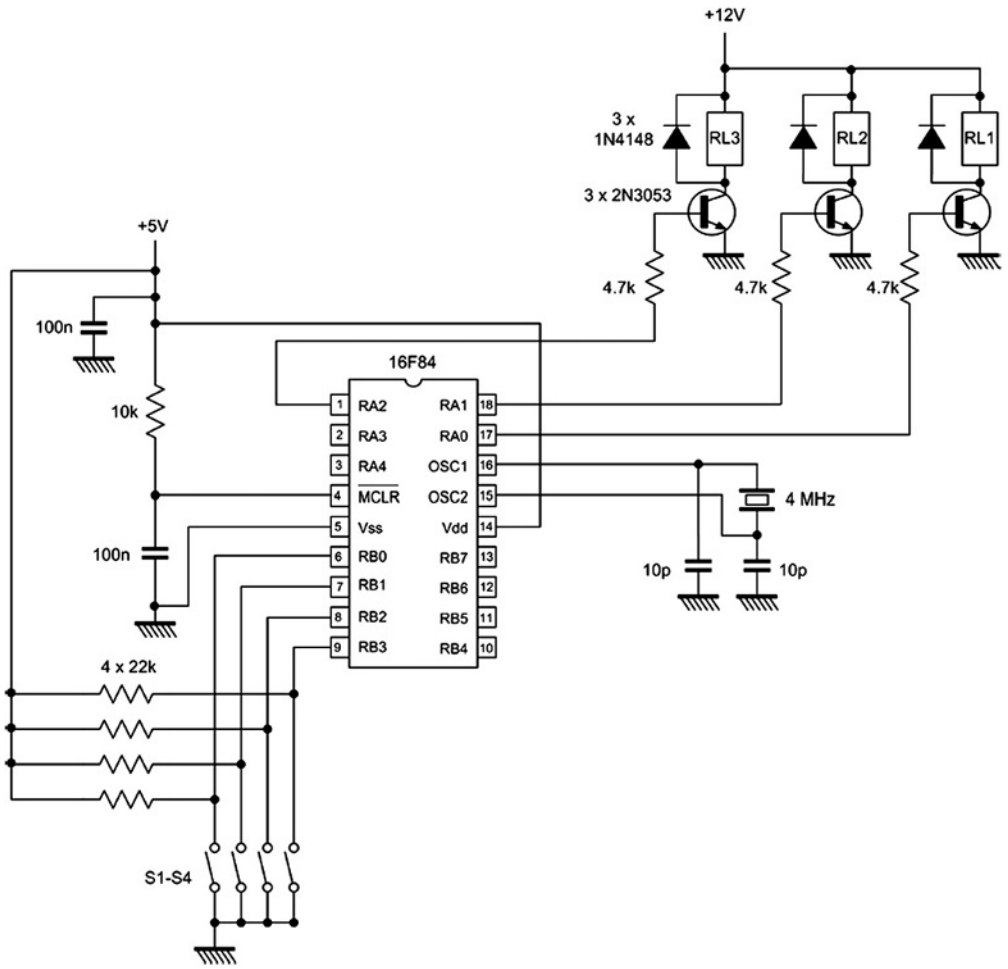


Fig. 9.2.4 A simple microcontroller application based on the PIC16F84 device with four switch inputs (S1–S4) to control three relays (RL1–RL3). The program is automatically reset at power-up.

- Two analog comparators with programmable on-chip voltage reference, programmable input multiplexing, and externally accessible comparator output.
- 25 mA current sink max. per pin
- 20 mA current source per pin
- 8-bit timer/counter with prescaler
- Power-on reset (POR)
- Power-up timer (PWRT)
- Oscillator start-up timer (OST)
- Watchdog timer (WDT) with own on-chip RC oscillator
- Power saving 'sleep' function
- In-circuit serial programming (via two pins)
- Single supply 5 V in-circuit serial programming
- Selectable oscillator options
- Self-programmable under software control
- 15 interrupts
- Operating frequency: DC to 20 MHz
- Supply voltage range: 2.0–5.5 V
- Low power consumption

9.2.6 PIC programming

PIC programming can be carried out using assembly language, C (see Table 9.2.2), BASIC, and Flowcode or a mixture of these languages. Software and a programmer (see Figures 9.2.6 and 9.2.7) will be required to do this. The two most commonly used methods of programming a PIC chip are shown in Figure 9.2.8. In Figure 9.2.8(a) a dedicated PIC programmer is used whilst in Figure 9.2.8(b) the PIC is programmed whilst resident in the target system. In either case, the programming software and source code is resident on a PC and downloaded as *hex code* into the PIC. The process of generating the source code, compiling and/or assembling

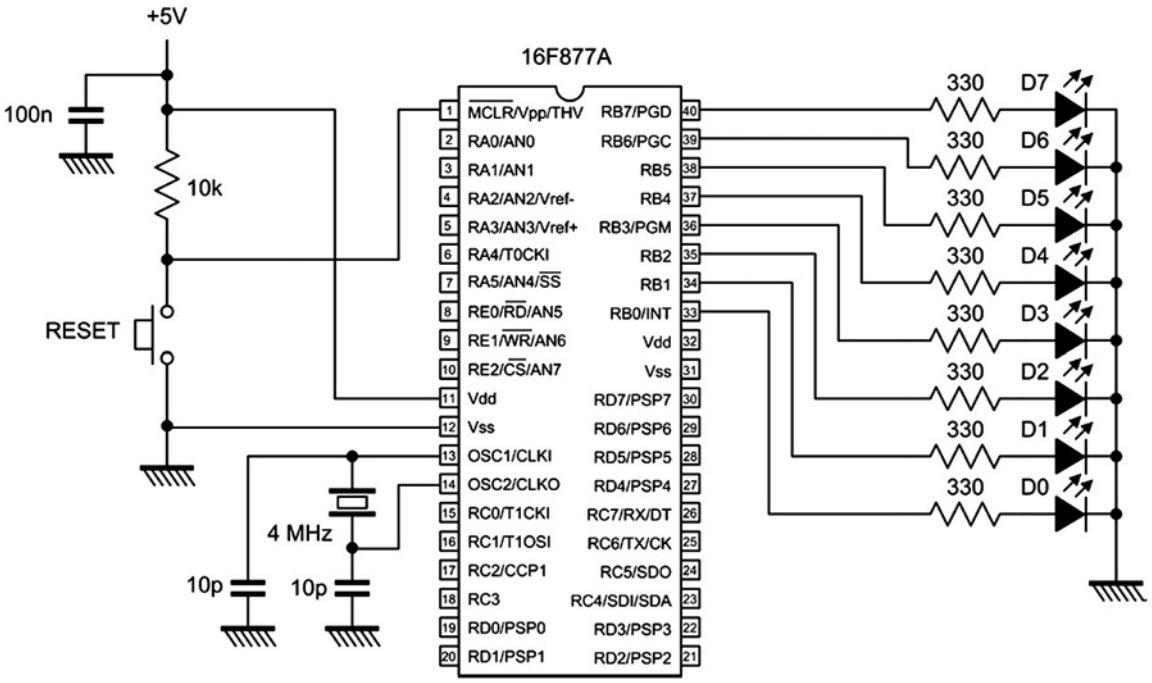


Fig. 9.2.5 A simple microcontroller application based on the 16F877A device. This application uses eight LED outputs (D0–D7). Note that the circuit must be manually reset at power-up.

it into hex, code (see Figure 9.2.10) is invariably performed by software known as an *Integrated Development Environment* (IDE), as shown in Figure 9.2.9. Software is also available that can be used to simulate a PIC (see Figures 9.2.11–9.2.13) and full development systems are also available (see Figures 9.2.14 and 9.2.15).

9.2.7 Flowcode

Flowcode is a novel programming system that is designed to allow those new to PIC programming to develop complete applications with little or no knowledge of any higher level languages. However it is possible to embed code written in other languages into Flowcode programs.

Programs written in C and Assembly code can be easily embedded in Flowcode. This will allow you to incorporate more complex and sophisticated code in applications where this might become necessary. Note however that the embedded code cannot be checked or simulated by Flowcode but instead it will be passed on to the C compiler when the flowchart is compiled. It is therefore important to verify that the C code entered is correct, as syntax errors will cause the compilation of the whole flowchart to fail.

To access Flowcode variables, macro functions and connection points, it is necessary to colour the variable in your C code with the prefixes FCV_, FCM_ and FCC_MacroName_, respectively. For example, to use a Flowcode variable called DELAY in your C code, you must refer to it using FCV_DELAY. Note that all Flowcode defined variables must be entered in upper case.

To call a Flowcode macro called TEST in your C code, you must call FCM_TEST();. To jump to a connection point called A, defined in a Flowcode macro called TEST, your C code must be goto FCC_TEST_A;. Connection points defined in the main flowchart of a Flowcode file are prefixed FCC_Main_.

Assembly code (see Table 9.2.3) can be added into the code field in a C assembly code wrapper. For a single line of code you can use the asm operator in front of each instruction (e.g. asm movlw 5) or you can enclose several statements within an asm block. This makes adding assembly language routines very simple.

9.2.8 Practical investigation

9.2.8.1 Objective

To investigate the operation of a PIC16F877A microcontroller using a simulator to run a simple program.

9.2.8.2 Components and test equipment

Oshonsoft PIC Simulator or similar virtual Integrated Development Environment (IDE). The Oshonsoft IDE can be downloaded from www.Oshonsoft.com.

Table 9.2.2 Switch and LED PIC test routine in C

```
//Defines for the PIC microcontroller
char PORTC@0x07;
char TRISC@0x87;
char PORTD@0x08;
char TRISD@0x88;
char PORTE@0x09;
char TRISE@0x89;

//PIC Functions
#pragma CLOCK_FREQ 4000000
#define P16F877A
#include <system.h>
#define MX_EE
#define MX_EE_TYPE2
const char MX_EE_SIZE = 256;
#define MX_SPI
#define MX_SPI_C
#define MX_SPI_SDI 4
#define MX_SPI_SDO 5
#define MX_SPI_SCK 3
#define MX_UART
#define MX_UART_C
#define MX_UART_TX 6
#define MX_UART_RX 7

//Macro function declarations

//Variable declarations
char FCV_SWITCHSTATE;

//Macro implementations


void main ()
{
//PIC Initialisation
adconl = 0x07;

//Interrupt initialisation code
option_reg = 0xC0;

while(1)
  {
     TRISA = TRISA | 0xff;
     FCV_SWITCHSTATE = PORTA;
     TRISB = 0x00;
     PORTB = FCV_SWITCHSTATE;
  if ((FCV_SWITCHSTATE) = = 0)
     break;
  }
mainendloop: goto mainendloop;
}
```

9.2.8.3 System configuration and required operation

The program is to simulate the operation of a PIC16F877A microcontroller and is to read an analog voltage present on the AN0 analog input and display the 8-bit binary conversion result on PORTB.

The analog-to-digital converter (ADC) BASIC source code used in the Practical investigation (shown in Table 9.2.4) is available for downloading from the companion website.

9.2.8.4 Procedure

Start the PIC Simulator (see Figure 9.2.17), click on Options and then Select Microcontroller. Select the 'PIC16F877' from the list and click on the Select button. Next click on File and Load Program. Select the file named adc.hex (generated as a result of compiling the BASIC source code shown in Table 9.2.4) and click on Open. This will load the program into the PIC's program memory.

Click on Tools and Microcontroller View in order to open the window shown in Figure 9.2.18). If necessary,

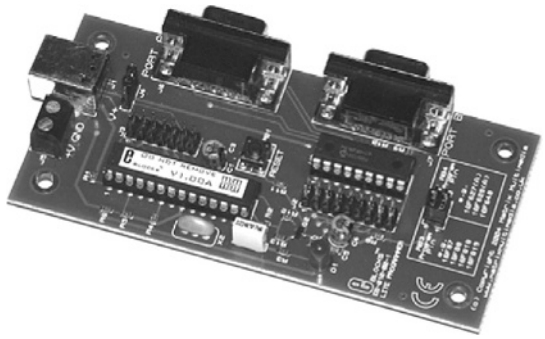


Fig. 9.2.6 The E-blocks low-cost PIC microcontroller 'lite' programmer. This device is suitable for programming 18-pin PIC chips and it connects to a PC via a USB port. The board has two I/O ports for connecting to a wide variety of external peripheral modules.

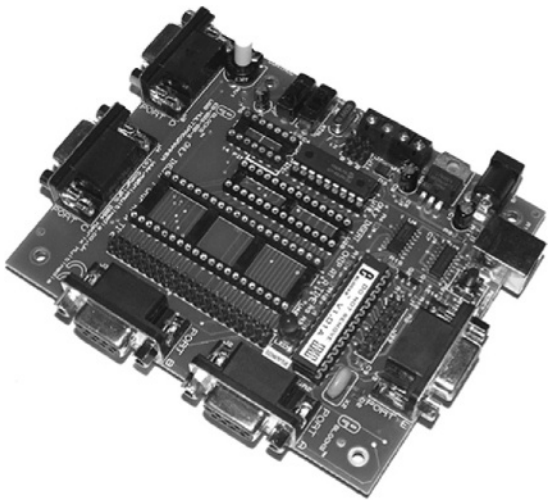


Fig. 9.2.7 E-blocks PIC microcontroller/microprogrammer. This versatile device connects to a PC via USB and can be used to program most 8-, 14-, 28- and 40-pin PIC chips. The board has five I/O ports for connection to a wide variety of external peripheral modules.

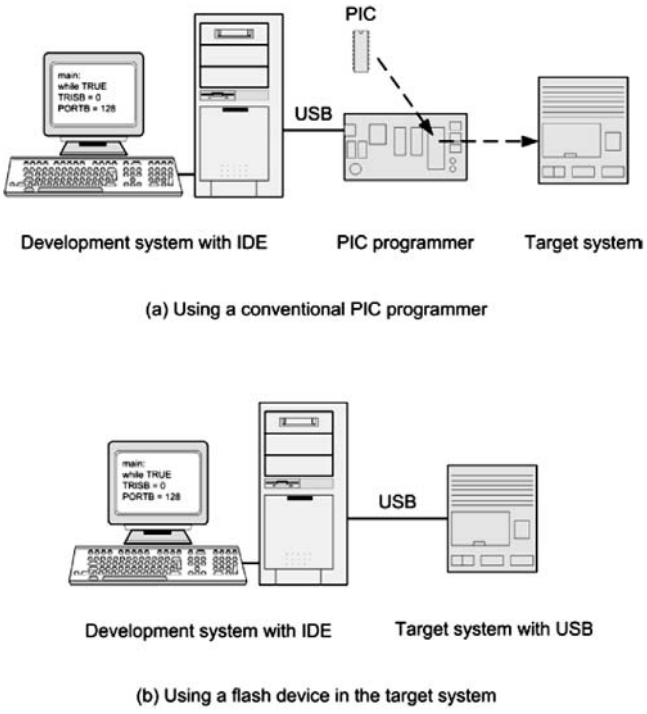


Fig. 9.2.8 Two methods of programming a PIC device.

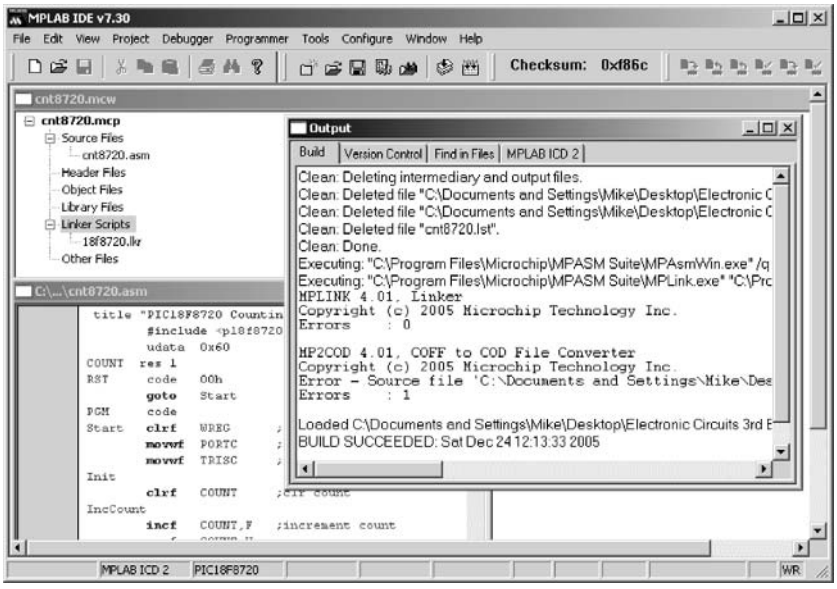


Fig. 9.2.9 The MPLAB Integrated Development Environment (IDE).

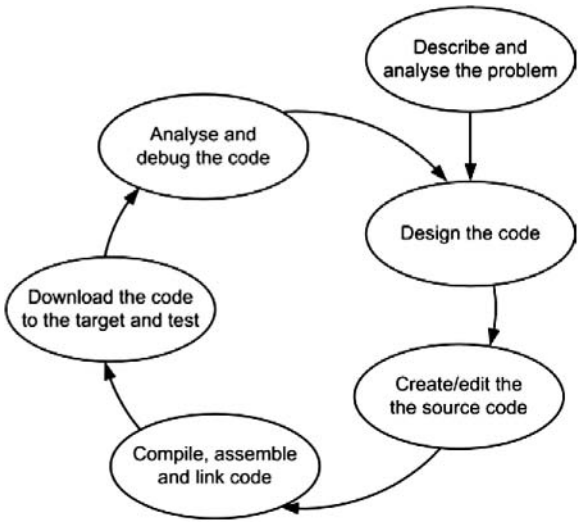


Fig. 9.2.10 The PIC software development cycle.

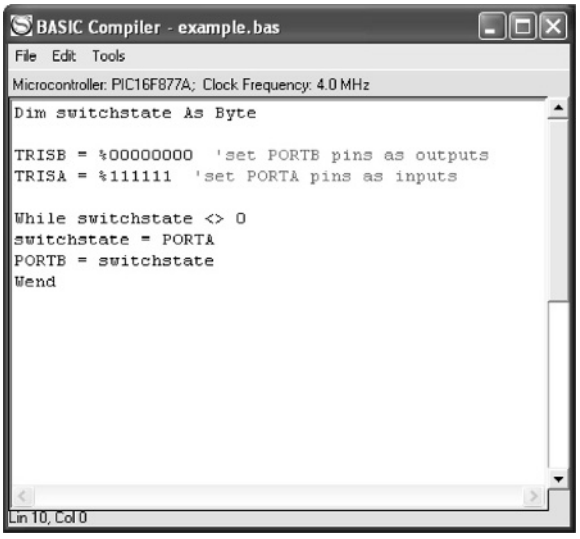


Fig. 9.2.12 The switch and LED test program using the Oshonsoft PIC Simulator.

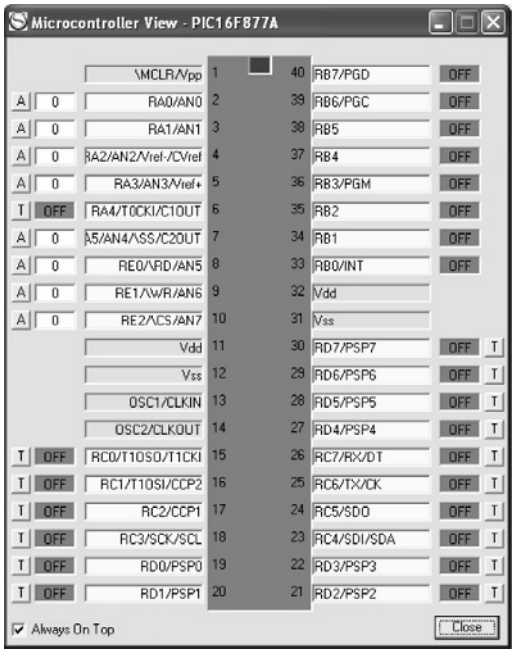


Fig. 9.2.11 The Oshonsoft PIC Simulator showing the Microcontroller View window in which the current state of each pin is displayed.

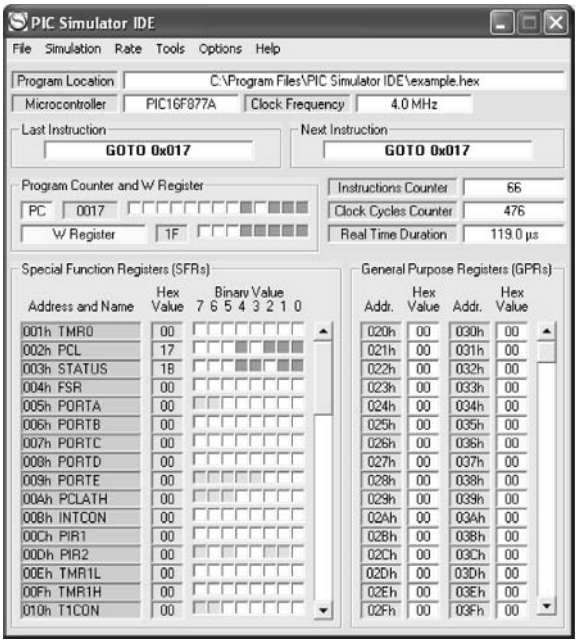


Fig. 9.2.13 The main Oshonsoft PIC Simulator screen showing the state of the Special Function Registers (SFRs) and General Purpose Registers (GPRs). Note also the contents of the Program Counter and W Register.

Table 9.2.3 PIC16C/F84 instruction set

Command	Operation	Syntax	Status register bit(s) affected
ADDLW	ADD Literal to W	ADDLW K	C, DC, Z
ADDWF	ADD W to F	ADDWF F, D	C, DC, Z
ANDLW	AND Literal and W	ANDLW K	Z
ANDWF	AND W with F	ANDWF F, D	Z
BCF	Bit Clear F	BCF F,B	None
BSF	Bit Set F	BSF F,B	None
BTFSC	Bit Test, Skip if Clear	BTFSC F,B	None
BTFSS	Bit Test, Skip if Set	BTFSC F,B	None
CALL	Subroutine Call	CALL K	None
CLRF	Clear F	CLRF F	Z
CLRW	Clear W register	CLRW	Z
CLRWDT	Clear Watchdog Timer	CLRWDT	TO, PD
COMF	Complement F	COMF F,D	Z
DECF	Decrement F	DECF F,D	Z
DECFSZ	Decrement F, Skip if 0	DECFSZ F,D	None
GOTO	Unconditional branch	GOTO K	None
INCF	Increment F	INCF F,D	Z
INCFSZ	Increment F, Skip if 0	INCFSZ F,D	None
IORLW	Inclusive OR Literal with W	IORLW K	Z
IORWF	Inclusive OR W with F	IORWF F,D	Z
MOVF	Move F	MOVF F,D	Z
MOVLW	Move Literal to W	MOVLW K	None
MOVWF	Move W to F	MOVWF F	None
NOP	No Operation	NOP	None
RETFIE	Return from Interrupt	RETFIE	None
RETLW	Return with Literal in W	RETLW K	None
RETURN	Return from Subroutine	RETURN	None
RLF	Rotate Left through Carry	RLF F,D	C
RRF	Rotate Right through Carry	RRF F,D	C
SLEEP	Power-down	SLEEP	TO, PD
SUBLW	Subtract W from Literal	SUBLW K	C, DC, Z

Table 9.2.3 PIC16C/F84 instruction set—cont'd

Command	Operation	Syntax	Status register bit(s) affected
SUBLW	Subtract W from F	SUBWF F,D	C, DC, Z
SWAPF	Swap F	SWAPF F,D	None
XORLW	Exclusive OR Literal with W	XORLW K	Z
XORWF	Exclusive OR W with F	XORWF F,D	Z

Abbreviations

C Carry
b Bit address
C Carry
d Destination
DC Digit Carry
f Register File Address
k Literal or label
PD Power-down
TO Time-out
W Working Register
Z Zero

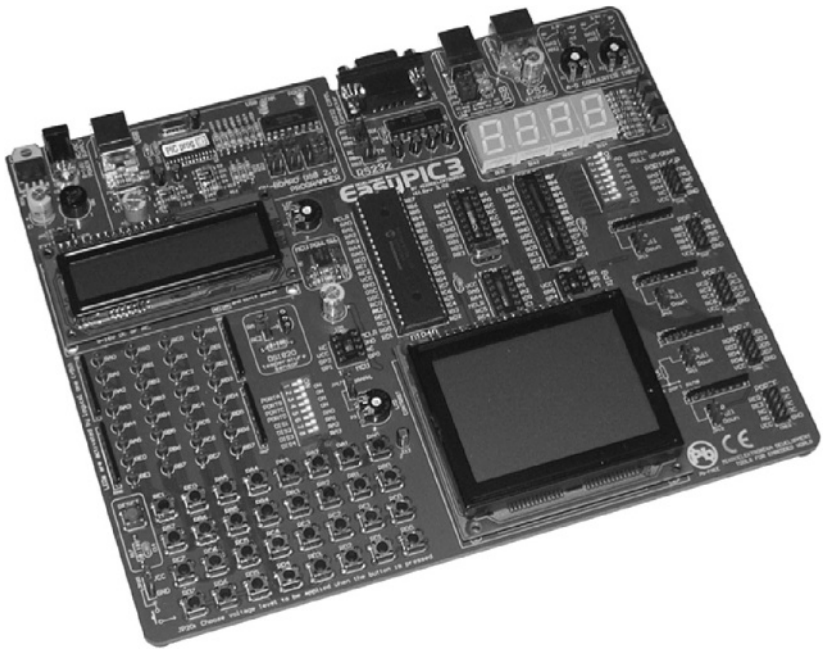


Fig. 9.2.14 The MicroElektronika EasyPIC3 PIC development system. This system supports the PIC10F, 12F, 16F and 18F families and is connected to a PC via a USB port.

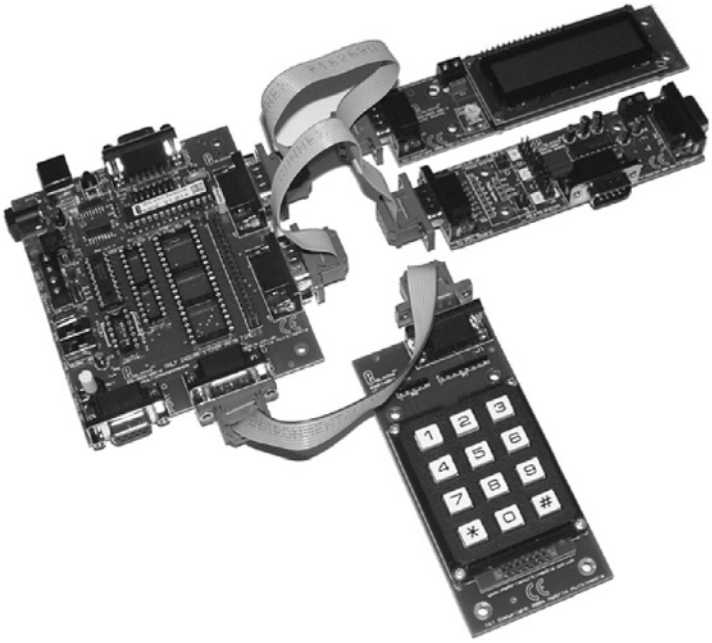


Fig. 9.2.15 The E-blocks system. The microcontroller system shown here uses a PIC programmer (left), keypad (bottom), LCD display (top) and serial communications interface (right).

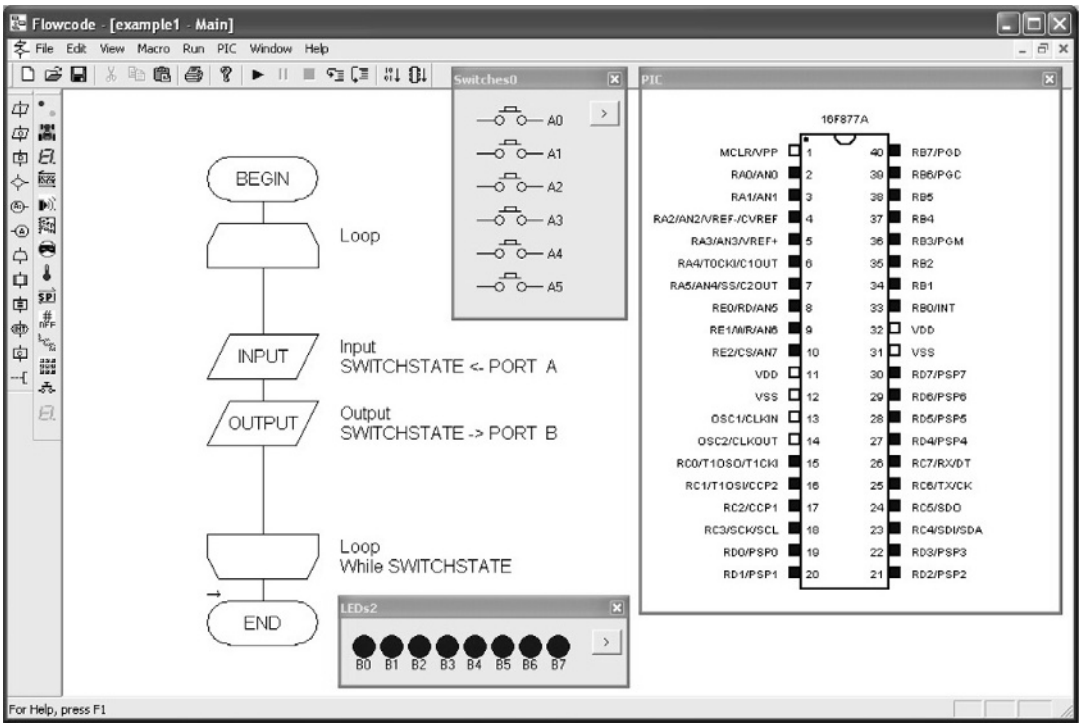


Fig. 9.2.16 The switch and LED test program using Flowcode. Note how the flow chart symbols are simply dragged and dropped into the main program window from the toolbar on the left.

Table 9.2.4 ADC program used in the practical investigation

```
Symbol ad action = ADCON0.GO_DONE 'set new name for A/D conversion start bit
Symbol display = PORTB 'set new name for PORTB used to display the conversion result

TRISB = %00000000 'set PORTB pins as outputs
TRISA = %111111 'set PORTA pins as inputs
ADCON0 = 0 × c0 'set A/D conversion clock to internal source
ADCON1 = 0 'set PORTA pins as analog inputs
High ADCON0.ADON 'turn on A/D converter module

main:
Gosub getadresult 'go to conversion routine
display = ADRESH 'display the result of the conversion
Goto main 'repeat forever
End

getadresult: 'conversion routine
High ad_action 'start the conversion
While ad_action 'wait until conversion is completed
Wend
Return
```

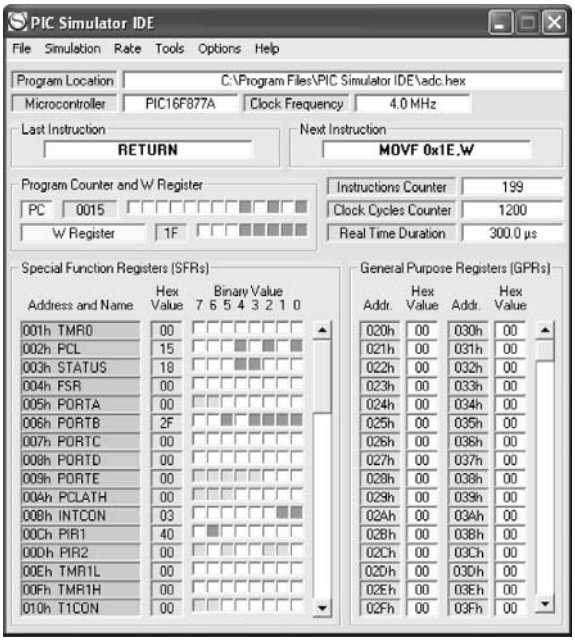


Fig. 9.2.17 PIC Simulator IDE used in the Practical investigation. This screen shows the state of the PIC with the Program Counter at 0015.

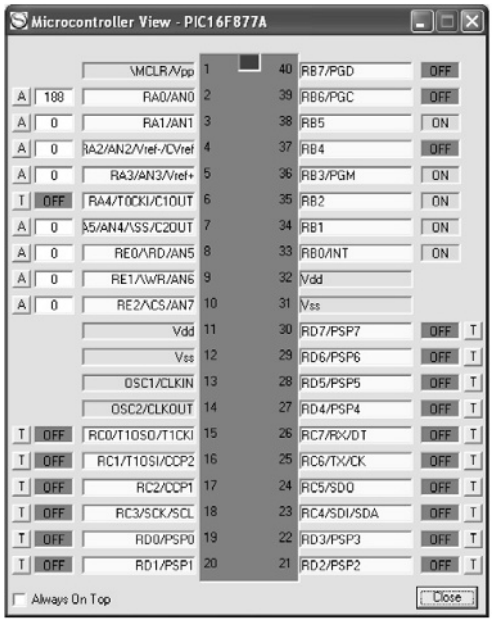


Fig. 9.2.18 Microcontroller View showing the state of each of the PIC's pins during program execution.

before clicking on Simulation and Start. The simulation will start immediately.

To set an analog input voltage, click on the analog (A) button associated with the RA0/AN0 pin then use the slider to change the analog value on this pin and click on Accept button to accept the value. The state of the PORTB pins should immediately reflect this value. Repeat with several different analog input values and at reduced simulation speeds in order to examine the effect of each program instruction.

9.2.8.5 A further program

If you have been able to run the simple ADC program successfully you might like to try your hand at entering, compiling and running a program that will loop continuously reading the state of the switches on Port A and output the result to Port B (see Figure 9.2.12).

This page is intentionally left blank

Chapter 9.3

9.3

Introduction to EMC

Tim Williams

9.3.1 What is EMC?

Electromagnetic interference (EMI) is a serious and increasing form of environmental pollution. Its effects range from minor annoyances due to crackles on broadcast reception, to potentially fatal accidents due to corruption of safety-critical control systems. Various forms of EMI may cause electrical and electronic malfunctions, can prevent the proper use of the radio frequency (RF) spectrum, can ignite flammable or other hazardous atmospheres, and may even have a direct effect on human tissue. As electronic systems penetrate more deeply into all aspects of society, both the potential for interference effects and the potential for serious EMI-induced incidents increase.

Electromagnetic *compatibility* (EMC), then, is the absence of effects due to EMI. The definition of EMC, as it appears in the International Electrotechnical Vocabulary is:

> *The ability of a device, equipment or system to function satisfactorily in its electromagnetic environment without introducing intolerable electromagnetic disturbance to anything in that environment.*

Some reported examples of electromagnetic incompatibility are:

- in Norfolk, various makes of car would "go crazy" when they passed a particular air defence radar installation – dashboard indicators dropping to zero or maximum, lights and engines cutting out;
- on one type of car, the central door locking and electric sunroof would operate when the car's mobile transmitter was used;
- new electronic push-button telephones installed near the Brookmans Park medium wave transmitter in North London were constantly afflicted with BBC radio programmes;
- interference to aeronautical safety communications at a US airport was traced to an electronic cash register a mile away;
- the instrument panel of a well-known airliner was said to carry the warning "ignore all instruments while transmitting HF";
- electronic point-of-sale units used in shoe, clothing and optician shops (where thick carpets and nylon-coated assistants were common) would experience lock up, false data and uncontrolled drawer openings;
- when a piezo-electric cigarette lighter was lit near the cabinet of a car park barrier control box, the radiated pulse caused the barrier to open and drivers were able to park free of charge;
- lowering the pantographs of electric locomotives at British Rail's Liverpool Street station interfered with newly installed signalling control equipment, causing the signals to "fail safe" to red;
- a digital TV set-top box initiated an air–sea rescue operation in Portsmouth harbour by creating an emission on the distress frequency;
- two Navy warships nearly collided when the radar transmissions of the frigate HMAS Anzac disabled the steering of the minehunter HMAS Huon, Huon passing ahead of Anzac "at close range".

Many other examples have been collected over the years; the "Banana Skins" column in the *EMC Journal*, collated by Keith Armstrong, is a fruitful source, and the EMC group of the former UK Radiocommunications Agency commissioned an EMC Awareness web page introducing

EMC for Product Designers; ISBN: 9780750681704
Copyright © 2007 Elsevier Ltd. All rights of reproduction, in any form, reserved.

the subject, which also contains a number of examples. Here are a few issues in more detail.

9.3.1.1 Portable electronic devices in aircraft

Mobile cellular telephones are rapidly establishing themselves, through their sheer proliferation, as a serious EMC threat. Passengers boarding civil airliners are now familiar with the announcement that the use of such devices is not permitted on board. They may be less familiar with why this is regarded as necessary. The IFALPA International Quarterly Review reported 97 EMI-related events due to passenger "carry-on" electronic devices since 1983. To quote the Review:

> *…By 1990, the number of people boarding aeroplanes with electronic devices had grown significantly and the low-voltage operation of modern aircraft digital electronics were potentially more susceptible to EMI.*
>
> *A look at the data during the last ten years indicates that the most likely time to experience EMI emissions is during cruise flight. This may be misleading, however. During the last three years, 43% of the reported events occurred in cruise flight while an almost equal percentage of events occurred in the climb and approach phases.*
>
> *Of particular note: during the last three years the number of events relating to computers, compact disc players, and phones has dramatically increased and these devices have been found to more likely cause interference with systems which control the flight of the aircraft.*
>
> *Recognising an apparent instrument or autopilot malfunction to be EMI related may be difficult or impossible in many situations. In some reported events the aircraft was off course but indications in the cockpit displayed on course. Air traffic controllers had to bring the course deviations to the attention of the crews. It is believed that there are EMI events happening that are not recognized as related to EMI and therefore not reported.*

Particular points noted by the Review were that:

- events are on the rise;
- all phases of flight are exposed (not just cruise);
- many devices may cause EMI (phones, computers, CD players, video cameras, stereos);
- often there will be more than one device on a flight;
- passengers will turn on a device even after being told to turn it off[1];
- passengers will conceal usage of some devices (phones, computers);
- passengers will turn devices on just after takeoff and just prior to landing;
- phones are a critical problem;
- specific device type and location should be recorded and reported by the crew;
- when the emitting EMI device is shut off, the aircraft systems return to normal operation (in the case of positioning errors a course change may be necessary);
- flight attendants should be briefed to recognize possible EMI devices.

In 2000, the Civil Aviation Authority carried out tests on two aircraft parked at Gatwick, which reinforces the ban on the use of mobile phones while the engine is running. The tests revealed that interference levels varied with relatively small changes in the phone's location, and that the number of passengers on the flight could affect the level, since they absorbed some of the signal. Further testing showed that at the GSM mobile frequencies it was possible to create the following interference effects:

- compass froze or overshot actual magnetic bearing;
- instability of indicators;
- digital VOR (VHF Omnidirectional Ranging, an aeronautical navigation aid using the VHF spectrum) navigation bearing display errors up to 5 degrees;
- VOR navigation To/From indicator reversal;
- VOR and ILS (Instrument Landing System, an aeronautical navigation aid using the VHF spectrum) course deviation indicator errors with and without a failure flag;
- reduced sensitivity of the ILS Localizer receiver;
- background noise on audio outputs.

Nevertheless, there is considerable public pressure to allow use of cellphones on board, and the fact that more often than not interference *doesn't* actually create problems has led to the perception that there *isn't* a problem. To deal with both of these issues, some airlines are trialling a system which actually allows cellphones to be used via a picocell base station on the

[1] Especially if they regard their need for personal communication as more important than a mere request from the crew. An aircraft carrying a German Foreign Minister had to make an emergency landing "after key cockpit equipment cutout" as noted in one of the reports. It was claimed that mobile phone transmissions could be the only explanation and it was said that, "despite repeated requests from the crew, there were still a number of journalists and foreign office personnel using their phones".

aircraft, communicating via satellite with the ground networks. The EMC implication of this is that because the base station is very local, the phones are able to transmit at their minimum power, thus (hopefully) eliminating EMC interactions. Even so, there will almost certainly be restrictions on use, only above 10,000 feet and not during takeoff and landing.

9.3.1.2 Interference to medical devices

Another critical area with potentially life-threatening consequences is the EMC of electronic medical devices. A 1995 review article described three incidents in detail and listed more than 100 EMI problems that were reported to the US Food & Drug Administration between 1979 and 1993. It states bluntly that:

> *EMI-related performance degradation in electronic medical devices has resulted in deaths, serious injuries, and the administration of inappropriate and possibly life-threatening treatment.*

The detailed case studies were as follows:

- *Apnea monitors*: The essential function of an apnea monitor is to sound an alarm when breathing stops; the devices are used in hospitals and frequently prescribed for home use in the case of infants who either have exhibited or are at risk of experiencing prolonged apnea. After numerous reports of unexplained failure on the part of apnea monitors to alarm even upon death, their susceptibility to radiated RF was evaluated by the US Center for Devices and Radiological Health (CDRH). Most commercial apnea monitors were found to erroneously detect respiration when exposed to relatively low-field strengths, a situation that could result in failure to alarm during apnea. Most monitors were found to be susceptible above 1 V/m; one particular model was susceptible to pulsed fields above 0.05 V/m.
- *Anaesthetic gas monitor*: The CDRH received several reports of erroneous displays and latch-up of an anaesthetic gas monitor during surgery. None of the reports mentioned EMI as a possible cause. FDA investigators found that the manufacturer had a list of 13 complaint sites, and his own investigations revealed that interference from certain types of electrosurgery units disrupted the communication link between the monitor and a central mass spectrometer, causing the monitor to fail to display the concentration of anaesthetic gas in the operating room during surgery.
- *Powered wheelchairs*: A QA manager at a large wheelchair manufacturer had received reports of powered wheelchairs spontaneously driving off kerbs or piers when police or fire vehicles, harbour patrol boats, or CB or amateur radios were in the vicinity. Though CDRH databases showed reports of unintended motion – in several cases involving serious injury – none of these incidents had been attributed to EMI. When CDRH investigated the EMI susceptibility of the motion controllers on various makes of powered wheelchairs and scooters, they discovered susceptibilities in the range of 5–15 V/m. At the lower end of the range, the electric brakes would release, which could result in rolling if the chair happened to be stopped on an incline; as the field strength at a susceptible frequency was increased, the wheels would actually begin turning, with the speed being a function of field strength.

Another issue is the effect on hearing aids:

> *The problem of interference to hearing aids has been known for some time. Digital mobile phones use a form of radio transmission called Time Division Multiple Access (TDMA), which works by switching the radio frequency carrier rapidly on and off. If a hearing aid user is close to a digital mobile telephone, this switching of the radio frequency carrier may be picked up on the circuitry of the hearing aid. Where interference occurs, this results in a buzzing noise which varies from very faint to maximum volume of the aid… [A specialist standards panel] has determined that, although digital mobile telephones are being looked at as the source of likely interference, all radio systems using TDMA or similar transmissions are likely to cause some interference.*
>
> BSI News December 1993

These are all examples of the lack of a product's "fitness for purpose": that is, to operate correctly and safely in its intended environment, which includes the electromagnetic environment. There are clear safety implications in the reports.

9.3.1.2.1 Hospital and emergency service radio management

Many types of hospital equipment are susceptible to RF radiation from hand-portable mobile radio transmitters – diagnostic equipment such as ECGs, EEGs, pulse oximeters and other physiological monitoring equipment; and therapeutic equipment such as infusion pumps, ventilators and defibrillators. Physiological (patient-coupled) monitoring equipment is very sensitive and hence very susceptible, although for every device type, some models consistently perform better than average (they exhibit good EMC design). The type of modulation employed by the mobile transmitter can be

significant. For example, an external pacemaker withstood a GSM signal (modulated at 217 Hz) at 30 V/m field strength, but TETRA modulation (17 Hz) caused interference at 3 V/m.

This is of particular concern for ambulances, which in Europe are mandated to use the TETRA system for emergency communications, but which also carry an array of patient-coupled instrumentation for life support purposes. This has led to the UK's Medicines and Healthcare Products Regulatory Agency (MHRA) recommending as follows:

- the use of portable handsets and cellphones inside ambulances should be restricted;
- special precautions are needed if a patient with an external pacemaker is being transported;
- displaying warning notices, providing staff training, and relocating parking bays are possible actions if risks of interference prove unacceptable when emergency vehicles are parked immediately outside patient treatment areas;
- caution should be exercised when treating patients with medical devices at the scene of an accident if an emergency vehicle is nearby;
- mobile data terminals should be subjected to any restrictions which are locally applied to cellphones.

Various studies have tested medical devices and recommend that a distance of 1–1.5 m be maintained between typical hand-portable transmitters and medical equipment. The MHRA tested 178 different models of medical device using a wide range of radio handsets. Overall, in 23% of tests medical devices suffered EMI from handsets. Forty-three per cent of these interference incidents would have had a direct impact on patient care, and were rated as serious. Only 4% exhibited effects with cellphones at 1 m distance, although at that distance emergency and security handsets had much greater effects.

The difficulty with controlling the use of radio communications in hospitals and other medical situations is well illustrated in the MHRA's guidance document, which itself refers to an ISO technical report on the subject:

> *Overly-restrictive policies may act as obstacles to beneficial technology and may not address the growing need for personal communication of patients, visitors and the workforce. At the other extreme, unmanaged use of mobile communications can place patients at risk.*

The guidance stresses the need for an effective policy for healthcare providers to manage the use of the RF spectrum in their own sites. This includes considering areas where medical devices will not be affected and therefore no restrictions apply, and other areas where authorized staff can use communication devices authorized by the hospital. Incidents should be reported when a medical device is suspected to have suffered EMI.

9.3.1.2.2 Diathermy and electrosurgery

As well as radio communications, medical diathermy and electrosurgery are well known as a source of significant interference problems that most surgeons simply learn to cope with. Medical diathermy (tissue heating) used for physiotherapy typically operates at 27 MHz with RF powers up to 400 W, although modern pulsed diathermy uses average RF powers around 40 W; but these levels are more than enough to interfere with many kinds of electro-medical equipment, particularly monitors.

9.3.1.3 Thermostats

Thermostats and other automatic switching contacts of all sorts are a major source of noise complaints, particularly when they are faulty. The former UK Radiocommunications Agency dealt with many cases of interference caused by thermostats or the radio-suppression components fitted to them. In about 90% of these cases, the interference is attributable to thermostats in gas boilers. It seems that, as these operate in a heat-stressed environment, they are prone to more rapid deterioration than other domestic thermostats such as room thermostats, cylinder thermostats and diverter switching valves. Sometimes the offending thermostat is found in the house that is suffering the interference, although there have been cases where the source of the interference has been found some distance away.

New domestic appliances are required to pass tests for "discontinuous disturbance" emissions (the current harmonized standard is EN 55014-1), but this does not guarantee that such products will remain noise-free after many years of operation. The limits for RF emissions are related in a complex way to the repetition rate and duration of the automatic switching event.

An example is the interference signal generated from a boiler gas control valve and its associated thermostat switching from stand-by to ON and vice versa. The low-power single-phase arc causes a short burst of radiation. When the thermostat is malfunctioning this burst of radiation can be heard as a rough rasping noise which typically lasts for a few seconds but may last for 20 s or more. It repeats typically every 10 min but in some cases, a faulty thermostat may arc several times per minute. This kind of interference, which is intermittent in nature, is mostly noticed in relation to the reception of analog TV signals at 470–850 MHz and sometimes on FM radio at 88–108 MHz.

Replacing the faulty thermostat will normally resolve the problem, but a better solution is to fit suppression to all

such switching contacts. This prevents the arc forming at the instant of switching and if properly designed has the side effect of lengthening the contact life, but the added cost is usually viewed unfavourably by manufacturers.

9.3.1.4 The quacking duck

In a lighter vein, probably the least critical EMC problem this author has encountered is the case of the quacking duck: there is a toy for the under-fives, which is a fluffy duck with a speech synthesizer which is programmed to quack various nursery rhyme tunes. It does this when a certain spot (hiding a sensor) on the duck is pressed, and it shouldn't do it otherwise. Whilst it was in its Christmas wrapping in our house, which is not electrically noisy, it was silent. But when it was taken to our daughter's house and left in the kitchen on top of the fridge, next to the microwave oven, the Christmas present quacked apparently at random and with no one going near it. Some disconcerting moments arose before it was eventually explained to the family that this was just another case of bad EMC and that they shouldn't start to doubt their sanity!

9.3.2 Compatibility between and within systems

The threat of EMI is controlled by adopting the practices of EMC, as defined earlier. The concept of EMC has two complementary aspects:

- it describes the ability of electrical and electronic systems to operate without interfering with other systems;
- it also describes the ability of such systems to operate as intended within a specified electromagnetic environment.

Thus it is closely related to the environment within which the system operates. Effective EMC requires that the system is designed, manufactured and tested with regard to its predicted operational electromagnetic environment: that is, the totality of electromagnetic phenomena existing at its location. Although the term "electromagnetic" tends to suggest an emphasis on high-frequency field-related phenomena, in practice the definition of EMC encompasses all frequencies and coupling paths, from DC through mains supply frequencies to RF and microwaves. And "phenomena" is not restricted to radio-based phenomena but also transient events and power-related disturbances.

9.3.2.1 Intra-system EMC

There are two approaches to EMC. In one case the nature of the installation determines the approach. EMC is especially problematic when several electronic or electrical systems are packed into a very compact installation, such as on board aircraft, ships, satellites or other vehicles. In these cases susceptible systems may be located very close to powerful emitters and special precautions are needed to maintain compatibility. To do this cost-effectively calls for a detailed knowledge of both the installation circumstances and the characteristics of the emitters and their potential victims. Military, aerospace and vehicle EMC specifications have evolved to meet this need and are well established in their particular industry sectors.

This, then, can be characterized as an int*ra*-system approach: the EMC interactions occur between parts of the overall system, the whole of which is amenable to characterization. It may not be necessary or desirable to draw a boundary around individual products in the system, but rather to consider how they affect or are affected by other parts of the same system. Mitigation measures can be applied as easily, and sometimes more easily, at the system level as at the equipment level.

9.3.2.2 Inter-system EMC

The second approach assumes that the system will operate in an environment which is electromagnetically benign within certain limits, and that its proximity to other sensitive equipment will also be controlled within limits. This approach is appropriate for most electronics used in homes, offices and industry, and similar environments. So for example, most of the time a personal computer will not be operated in the vicinity of a high-power radar transmitter, nor will it be put right next to a mobile radio receiving antenna. This allows a very broad set of limits to be placed on both the permissible emissions from a device and on the levels of disturbance within which the device should reasonably be expected to continue operating. These limits are directly related to the class of environment – domestic, commercial, industrial, etc. – for which the device is marketed. The limits and the methods of demonstrating that they have been met form the basis for a set of standards, some aimed at emissions and some at immunity, for the EMC performance of any given product *in isolation*. This makes it an int*er*-system approach rather than an int*ra*-system approach, and means that a necessary part of the process is defining the boundary of the product – easy for typical commercial electronic devices, harder when it comes to installations.

Note that compliance with such standards will not guarantee EMC under all conditions. Rather, it establishes a probability (hopefully very high) that equipment will not cause interference nor be susceptible to it when operated under *typical* conditions. There will inevitably

be some special circumstances under which proper EMC will not be attained – such as operating a computer within the near field of a powerful transmitter – and extra protection measures must be accepted.

9.3.2.3 When intra-system meets inter-system

Difficulty arises when these two approaches are confused one with the other, or at the interface where they meet. This can happen when commercial equipment is used in other environments, for instance on vehicles or in aircraft, and we get the issues referred to earlier, for instance with passenger electronic devices; the PED might be compliant with its normal requirements but that isn't necessarily relevant to its use in these different surroundings. Military projects might require commercial-off-the-shelf (COTS) products to be procured, but their EMC performance requirements are substantially mismatched to military needs. Grounding and bonding techniques which are necessary and appropriate for intra-system requirements can be misapplied to attempt to meet the EMC Directive.

From the product designer's point of view, many of the necessary techniques are similar or common to both approaches, but there are instances where they diverge and so we need to be clear about which approach is being considered in any given case.

9.3.3 The scope of EMC

9.3.3.1 Malfunction of control systems

Solid state and especially processor-based control systems have taken over many functions which were earlier the preserve of electromechanical or analog equipment such as relay logic or proportional controllers. Rather than being hard-wired to perform a particular task, programmable electronic systems rely on a digital bus-linked architecture in which many signals are multiplexed onto a single hardware bus under software control. Not only is such a structure more susceptible to interference, because of the low level of energy needed to induce a change of state, but also the effects of the interference are impossible to predict; a random pulse may or may not corrupt the operation depending on its timing with respect to the internal clock, the data that is being transferred and the program's execution state. Continuous interference may have no effect as long as it remains below the logic threshold, but when it increases further the processor operation will be completely disrupted. With increasing functional complexity comes the likelihood of system failure in complex and unexpected failure modes.

Clearly the consequences of interference to control systems will depend on the value of the process that is being controlled. In some cases disruption of control may be no more than a nuisance, in others it may be economically damaging or even life threatening. The level of effort that is put into assuring compatibility will depend on the expected consequences of failure.

9.3.3.1.1 Phenomena

Electromagnetic phenomena which can be expected to interfere with control systems are:

- supply voltage interruptions, dips, surges and fluctuations;
- fast transient overvoltages (spikes and surges) on supply, signal and control lines;
- RF fields, both pulsed (radar) and continuous, coupled directly into the equipment or onto its connected cables;
- electrostatic discharge (ESD) from a charged object or person;
- low-frequency magnetic or electric fields.

The stress levels that can occur for each of these phenomena depend largely on the local environment, with industrial areas expected to be more stressful than, say, residential or commercial. Special environments such as military, automotive and aerospace can be expected to have even higher levels of some phenomena. The levels are reflected in the standards for immunity testing.

Note that we are not directly concerned with the phenomenon of component damage due to ESD, which is mainly a problem of electronic production. Once the components are assembled into a unit they are protected from such damage unless the design is particularly lax. But an ESD transient can corrupt the operation of a microprocessor or clocked circuit just as a transient coupled into the supply or signal ports can, without actually damaging any components (although this may also occur), and this is properly an EMC phenomenon.

9.3.3.1.2 Software

Malfunctions due to faulty software may often be confused with those due to EMI. Especially with real-time systems, transient coincidences of external conditions with critical software execution states can cause operational failure which is difficult or impossible to replicate, and the fault may survive development testing to remain latent for years in fielded equipment. The symptoms – system crashes, incorrect operation or faulty data – can be identical to those induced by EMI. In fact you may only be able to distinguish faulty software from poor EMC by characterizing the environment in which the system is installed.

9.3.3.2 Immunity of data and programme processing

A very large class of EMC problems comes not with the impact of disturbance on electronic controls, but on various types of system which process audio, video material or data – telecoms and consumer applications being the main areas. Interference here manifests as corruption of data or degradation of the audio or video programme that is being enjoyed.

The phenomena which cause the effects are the same as those discussed above in Section 9.3.3.1.1. In the worst case these can cause complete loss of function; as before, in digitally based systems this is usually a result of transient interference interrupting the processor program control. But before that level is reached, lower levels of transients and continuous RF disturbances can affect the analog sections of the equipment and cause annoying effects on the programme material itself, particularly demodulation effects which appear as distortion of the video picture or extraneous audible noise on an audio programme. The buzzing noise that occurs when you place a mobile cellphone next to your radio or TV set, which warns you that it is about to ring, or when you are engaged on a call, is an example of this phenomenon. Other examples which were more prevalent in the days before RF immunity requirements became well established are the pickup of taxi mobile transmitters on your home hi-fi, or the susceptibility of early electronic telephones to the programmes of high-power local broadcasting transmitters.

Because of the heavy reliance of developed societies on data communications, the impact of interference on telecommunications networks is a critical issue and the network equipment itself must be capable of withstanding the full range of phenomena at quite substantial levels; for instance the lightning surge immunity of exchange equipment and mobile phone base stations, both at the equipment and the system level, is carefully specified by the network operators.

9.3.3.3 Interference with radio reception

Bona fide users of the radio spectrum have a right to expect their use not to be affected by the operation of equipment which is nothing to do with them. Typically, received signal strengths of wanted signals vary from less than a microvolt to more than a millivolt, at the receiver input. If an interfering signal is present on the same channel as the wanted signal then the wanted signal will be obliterated if the interference is of a similar or greater amplitude. The acceptable level of co-channel interference (the "protection factor") is determined by the wanted programme content and by the nature of the interference. Continuous interference on a high-fidelity broadcast signal would be unacceptable at very low levels, whereas a communications channel carrying compressed voice signals can tolerate relatively high levels of impulsive or transient interference.

Analog sound and TV broadcasting are being replaced by digital broadcasting like Digital Radio Mondial (DRM) which is intended to replace the AM radio in the MF and HF bands, Digital Audio Broadcasting (DAB or T-DAB) operated in the VHF and UHF bands, and Digital Video Broadcasting Terrestrial (DVB-T) operated in the UHF bands. These digital radio services require lower RF protection ratios (17 dB for DRM, 20 dB for DVB-T and 28 dB for DAB) than radio services with analog modulation (which need protection ratios of about 27 dB for AM, about 48 dB for FM and about 58 dB for TV). On the other hand, the transition between no interference and unacceptable interference is narrower than for analog modulation: digital systems do not fail gracefully.

9.3.3.3.1 Setting the limits

Radiated interference, whether intentional or not, decreases in strength with distance from the source. For radiated fields in free space, the decrease is inversely proportional to the distance provided that the measurement is made in the far field (see Section 10.1.4.2 for a discussion of near and far fields). As ground irregularity and clutter increase, the fields will be further reduced because of shadowing, absorption, scattering, divergence and defocussing of the diffracted waves. Annex D of CISPR 11 suggests that for distances greater than 30 m over the frequency range 30–300 MHz, the median field strength varies as $1/d^n$ where n varies from 1.3 for open country to 2.8 for heavily built-up urban areas ($n = 1$ for ideal, far field, free-space propagation). An average value of $n = 2.2$ can be taken for approximate estimations; thus, increasing the separation by 10 times would give a drop in interfering signal strength of 44 dB. In the near field – for a frequency of 30 MHz the transition distance is 1.6 m, extending further away proportionally as the frequency decreases – the factor n depends also on the detailed characteristics of the emitting source and cannot be generalized, but remains theoretically in the range 2–3.

Below 30 MHz the dominant method of coupling out of the interfering equipment is via its connected cables, and therefore the radiated field limits are translated into equivalent voltage or current levels that, when present on the cables, correspond to a similar level of threat to HF and MF reception. Electric or magnetic field coupling to nearby antennas in this frequency range is in most cases of minor importance

compared with conduction coupling to the mains input of the affected receiver.

Limits for unintentional emissions are based on the acceptable interfering field strength that is present at the receiver – that is, the minimum wanted signal strength for a particular service modified by the protection ratio – when a nominal distance separates it from the emitter. This will not protect the reception of very weak wanted signals nor will it protect against the close proximity of an interfering source, but it will cover the majority of interference cases and this approach is taken in all those standards for emission limits that have been published for commercial equipment by CISPR. CISPR 16–4–4 gives an account of how such limits are derived, including the statistical basis for the probability of interference occurring.

Historically, the rationale for particular limits has relied upon statistics of interference complaints. A draft amendment to CISPR 16–4–4 points out that with the advent of digital communications this is becoming harder to rely on:

> *If a digital mobile phone or a wireless LAN receiver cannot receive the signal from the nearest base station or access point because of an unwanted emission from a nearby equipment, the user will never suspect this equipment and will even not consider the possibility of an interference to occur. He will assume that the coverage of the network is poor and will move to another place to make his call or to get his connection. Furthermore, as these systems are generally frequency agile, if one channel is interfered with, the system will choose another channel, but if all other channels are occupied, then the phone will indicate that the network is busy, and once again, the user will think the network capacity is not large enough to accommodate his call, but he will never suspect an EMC problem.*
>
> *Generally for analogue systems, one can hear the interference. With digital and mobile systems, interference is much less noticeable*[2] *(muting in audio reception, or frozen images on the TV set for DVB). In addition, modern digital modulations implement complex escape mechanisms (data error correction, frequency agile systems, …) so that the system can already be permanently affected from an EMC point of view before an interference case actually is detected.*
>
> *The evolutions detailed above – generalisation of mobile use of radio receivers and move from analogue to digital radio services – will not reduce the number of interference situations, but continues to decrease the probability of getting significant numbers of interference complaints indicating an existing EMC problem.*

9.3.3.3.2 Malfunction versus spectrum protection

It should be clear from the foregoing discussion that RF emission limits are not determined by the need to guard against malfunction of equipment which is not itself a radio receiver. As discussed in the last section, malfunction requires fairly high energy levels – RF field strengths in the region of 1–10 V/m for example. Protection of the spectrum for radio use is needed at much lower levels, of the order of 10–100 μV/m – ten to a hundred thousand times lower. RF incompatibility between two pieces of equipment neither of which intentionally uses the radio spectrum is very rare. Normally, equipment immunity is required from the local fields of intentional radio transmitters, and unintentional emissions must be limited to protect the operation of nearby intentional radio receivers. The two principal EMC aspects of emissions and immunity therefore address two different issues.

9.3.3.3.3 Free radiation frequencies

Certain types of equipment, collectively known as industrial, scientific and medical (ISM) equipment, generate high levels of RF energy but use it for purposes other than communication. Medical diathermy and RF heating apparatus are examples. To place blanket emissions limit on this equipment would be unrealistic. In fact, the International Telecommunications Union (ITU) has designated a number of frequencies specifically for this purpose, and equipment using only these frequencies (colloquially known as the "free radiation" frequencies) is not subject to emission restrictions. Table 9.3.1 lists these frequencies.

9.3.3.3.4 Co-channel interference, spurious emissions and blocking

A further problem with radio communications, often regarded as an EMC issue, is the problem of co-channel interference from unwanted transmissions. This is caused when two radio systems are authorized to use the same frequency on the basis that there is sufficient distance between the systems, but abnormal propagation

[2] It is questionable that the interference is less "noticeable"; on the other hand, it is harder to attribute the noticeable effects unequivocally to interference.

Table 9.3.1 ITU designated industrial, scientific and medical free radiation frequencies

Centre frequency, MHz	Frequency range, MHz
6.780	6.765–6.795[a]
13.560	13.553–13.567
27.120	26.957–27.283
40.680	40.66–40.70
433.920	433.05–434.79[a]
2450	2400–2500
5800	5725–5875
24,125	24,000–24,250
61,250	61,000–61,500[a]
122,500	122,000–123,000[a]
245,000	244,000–246,000[a]

[a] Maximum radiation limit under consideration.
The frequency range 902–928 MHz is also allowed in Region 2 only (the Americas, Canada and Greenland).
Source: CISPR 11.

conditions increase the signal strengths to the point at which interference is noticeable. This is essentially an issue of spectrum utilization.

A transmitted signal may also overload the input stages of a nearby receiver which is tuned to a different frequency and cause desensitization or distortion of the wanted signal. Transmitter outputs themselves will have spurious frequency components present as well as the authorized frequency, and transmitter type approval has to set limits on these spurious levels. This is more properly viewed as a spectrum management concern and is sometimes called "mutual interference".

Issues which have to be addressed are:

- spurious and harmonic emissions from the transmitters: emissions of any frequency which is not intended;
- correct frequency and power control of transmitters;
- spurious responses of the receivers: reception of any frequency which is not intended;
- adjacent channel blocking of receivers: unintended deafness induced by loud signals on nearby channels;
- sidelobe antenna patterns of both transmitters and receivers;
- intermodulation effects in transmitters, receivers and nearby structures, which create different frequencies from the interaction of existing ones.

9.3.3.4 Disturbances of the mains supply

Mains electricity suffers a variety of disturbing effects during its distribution. These may be caused by sources in the supply network or by other users, or by other loads within the same installation. A pure, uninterrupted supply would not be cost-effective; the balance between the cost of the supply and its quality is determined by national regulatory requirements, tempered by the experience of the supply utilities. Typical disturbances are:

- *Medium-term voltage variations*: The distribution network has a finite source impedance and varying loads will affect the terminal voltage. Not including voltage drops within the customer's premises, an allowance of ±10% on the nominal voltage will cover normal variations in Europe.
- *Voltage fluctuations*: Short-term (sub-second) fluctuations with quite small amplitudes are annoyingly perceptible as flicker on electric lighting, though they are comfortably ignored by electronic power supply circuits. Generation of flicker by high-power load switching is subject to regulatory control.
- *Voltage dips and interruptions*: Faults on power distribution systems cause up to 100% voltage drops but are cleared quickly and automatically by protection devices, and throughout the rest of the system the voltage immediately recovers. Most consumers therefore see a short voltage dip. The frequency of occurrence of such dips depends on location and seasonal factors.
- *Waveform distortion*: At source, the AC mains is generated as a pure sine wave but the reactive impedance of the distribution network together with the harmonic currents drawn by non-linear loads causes voltage distortion. Power converters and electronic power supplies are important contributors to non-linear loading. Harmonic distortion may actually be worse at points remote from the non-linear load because of resonances in the network components. Not only must non-linear harmonic currents be limited but equipment should be capable of operating with up to 10% total harmonic distortion in the supply waveform.
- *Transients and surges*: Switching operations generate transients of a few hundred volts as a result of current interruption in an inductive circuit. These transients normally occur in bursts and have rise-times of no more than a few nanoseconds, although the finite bandwidth of the distribution network will quickly attenuate all but local sources. Rarer high-amplitude spikes in excess of 2 kV may be observed due to fault conditions. Even higher voltage surges

due to lightning strikes occur, most frequently on exposed overhead line distribution systems in rural areas.

All these sources of disturbance can cause malfunction in systems and equipment that do not have adequate immunity.

9.3.3.4.1 Mains signalling

A further source of incompatibility arises from the use of the mains distribution network as a telecommunications medium, or mains signalling (MS). MS superimposes signals on the mains in the frequency band from 3 to 150 kHz and is used both by the supply industry itself and by consumers. Unfortunately this is also the frequency band in which electronic power converters – not just switch-mode power supplies, but variable speed motor drives, induction heaters, fluorescent lamp inverters and similar products – operate to their best efficiency. There are at present very few pan-European standards which regulate conducted emissions on the mains below 150 kHz, although EN 50065-1 sets the frequency allocations and output and interference limits for MS equipment itself. Overall, compatibility problems between MS systems and such power conversion equipment can be expected to increase.

9.3.3.5 Powerline telecoms

Power Line Telecommunication (PLT or PLC, Power Line Communications, or Broadband over Power Line, BPL, in the USA) is a means of transmitting broadband data over the installed base of mains electricity supply cables. It can be used in two ways:

- access to the home, to deliver the data connection from the service provider;
- networking within the home, for data interconnection between mains-connected devices.

The two applications use different frequency ranges: low frequency (1.6–10 MHz) for access, and high frequency (10–30 MHz) for in-home, as specified for Europe in ETSI TS 101 867. The generally accepted power level for adequate operation of a PLT system is −50 to −40 dBm/Hz. Measured in a 9 kHz bandwidth, as is the normal method for interference measurements at these frequencies, this implies a power level of around −10 to 0 dBm, which across the differential 100 Ω resistance of the power network is 100–110 dBμV (0.1–0.32 V). This compares with the allowed levels for conducted emissions in the domestic environment, with which most if not all electronic product designers are familiar, of 60 dBμV in a comparable frequency range – one hundred times lower.

9.3.3.5.1 The politics of PLT

Because it can deliver domestic broadband access as an alternative to other providers such as cable and telephone companies, PLT is viewed favourably by regulators on the grounds of extending competition. The local loop, or the "last mile" (delivery of the broadband data finally into the home or office), appears as a bottleneck in the process of liberalizing the competitive environment for the telecommunications infrastructure, particularly in breaking the perceived stranglehold of the "incumbents" (pre-existing telecom providers). Hence any technology which promises to unblock this bottleneck is encouraged by the European authorities. PLT is clearly such a technology.

But meanwhile, some European nations have implemented regulations which would allow them to control it if there was any threat of interference becoming widespread. In Germany, the standard NB30 put down radiated emission limits in the 1.6–30 MHz range. In the UK, the Radiocommunications Agency standard MPT1570 was also published, though it covered a lower frequency range. Naturally, this put a brake on PLT activity in these countries, since investors were wary of supporting systems which might quickly turn out to be illegal, and it also meant that there were differences in approach across the European Union.

In 2001 the European Commission placed a mandate on the standard bodies ETSI and CENELEC (mandate M/313) to create a standard for the EMC of Telecommunications Networks. This was addressed by a Joint Working Group of the two bodies but the difficulties involved, particularly that of finding agreement on a set of limits for radiated emissions from the network which would satisfy all participants, meant that the work on it eventually stalled. The Commission subsequently issued a Recommendation in April 2005 which includes the following wording:

> *2. … Member States should remove any unjustified regulatory obstacles, in particular from utility companies, on the deployment of broadband powerline communications systems and the provision of electronic communications services over such systems.*
>
> *3. Until standards to be used for gaining presumption of conformity for powerline communications systems have been harmonised under Directive 89/336/EEC, Member States should consider as compliant with that Directive a powerline communications system which is:*
>
> - *made up of equipment compliant with the Directive and used for its intended purpose,*

- *installed and operated according to good engineering practices designed to meet the essential requirements of the Directive.*

The documentation on good engineering practices should be held at the disposal of the relevant national authorities for inspection purposes as long as the system is in operation.

We will meet the phrase "good engineering practices" elsewhere (Section 9.4.2.5.4) in the context of fixed installations.

9.3.3.5.2 The use of the HF spectrum

The slice of spectrum from about 1 to 30 MHz (MF and HF) is unique in that it can support long-distance communication, and so it is important to broadcasters and many other users. Sky-wave propagation in the HF bands enables an international broadcaster to reach a target country without having a transmitter within the target area. This has political consequences, since it means that an audience can be reached without the cooperation of that country's authorities – which cannot be said for other kinds of access, including any kind of internet delivery. The BBC's World Service, for instance, is broadcast on several HF frequencies and is heard by many people in countries that have no free media of their own.

As well as broadcasting, aeronautical and marine communications use the HF band for long-distance communication when the mobile station is out of reach of ground-based VHF stations, which is a large proportion of their journeys.

9.3.3.5.3 PLT's interference capability

Interference from PLT systems stands outside the general regime of interference control, which as we will see applies limits to the amount of noise injected onto the mains supply. The principal emissions are deliberately injected onto the supply wiring, rather than accidentally as is the case with other sources such as fluorescent light inverters or computer power supplies. From access-PLT systems, the interference will be largely continuous rather than intermittent, and will potentially affect all households being supplied from a substation in a PLT-active zone, whether they are a subscriber or not. Even in-home systems could interfere with other parties connected to the same service entrance.

The interference signal will stretch broadband across the whole of the spectrum occupied by the modem's output, and within a given region of spectrum it will be impossible to tune it out. In the quiescent state some systems will create a pulsing type of signal which may or may not be subjectively less annoying than the continuous noise which occurs when the system is actually passing data. Some systems may use low-frequency carriers such that a continuous audible tone is present across the spectrum.

9.3.3.5.3.1 Dependence on quality of wiring

The mains supply wiring both to and within a domestic house was never intended to carry high frequencies. At some frequencies the signal may be transmitted with little loss, but at others the attenuation can be severe, and this characteristic can change with time as users plug various appliances into the mains supply. This means that in order to work at all, the amplitude of the signal must be high enough to ride over any interference already present on the network, and must adapt to time-dependent changes in this interference and the network attenuation. Current-generation PLT systems are designed to do this.

A critical parameter which determines the amount of radiation that the mains wiring creates is the "Longitudinal Conversion Loss" (LCL) of the cable. Simply put, this is the ratio between the signal level which appears across the wires, intentionally, due to the desired data transmission, and which to a first order does not radiate; and the signal level in common mode – all wires together – which represents the leakiness of the cable and which contributes the lion's share of the radiation. Data cables which carry broadband signals, of which Ethernet is the most typical example, are very tightly specified for good LCL, which ensures that the RF leakage from the data signal is kept low. This is not the case for mains wiring. The most important aspect of cable design which affects LCL is the physical balance of the two wires that make up the cable. Each conductor must be tightly coupled to the other so that they cancel each other's interaction with the environment. Data cables are tightly twisted in a controlled way to achieve this. Not only is mains wiring not controlled in this way, but also it is commonly installed in direct contravention of this principle.

9.3.3.5.4 Is PLT the same as other interferers?

PLT supporters propose that there should be parity (at least) between the emissions compliance requirements that a PLT system has to meet, and those applied to other devices such as information technology, lighting or household appliances. CISPR conducted limits, it is said, have been adequate to protect the HF spectrum so far and therefore any system limits should be no more onerous than levels derived from these.

This argument overlooks a number of important points:

- A victim won't be able to get away from PLT interference. When a whole street or a whole building is wired for PLT, it will be pervasive and repositioning the victim will not work. CISPR limits assume that

mitigation by separation from a local interferer is possible.
- PLT is always on. CISPR limits incorporate a relaxation which takes into account the probability of non-coincidence in time of source and victim. For PLT, this relaxation must be ignored.
- EMC engineers know that most products which comply with CISPR conducted limits do so with a good margin, often at least 20 dB, in the frequency range above 2 MHz. If CISPR limits do indeed protect HF reception, this factor should not be overlooked, since such a margin will not be enjoyed by a PLT product.

In fact, PLT modems are unable to operate anywhere near the mains conducted emissions limits in force in CISPR at the moment.

9.3.3.5.5 Cumulative effects

The threat to victim receivers in close proximity to the PLT system is not the only one that concerns radio administrations. If PLT were to be widely implemented within any country, the total radiated power available would be sufficient to increase the radio noise floor at distances remote from the source, potentially even in other countries. If, say, an entire city was to be wired for PLT, this could form an aggregate transmitter whose RF energy was reflected from the ionosphere and illuminated a continent. In addition, an aircraft flying over such a city might find that its ability to receive HF signals was curtailed. The UK's Civil Aviation Authority has expressed its concern that "aeronautical services are under threat from cabled telecommunications services." Established HF propagation models exist for this phenomenon and a number of studies have been carried out to try and model the possible outcome. The problem with any such study is that for the time being it must remain theoretical, since it is impossible to validate the radiation (as opposed to propagation) models used for prediction until there are sufficient installed systems to be statistically acceptable; but by then the roll out may be so advanced that it will be impossible to stop it.

9.3.3.5.6 Compliance status of PLT devices

The EC's Recommendation on PLT quoted above refers to a system being "made up of equipment compliant with the Directive". Here is the nub of the question: how can PLT modems be made compliant with the EMC Directive? It is the case that some PLT modems are already on the market in Europe and are CE Marked, which means that their manufacturers believe that they meet the essential requirements of the EMC Directive. But there are no standards specifically for such devices and for now, no such device could actually meet the usual standards. In fact, existing products appear to use the Technical Construction File route (under the first EMC Directive, see Section 9.4.1.3) which implies that a Competent Body has issued a certificate of compliance with the essential requirements. It is difficult to know how this has been justified.

9.3.3.5.7 Opening the floodgates

It must be assumed that the mains supply already carries noise from other apparatus 18 EMC for Product Designers which may approach the usual conducted emission limits, even if everything connected is in full compliance with the EMC Directive. For PLT to operate at all, its signals must be greater than this minimum noise level, almost by definition; and so it must breach these limits, again by definition. This is indeed so, by several tens of dB, and if it were not, PLT could not operate. Yet all other mains-connected equipment, such as ITE, medical products, household appliances, lighting and so forth, are subject to the standard limits.

What is to prevent the manufacturers of such equipment, which after all forms the vast bulk of products placed on the market within the EU, from demanding to know why PLT has received such special treatment? If PLT can flagrantly flout the limits, they would say, so can we. But of course, were they to do that, it would open the floodgates to an uncontrolled escalation of interference on the mains wires. More bluntly, it would drive a horse and cart through the principles of interference control established over decades.

This exposes a contradiction at the core of the case for PLT. It can only operate if it is indeed granted special status to apply RF disturbances to the mains lines. It must, in fact, be regarded as a special case in the context of the EMC Directive. It cannot possibly comply with the requirement not to generate an electromagnetic disturbance exceeding "a level allowing radio and telecommunications equipment and other apparatus to operate as intended" because, since the limits are set to achieve this requirement, it must itself exceed those limits and therefore breach the requirement.

In fact, it seems at the time of writing that PLT is losing its attractiveness both to the regulators and to its hopeful operators, mainly because of the regulatory EMC problems rehearsed above. Probably, it will be relegated to a curiosity in years to come, superseded by less polluting, more reliable and more effective wired and wireless ways of delivering broadband internet access.

9.3.3.6 Other EMC issues

The issues discussed above are those which directly affect product design to meet commercial EMC

requirements, but there are some other aspects which should be mentioned briefly.

9.3.3.6.1 Electro-explosive devices and flammable atmospheres

The first is the hazard of ignition of flammable atmospheres in petrochemical plant, or the detonation of electro-explosive devices (EEDs) in places such as quarries, due to incident RF energy. A strong electromagnetic field will induce currents in large metal structures which behave as receiving antennas. A spark will occur if two such structures are in intermittent contact or are separated. If flammable vapour is present at the location of the spark, and if the spark has sufficient energy, the vapour will be ignited. Different vapours have different minimum ignition energies, hydrogen/air being the most sensitive. The energy present in the spark depends on the field strength, and hence on the distance from the transmitter, and on the antenna efficiency of the metal structure. BS 6656 discusses the nature of the hazard and presents guidelines for its mitigation.

Similarly, EEDs are typically connected to their source of power for detonation by a long wire, which can behave as an antenna. Currents induced in it by a nearby transmitter could cause the charges to explode prematurely if the field was strong enough. As with ignition of flammable atmospheres, the risk of premature detonation depends on the separation distance from the transmitter and the efficiency of the receiving wire. EEDs can if necessary be filtered to reduce their susceptibility to RF energy. BS 6657 discusses the hazard to EEDs.

9.3.3.6.2 Data security

The second aspect of EMC is the security of confidential data. Low-level RF emissions from data-processing equipment may be modulated with the information that the equipment is carrying – for instance, the video signal that is fed to the screen of a VDU.

These signals could be detected by third parties with sensitive equipment located outside a secure area and demodulated for their own purposes, thus compromising the security of the overall system. This threat is already well recognized by government agencies and specifications for emission control, under the Tempest scheme, have been established for many years. Commercial institutions, particularly in the finance sector, are now beginning to become aware of the problem.

9.3.3.6.3 Electromagnetic weapons

The idea that an intense broadband radiated pulse could be generated intentionally, and used to upset the operation of all potentially susceptible electronics within a certain range, is gaining credence. Because of the almost universal social reliance on electronic systems, an attack that simultaneously crashed many computer networks could indeed have substantial consequences. It is known that US and other military researchers are working on such technology, but we can also imagine less sophisticated devices being within reach of many other organizations or individuals.

The more sensationalist press, of course, has a field day with this idea – phrases such as "frying computer chips" are used with abandon. Realistically, the amount of energy needed to generate a wide-area pulse would be so enormous that only disruption, not damage, is at all likely. This is precisely the effect of a high-altitude nuclear explosion, which generates a sub-nanosecond nuclear electromagnetic pulse that is disruptive over an area of hundreds of square kilometres. The idea that attracts military researchers now is to do this more discreetly. The limitation of any such weapon is its uncertainty. Unless you know exactly what kind of electronics you are attacking, and how well protected it is, it is hard to predict the damage that the weapon will cause. Equipment that is immune to a local ESD (as described in these pages) is likely to have good immunity to electromagnetic warfare.

On the other hand, radio communications are easy to disrupt by jamming transmissions and this subject is well established in the military canon; there are military aircraft that are specifically designed to carry high-power broadband jammers and it is not hard to imagine that they are used with some enthusiasm in "theatres of war" (it is said that one such aircraft suffered an engine shutdown when its jammers were turned on in earnest). Under such circumstances, the issue of the commercial application of radiated emission limits is of no more than academic interest.

9.3.3.7 The compatibility gap

The increasing susceptibility of electronic equipment to electromagnetic influences is being paralleled by an increasing pollution of the electromagnetic environment. Susceptibility is a function partly of the adoption of VLSI technology in the form of microprocessors, both to achieve new tasks and for tasks that were previously tackled by electromechanical or analog means, and the accompanying reduction in the energy required of potentially disturbing factors. It is also a function of the increased penetration of radio communications, and the greater opportunities for interference to radio reception that result from the co-location of unintentional emitters and radio receivers.

At the same time more radio communications mean more transmitters and an increase in the average RF field strengths to which equipment is exposed. A study has

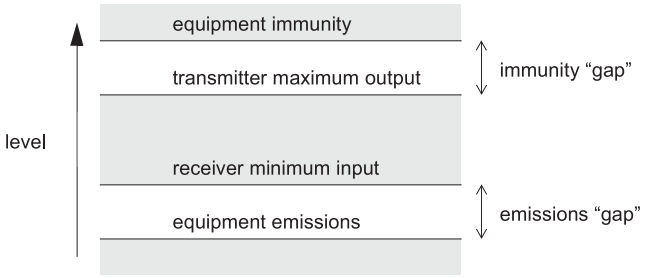


Fig. 9.3.1 The EMC gap.

been reported which quantified this exposure for a single site at Baden, Switzerland, for 1 year; this found the background field strength in the shortwave band regularly approaching, and occasionally exceeding, levels of 1 V/m. Also, the proliferation of digital electronics means an increase in low-level emissions which affect radio reception, aptly described as a form of electromagnetic "smog".

These concepts can be graphically presented in the form of a narrowing EMC gap, as in Figure 9.3.1. This "gap" is of course conceptual rather than absolute, and the phenomena defined as emissions and those defined as immunity do not mutually interact except in rare cases. But the maintenance of some artificially defined gap between equipment immunity and radio transmissions on the one hand, and equipment emissions and radio reception on the other, is the purpose of the application of EMC standards, and is one result of the enforcement of the EMC Directive.

9.3.4 Electromagnetic fields and human health

This chapter is about EMC in electronic products, but one subject which any EMC engineer is regularly quizzed about is the related issue of the direct health hazards of electromagnetic fields. Apart from its application under the R&TTE Directive and relevance to radiated RF immunity tests at high levels, it is not a subject that will take up a lot of space.

It is undoubtedly the case that high levels of electromagnetic energy can create effects in human tissue. This then implies that below some level the EM energy can be regarded as "safe" and above it as "hazardous", and that exposure to it should be controlled. The exposure levels to which most if not all regulatory limits adhere are laid down in the ICNIRP Guidelines.

9.3.4.1 ICNIRP basic restrictions

The International Commission on Non-Ionising Radiation Protection is a nongovernmental expert body and its guidance document refers to the known short-term and immediate health effects. These are:

- stimulation of peripheral nerves and muscles;
- shocks and burns through touching conducting parts;
- tissue heating due to energy absorption.

For absorption of energy, EM fields can be divided into four frequency ranges:

- frequencies from about 100 kHz to less than about 20 MHz, at which absorption in the trunk falls rapidly with decreasing frequency, and significant absorption may occur in the neck and legs;
- frequencies in the range from about 20 to 300 MHz, at which relatively high absorption can occur in the whole body, and to even higher values if partial-body (e.g. head) resonances are considered;
- frequencies in the range from about 300 MHz to several GHz, at which significant local, non-uniform absorption occurs;
- frequencies above about 10 GHz, at which energy absorption occurs primarily at the body surface.

The ICNIRP guidelines' levels are set on the following basis:

> *Available experimental evidence indicates that the exposure of resting humans for approximately 30 minutes to EMF producing a whole-body SAR of between 1 and 4 W kg^{-1} results in a body temperature increase of less than 1 °C. Animal data indicate a threshold for behavioral responses in the same SAR range. Exposure to more intense fields, producing SAR values in excess of 4 W kg^{-1}, can overwhelm the thermoregulatory capacity of the body and produce harmful levels of tissue heating.*
>
> *Many laboratory studies with rodent and non-human primate models have demonstrated the broad range of tissue damage resulting from either partial-body or whole-body heating producing temperature rises in excess of 1–2 °C. The sensitivity of various types of tissue to thermal damage varies widely, but the threshold for irreversible effects in even the most sensitive tissues is greater than 4 W kg^{-1} under normal environmental conditions. These data form the basis for an occupational exposure restriction of 0.4 W kg^{-1}, which provides a large margin of safety for other limiting conditions such as high ambient temperature, humidity, or level of physical activity.*

SAR stands for Specific Absorption Rate, quoted in watts per kilogram, and is a measure which is given as a Basic Restriction in the ICNIRP guidelines. Depending on frequency, the physical quantities used to specify the basic restrictions on exposure to EMF are current density,

SAR, and power density. Protection against adverse health effects requires that these basic restrictions are not exceeded.

9.3.4.1.1 Reference levels

"Reference levels" of exposure are provided for comparison with measured values of physical quantities (electric or magnetic field strength); compliance with all reference levels given in the ICNIRP guidelines will ensure compliance with basic restrictions.

In many circumstances actually measuring SAR is either not feasible (because it would imply invasive measurements in a human body) or would not be economically justified. In these cases the fields to which people are actually exposed can be measured directly and compared to the reference levels. It is important to note that these are not limits but levels above which you should investigate the exposure regime: provided that basic restrictions are met and adverse indirect effects can be excluded, these field strength values can be exceeded. Typical circumstances in which the reference levels are applied are in exposure of the public or workers to fields from base station transmitter installations (though not in evaluation of handheld transmitters). For EMC test labs, the radiated RF immunity test also provides an opportunity to consider them.

9.3.4.1.2 SAR measurement

Reference levels are not relevant for handheld transmitters since the coupling takes place in the near field of the transmitting antenna. For this situation, it is necessary to measure SAR directly using a phantom of a human head or body, filled with a tissue-simulating liquid (actually a mixture of sugar, salt and water) which gives the correct dielectric properties at the relevant frequency, and allows a specialized E-field probe to be positioned inside the head cavity. This method is now well established for mobile phones and is being extended to other types of radio transmitter which are used close to the body. It not only has some similarities to EMC measurements but also many more differences; relatively few generalist test labs can do it, and it is mostly carried out by the mobile phone manufacturers themselves.

9.3.4.2 Athermal effects

Effects at much lower levels of RF that may be experienced by some people but are not based on tissue heating (athermal effects) are still controversial and disputed, and are not covered by the ICNIRP guidance or indeed any other officially published and recognized documents. ICNIRP itself says "available data are insufficient to provide a basis for setting exposure restrictions" and this position has been maintained for some years, although numerous studies have been published which purport to show one effect or another on biological systems of various frequencies and modulation schemes, including those used by mobile phone transmissions. It seems that there are two problems with this body of research:

- much of the experimental data is open to criticism and has been found difficult or impossible to replicate;
- the implication for radio transmitting authorities, if low-level EM energy is ever conclusively shown to have serious health hazards, is such that many radio-based systems would be shutdown overnight, and this is a nettle that, unsurprisingly, noone in authority has seen fit to grasp.

Nevertheless, there is a substantial body of public opinion which believes that such a low-level hazard exists, and agitates against the continued extension of the radio transmitting infrastructure, particularly as evidenced by mobile phone base station installations. It is largely in (perhaps misplaced) response to this public opinion that standards for measuring the RF emissions of virtually any electronic apparatus have been drafted. Since these are essentially safety standards they are only relevant in the context of the R&TTE Directive.

This page is intentionally left blank

Chapter 9.4

9.4

The EMC Directive

Tim Williams

9.4.1 History

The relaxed EMC regime that existed throughout most of Europe up until the early 90s was overturned with the adoption by the European Commission on 1 January 1992 of the EMC Directive, 89/336/EEC. This was at the time widely regarded to be "the most comprehensive, complex and possibly contentious Directive ever to emanate from Brussels". It has eventually been superseded by its second edition, published at the very end of 2004. This chapter examines the provisions of the new Directive and how manufacturers will need to go about complying with it. But first, we have to understand the background to European Directives and the history of the first EMC Directive.

9.4.1.1 The new approach directives

Of the various aims of the creation of the Single European Market, the free movement of goods between European states is fundamental. All Member States impose standards and obligations on the manufacture of goods in the interests of quality, safety, consumer protection and so forth. If there are detailed differences in procedures and requirements, these act as technical barriers to trade, fragmenting the European market and increasing costs because manufacturers have to modify their products for different national markets.

For many years the Commission tried to remove these barriers by proposing Directives which gave the detailed requirements that products had to satisfy before they could be freely marketed throughout the Community, but this proved difficult because of the detailed nature of each Directive and the need for unanimity before it could be adopted. In 1985 the Council of Ministers adopted a resolution setting out a "New Approach to Technical Harmonization and Standards".

Under the "new approach", directives are limited to setting out the essential requirements which must be satisfied before products may be marketed anywhere within the EU. The technical detail is provided by standards drawn up by the European standards bodies CEN, CENELEC and ETSI. Compliance with these standards is expected to demonstrate compliance with the essential requirements of each Directive. All products covered by each Directive must meet its essential requirements, but all products which do comply, and are labelled as such, may be circulated freely within the Community; no Member State can refuse them entering on technical grounds. Decisions on new approach Directives are taken by qualified majority voting, eliminating the need for unanimity and speeding up the process of adoption.

A document (the "Blue Guide") was published in early 2000 by the European Commission setting out the way in which new approach Directives should be implemented in a relatively harmonized fashion. This says that a new approach Directive contains the following elements:

- the scope of the Directive;
- a statement of the essential *requirements*;
- the *methods of satisfying* the essential requirements;
- how *evidence of conformity* will be provided;
- what *transitional arrangements* may be allowed;
- a statement confirming entitlement to *free circulation*;
- a *safeguard procedure*, to allow Member States to require a product to be withdrawn from the market if it does not satisfy the essential requirements.

It is the responsibility of the European Commission to put forward to the Council of Ministers proposals for

EMC for Product Designers; ISBN: 9780750681704
Copyright © 2007 Elsevier Ltd. All rights of reproduction, in any form, reserved.

new Directives. Directorate-General III of the Commission has the overall responsibility for Directives in the electrical sector. The actual decision on whether or not to adopt a proposed Directive is taken by the Council of Ministers, by a qualified majority vote. Texts of Directives proposed or adopted are published in the *Official Journal of the European Union*. Consultation on draft Directives is typically carried out through European representative bodies and in working parties of governmental experts.

9.4.1.2 Background to the legislation

In the UK, previous legislation on EMC was limited in scope to radio communications. Section 10 of the Wireless Telegraphy Act 1949 enables regulations to be made for the purpose of controlling both radio and non-radio equipment which might interfere with radio communications. These regulations have taken the form of various Statutory Instruments (SIs), which cover interference emissions from spark ignition systems, electromedical apparatus, RF heating, household appliances, fluorescent lights and CB radio. The SIs invoked British Standards which are closely aligned with international and European standards. The power existed to make regulations regarding the immunity to interference of radio equipment but this was never used.

At the European level various Directives have been adopted over the years, again to control emissions from specific types of equipment. Directive 72/245 EEC, adopted in June 1972, regulates interference produced by spark ignition engines in motor vehicles. Directives 76/889 EEC and 76/890 EEC, amended by various other subsequent Directives, apply to interference from household appliances and portable tools, and fluorescent lamps and luminaires. These latter two were superseded and repealed by the first EMC Directive. Each Member State is required to implement the provisions of these Directives in its national legislation, as described above for the UK.

This previous legislation is not comparable in scope to the EMC Directive, which covers far more than just interference to radio equipment, and extends to include immunity as well as emissions.

9.4.1.3 The first EMC Directive

89/336/EEC became fully operational in 1996 after an extended transition period lasting 4 years. During the initial period, manufacturers had to follow a fairly steep learning curve to get to grips both with the legal requirements and the technical subject of EMC: many had never had to deal with the issue before. The European Commission clearly assumed that manufacturers would be able to assume responsibility for the compliance of their products from the start, but in fact for many this was a new subject and a substantial extra burden, and the implementation of the Directive was at best patchy for several years. A detailed guidance document was published by the Commission in 1997 which cleared up many of the misunderstandings that had been created, and by the turn of the century the understanding of the Directive within electrical and electronics companies had largely settled down.

Development of enforcement followed a similar uneven course. The Directive itself did not mandate any particular approach to enforcement and so different Member States took different approaches, depending on their particular culture and attitude towards European-initiated laws. This led to a noticeable variation in the gradient of the playing field across Europe, somewhat along the lines of the "traffic lights" principle[1]. In the UK, the job of enforcement was given to the Trading Standards authorities, more at home with dealing with rogue traders and the safety of consumer products like toys; only a few Trading Standards Officers would consider themselves even slightly familiar with the issues of EMC, and this combined with under-resourcing has meant that enforcement of the Directive in the UK has been somewhat limited in scope.

In fact, a large part of the pressure forcing companies to comply has come not from the enforcement authorities, but from other customers. For instance, in the consumer sector most of the market for electrical goods is dominated by a few large national or international retailers, and these demand correct declarations of conformity as part of their conditions of trade. Many non-consumer products such as industrial instrumentation are sold not to the end user but to a systems integrator, and again these have found that they need to ask for EMC compliance from their suppliers as part of their own compliance stance.

The first EMC Directive had two alternative routes to compliance: application of and self certification to harmonized standards, which was used in the vast majority of cases; and generation and assessment of a Technical Construction File, which was meant to be used if the manufacturer didn't or couldn't use standards, and which involved a review by a Competent Body.

[1] In Northern European countries, compliance with traffic lights is mandatory. In Central Europe, they are advisory. In Southern Europe, traffic lights are for decoration only.

9.4.1.3.1 SLIM

In 1998 the EMC Directive was subject to a review under the SLIM (Simpler Legislation for the Internal Market) process. Ten governmental experts were charged with making recommendations for improvements and simplifications to the Directive. Their report looked at the following areas amongst others:

- functional safety – explicitly, not to be part of the EMC Directive;
- immunity requirements – to be retained, but clarified;
- standards – a strategic review panel to be set up;
- large machines and installations – not subject to conformity assessment;
- definitions of certain types of environment.

The SLIM recommendations have formed the foundation for a complete rewrite of the EMC Directive, but it would be wrong to suggest that everything the SLIM group recommended has been taken on board unaltered in the new edition. Indeed, the second edition has introduced some new aspects which need a whole new raft of interpretation.

9.4.1.3.2 Other Directives

Apart from the EMC Directive, other new approach Directives adopted which may affect some sectors of the electrical and electronic engineering industry include:

- Toy safety (88/378/EEC)
- Non-automatic weighing machines (90/384/EEC)
- Medical devices (93/42/EEC)
- Active implantable electromedical devices (90/385/EEC)
- Machinery safety (89/392/EEC)
- Gas appliances (90/396/EEC)
- Lifts (95/16/EC)
- Refrigeration appliances (96/57/EC)
- In vitro diagnostic medical devices (98/79/EC)
- Radio and telecommunications terminal equipment (99/5/EC).

Some of these contain their own EMC requirements which either partially or completely supplant the EMCD, as mentioned in Section 9.4.2.2.3. In addition to this list, there are two other Directives which are relevant although they are not strictly "new approach" Directives. These are the Low Voltage Directive (LVD) (73/23/EEC) and the Automotive EMC Directive (95/54/EC, now superseded by 2004/104/EC). The LVD is concerned with safety, not EMC, but as a result of the CE Marking Directive the CE mark attests to conformity with this Directive as well as any other applicable new approach Directives.

The Automotive EMC Directive is not a CE Marking Directive. It applies to vehicles and their electrical/electronic sub-assemblies and requires a form of type approval, signified by the 'e' marking.

9.4.2 The second EMC Directive

The second edition of EMC Directive 2004/108/EC was published on 31 December 2004[2]. The timescale for its adoption is as follows:

- Entry into force: 20 January 2005
- Transposition: Member States shall adopt and publish the provisions necessary to comply with this Directive by 20 January 2007. They shall apply those provisions as from 20 July 2007, and from that date equipment claiming compliance with 2004/108/EC can be put on the market.
- Directive 89/336/EEC is repealed as from 20 July 2007.
- Transitional provisions: Member States shall respect compliance with 89/336/EEC for equipment which was placed on the market before 20 July 2009.

9.4.2.1 What changes?

The essential requirements have not changed except for detailed wording, particularly that immunity now requires operation "without unacceptable degradation of its intended use". The routes to compliance however have been changed. Whereas before, if you didn't "apply" harmonized standards you had to use the Technical Construction File route and pass your TCF by a Competent Body, now, the TCF route has been swept away. The overarching requirement is for the manufacturer to perform and document an "EMC assessment" of his product. This may still involve compliance with harmonized standards: "The correct application of all the relevant harmonized standards whose references have been published in the OJEU shall be equivalent to the carrying out of the electromagnetic compatibility assessment." But it is also made clear that "Compliance with a harmonized standard is not compulsory", and the "presumption of conformity" which compliance offers is limited to the scope of the harmonized standard(s) applied and the relevant essential requirements covered by such harmonized standard(s).

[2] At the time of writing, the UK legislation and the EU Guidance for the second edition Directive had not been completed. The discussion in the rest of this chapter draws on draft legislation and guidance, and on consultation workshops that the UK DTI and EMCTLA sponsored in 2005 and 2006.

The real change here is that, if you don't apply correctly all the relevant harmonized standards, you can, indeed must, still carry out an EMC assessment; but there is no requirement to have this assessment reviewed by a third party such as a Competent Body. On the other hand, you may do so voluntarily, and the Directive allows for "Notified Bodies" to be appointed on pretty much identical terms to the old Directive's Competent Bodies, so that these organizations will most probably continue to offer their services under the new Directive. Their use, though, is entirely at the manufacturer's discretion.

The other major change is to the regime for fixed installations. Equipment that is regulated by the Directive now has two subsets: apparatus, and fixed installations. The change allows different regulatory provisions for each. "Apparatus" now covers virtually everything that was covered by the old Directive, including mobile installations and systems; such apparatus is subject to the full provisions of the Directive, but there is an explicit exemption now for benign apparatus. Fixed installations on the other hand are not subject to much of the conformity assessment procedures, but do have to follow other new requirements.

9.4.2.2 Scope, requirements and exceptions

2004/108/EC applies to apparatus which is liable to cause electromagnetic disturbance or which is itself liable to be affected by such disturbance. "Apparatus" is defined as any finished appliance or combination of appliances made commercially available as a single functional unit, and intended for the end user. Essentially, anything which is powered by electricity is covered, regardless of whether the power source is the public supply mains, a battery source or a specialized supply. Specific exceptions are discussed in Section 9.4.2.2.3.

An electromagnetic disturbance is any electromagnetic phenomenon which may degrade performance, without regard to frequency or method of coupling. Thus radiated emissions as well as those conducted along cables, and immunity from EM fields, mains disturbances, conducted transients and RF, electrostatic discharge and lightning surges are all covered. *No* specific phenomena are *excluded* from the Directive's scope.

9.4.2.2.1 Essential requirements

The essential requirements of the Directive (article 5 and annex I) are that the apparatus shall be so designed and manufactured, having regard to the state of the art, that:

- the electromagnetic disturbance it generates does not exceed a level above which radio and telecommunications equipment and other equipment cannot operate as intended;
- it has a level of immunity to the electromagnetic disturbance to be expected in its intended use, which allows it to operate without unacceptable degradation of its intended use.

The intention is to protect the operation not only of other users' radio and telecommunications equipment but any equipment which might be susceptible to EM disturbances, such as information technology or control equipment. At the same time, all equipment must be able to function correctly in whatever electromagnetic environment – that is, all electromagnetic phenomena observable in a given location – it might reasonably be expected to occupy. Notwithstanding these requirements, any Member State has the right to apply special measures with regard to the taking into service of apparatus, to overcome existing or predicted EMC problems at a specific site or to protect the public telecommunications and safety services. Compliance with the essential requirements will be demonstrated via the performance and documentation of an "EMC assessment". This is discussed in Section 9.4.3.

9.4.2.2.2 Sale and use of products

The Directive applies to all apparatus that is placed on the market or taken into service. The definitions of these two conditions do not appear within the text of the Directive but are the subject of several paragraphs in the 1997 Guidelines and in the Blue Guide.

9.4.2.2.2.1 Placed on the market

The "market" means the market in any or all of the European Economic Area (EEA); products which are found to comply within one state are automatically deemed to comply within all others. "Placing on the market" means the *first* making available of the product within the EEA, so that the Directive covers only new products manufactured within the EEA, but both new and used products imported from a third country. Products sold second hand within the EEA are outside its scope. Where a product passes through a chain of distribution before reaching the final user, it is the passing of the product from the manufacturer into the distribution chain which constitutes placing on the market. If the product is manufactured in or imported into the EEA for subsequent export to a third country, it has not been placed on the market.

The Directive applies to each individual item of a product type regardless of when it was designed, and whether it is a one-off or high volume product. Thus items from a product line that was launched at any time before 1996 had to comply with the provisions of the

first Directive after 1 January 1996. Put another way, there is no "grandfather clause" which exempts designs that were current before the Directive took effect. However, products already *in use* before 1 January 1996 do not have to comply retrospectively. The transitional provisions for the second EMCD are that you can use the methods of compliance of either, from 20 July 2007 up to 20 July 2009.

The picture becomes less clear for products which are modified before being resold. If used apparatus, which has been modified during its operational life, is placed on the EU market then the procedure will depend on the extent to which those modifications affect the EMC performance of the apparatus. If the apparatus has been maintained and also modified, but not to any significant degree in terms of EMC, then it is termed "reconditioned" and is not subject to the provisions of the Directive. If the apparatus has been modified more substantially then it is referred to as "upgraded" and the Directive does apply. The choice remains in the hands of whoever carries out the modifications and places the product on the market, who is expected to carry out a new EMC assessment – or at least to validate the conclusions of the original assessment.

If the manufacturer resides outside the EEA, then the responsibility for maintaining the certificate of conformity with the Directive rests with the person placing the product on the market for the first time within the EEA, i.e. the manufacturer's authorized representative or the importer. Any person who produces a new finished product from already existing finished products, such as a system builder, is considered to be the manufacturer of the new finished product.

9.4.2.2.2.2 Taken into service

"Taking into service" means the first *use* of a product in the EEA by its final user. If the product is used without being placed on the market, if for example the manufacturer is also the end user, then the protection requirements of the Directive still apply. This means that sanctions are still available in each Member State to prevent the product from being used if it does not comply with the essential requirements or if it causes an actual or potential interference problem.

9.4.2.2.3 Exceptions

There are a few specific exceptions from the scope of the Directive. Self-built amateur radio apparatus is specifically excluded. Military equipment is excluded, at least in the UK, as a result of an exclusion clause in the Treaty of Amsterdam which established the European Union, but equipment which has a dual military/civil use will be covered when it is placed on the civilian market. Equipment which falls within the scope of the R&TTE Directive is specifically excluded, since this Directive refers back to the EMCD's essential requirements for its EMC provisions. Another exclusion is aeronautical products, parts and appliances that are referred to in Regulation (EC) No. 1592/2002 on common rules in the field of civil aviation, i.e. equipment that will be used on aircraft. Equipment for display at trade fairs need not comply provided that it carries a warning notice, and provided that "adequate measures are taken to avoid electromagnetic disturbances".

9.4.2.2.3.1 Benign equipment

The most widely drawn exclusion is for electromagnetically benign equipment. This is defined as equipment, the *inherent qualities* of which are such that:

(a) it is incapable of generating or contributing to electromagnetic emissions which exceed a level allowing radio and telecommunication equipment and other equipment to operate as intended; and

(b) it will operate without unacceptable degradation in the presence of electromagnetic disturbance normally consequent upon its intended use.

Both conditions must be met, but if they are, such equipment is completely excluded from the application of the Directive. As a general rule, the requirements are contingent on the nature of the apparatus: such that it cannot create power supply or RF disturbances, and does not include active electronic components which could be affected by such disturbances. Most electronic equipment could not be regarded as benign, since it wouldn't meet one or both of these criteria, but some simple electrical apparatus such as switches, fuses, induction motors and filament lamps may be.

The picture is slightly clouded by the EC's draft guidance, which allows as examples "Protection equipment which only produce transitory disturbances of short duration" and "High voltage types of equipment in which possible sources of disturbances are due only to localized insulation stresses which may be the result of the ageing process and are anyway under the control of other technical measures", even though such devices do create disturbances, which can affect radio reception.

9.4.2.2.3.2 Other Directives

The only other exclusions are for those types of apparatus which are subject to EMC requirements in other Directives or regulations, to the extent that those Directives specifically lay down, in whole or part, the essential EMC requirements. These are:

- 90/385/EEC active implantable medical devices;
- 98/79/EC in vitro medical devices;
- 93/42/EEC medical devices;

- 72/245/EEC EMC of vehicles as amended by 95/54/EC and 2004/104/EC;
- 75/322/EEC EMC of agricultural or forestry tractors as amended by 2000/2/EC;
- 97/24/EC two- and three-wheel motor vehicles;
- 96/98/EC on marine equipment;
- 90/384/EEC non-automatic weighing instruments.

In addition, equipment covered by the Measuring Instruments Directive 2004/22/EC which is CE marked and "M" marked according to that Directive, is excluded in respect of the immunity of such equipment. This exclusion is somewhat more complex, since the MI Directive is voluntary for some instruments. For the MI and non-automatic weighing instruments Directives, the EMCD still applies as far as emissions requirements are concerned.

9.4.2.2.4 Components

The question of when does a "component" (which is not within the scope of the Directive) become "apparatus" (which is) has been problematical. The Commission's 1997 guidelines introduced the concept of the "direct function", that is, any function which fulfils the intended use specified by the manufacturer in the instructions to the end user. It is available without further adjustment or connections other than those which can be performed by a technically naive user. Any component without a direct function is clearly not apparatus and is therefore outside the scope of the Directive. Thus individual small parts such as ICs and resistors are definitely outside the Directive.

Unfortunately, the second edition of the Directive, in defining what is meant by a component, has completely ditched this distinction. The question has now become, is the device intended to be placed on the market for incorporation into an apparatus by an end user, or likely to follow this path even if not intended directly for the end user? If so, then it is apparatus and must follow the full procedure required by the Directive. If not, then such components must be intended for incorporation into apparatus by other manufacturers, who take on the responsibility for compliance of their final product.

A component may be complex provided that its only purpose is to be incorporated inside an apparatus, but the manufacturer of such a component must indicate to the equipment manufacturer how to use and incorporate it. The distinction is important for manufacturers of board-level products and other sub-assemblies that may appear to have their own function and are marketed separately, yet cannot be used separately from the apparatus in which they will be installed.

However, in the particular case of plug-in cards for personal computers, which are supplied by a third party for the user to insert, the situation has been clarified: although such boards clearly need a host computer to have any purpose, they are placed on the market for the final end user and therefore need to carry a CE mark. They will need to be tested in a "representative" host computer, and certified accordingly. This position has been carried through into the harmonized standards, particularly EN 55022, which includes specific advice on how to treat such modules.

For products which may be both supplied to OEMs for incorporation into other apparatus, and supplied to the end user – an example might be some types of industrial sensor – then the item becomes apparatus and needs separate certification. If the manufacturer can insist that the item is only ever sold to OEMs then it is a component. This distinction has been made by many suppliers to shrug off the responsibility of ensuring that their products are properly specified for EMC ("Oh no, the Directive doesn't apply to us, we make components"). But since their customers are demanding EMC performance specifications anyway, to help them meet their own responsibilities, this is not a sustainable position.

9.4.2.3 The CE mark and the paperwork

The manufacturer or his authorized representative is required to attest that the protection requirements of the Directive have been met. This requires two things:

- he issues a declaration of conformity, which must be kept available to the enforcement authority for 10 years following the last date of manufacture (a subtle change from the first Directive, which referred to the last date of placing on the market);
- he affixes the CE mark to the apparatus, or if this is not possible, to its packaging, instructions or guarantee certificate, in that order of priority.

9.4.2.3.1 The CE mark

A further Directive concerning the affixing and use of the CE mark was adopted in 1993. This Directive harmonized the provisions regarding CE marking among the various previous new approach Directives. The mark consists of the letters CE as shown in Figure 9.4.1. The mark should be at least 5 mm in height and be affixed "visibly, legibly and indelibly" but its method of fixture is not otherwise specified. Affixing this mark indicates conformity not only with the EMC Directive but also with the requirements of any other Directives relevant to the product which provide for CE marking – for instance, an electrical toy with the CE mark indicates compliance both with the Toy

Fig. 9.4.1 The CE mark.

Safety Directive and the EMC Directive. Many electrical products fall under the scope of the LVD and the CE mark also indicates compliance with this. But during the transition period of any such applicable Directive, the CE mark need not indicate compliance; those Directives which *are* complied with should be listed in the appropriate documentation, such as the declaration of conformity.

It's worth pointing out that the CE mark is not in any sense a quality mark, since all products that desire European market access must have it. Nor is it an approval mark, since in most cases there has been no third party oversight to check that it is valid. Its presence is largely meaningless in any technical sense; if you want to know anything about the performance of a product, you need to demand much more than simply that it carries the CE mark.

9.4.2.3.2 The D of C

The EC Declaration of Conformity is required in every case. It must include the following components:

- a reference to the EMC Directive;
- a description of the apparatus to which it refers;
- the name and address of the manufacturer and, where applicable, the name and address of his authorized representative;
- a dated reference to the specifications under which conformity is declared;
- the date of the declaration;
- an identification of the signatory empowered to bind the manufacturer or his authorized representative.

It should also, perhaps self-evidently, certify that the apparatus to which it relates conforms with the protection requirements. You can write a single D of C in such a way that it covers all Directives that apply; you don't need a separate declaration for each.

Under the EMC Directive, an actual copy of the D of C is not required to accompany each item of equipment, although under the R&TTE Directive it is. Perhaps to forestall anticipated problems with some countries' enforcement officers who may be less than fully familiar with the distinction, many manufacturers take the safe route of including it in the documentation anyway. There is no requirement to have the CE mark on the D of C, but neither is it forbidden.

9.4.2.3.3 Description

The description of the apparatus should be straightforward; assuming the equipment has a type number, then reference to this type number (provided that supporting documentation is available) should be the starting point. The new EMCD is more stringent with regard to identification than the old one, requiring that equipment is identified by "type, batch, serial number or any other information" which would identify it. The description on the D of C should then encompass the range of such identifications so that there is an unambiguous correlation between the two.

Difficulties can arise when the type is subjected to revision or modification. At what stage do modifications or updates result in a new piece of equipment that would require re-certification? If the declaration of conformity refers to the Widget 3000 with software version 1.0 launched in 1996, does it continue to refer to the Widget 3000S of 2003 with version 3.2? The sensible approach would be to determine whether the modifications had affected the EMC performance and if so, re-issue the declaration for the new product; but this will require that you re-test the modifications, with the attendant cost penalties, or you exercise some engineering judgement as to whether a minor change will affect performance. No general guidance can be given on this point, but it should be clear that the breadth of the EMC requirements means that very few modifications will have absolutely no effect on a product's EMC performance.

9.4.2.3.4 Signatory

The empowered signatory will not necessarily be competent to judge the technicalities of what is being declared. Normally this will be one of the directors of the manufacturing or importing company, since the phrase used in the Directive is "empowered to bind" the manufacturer or his authorized representative. In small companies the technical director will probably be close enough to the product in question to understand the detail of its EMC performance, but in medium or large-scale enterprises the directors will expect to rely on the technical advice of their product development and manufacturing engineers and/or the EMC test and management personnel. Such companies will have to define clearly the levels of responsibility that exist for each person involved in making the declaration.

9.4.2.3.5 Specifications

In most cases the references to the specifications under which conformity is declared, will be a list of the harmonized standards – with dates, to distinguish which version has been used – that have been applied to the

apparatus. If harmonized standards have been used only partially or not at all, a reference to the technical documentation (see Section 9.4.3.1.3) where the detailed technical EMC assessment is given should be included as well as a reference to any identifiable non-harmonized standards or specifications that have been applied.

The reference to specifications raises the question of whether you have to *test* to these specifications. Three possibilities are apparent. First, you may deem that the product intrinsically meets the requirements of the Directive and does not need testing. An example might be a simple linear unregulated stand-alone power supply which is below the power level at which harmonic currents are controlled. You may sometimes be able to convince your signatory that this is a competent engineering judgement, but many electronic products will not be able to follow this option.

Second, you may be able to make a declaration based on pre-existing test results. If for example you have already been conforming to existing non-harmonized standards, then you may be confident enough to state that the product will meet the appropriate harmonized standards without further testing, or with only partial testing.

The final option is to test fully to harmonized standards. For a sophisticated product this will be lengthy and expensive, and may involve some complex judgements as to what tests to apply. For new products though, testing will be essential. In fact, the new EMCD has subtly introduced a significant extra hurdle: it says (in the preamble)

> *Compliance with a harmonised standard means conformity with its provisions and demonstration thereof by the methods the harmonised standard describes or refers to.*

This appears to mean, in effect, that you must follow in every case the full compliance test method, in all its complexity and expense, whenever you are going to make a compliance declaration with a reference to a harmonized standard. The concept of precompliance testing, used successfully for many years in pursuance of the old Directive, has been closed off. However, what is almost certain to be the outcome of this is that pre-compliance testing will continue, supported by a brief statement in the technical documentation that harmonized standards have been applied "in part", and justifying the effectiveness of such partial application. In essence, the manufacturer will make an assessment of the risk of not following harmonized standards in full versus the risk (in terms of cost to the company) of doing so.

A fourth option, of course, is not to test at all; just make the declaration, stick on the CE mark and hope that nobody ever notices. A reputable company, of course, won't take this route, but the possibility of competitors doing so may be a factor in assessing your market position.

9.4.2.3.6 Other information

In keeping with the EC's general desire to keep the user informed, the second edition EMCD makes new requirements for information to be supplied with each apparatus. As well as identification of the apparatus and the name and address of the manufacturer, there should be:

- information on any specific precautions that must be taken when the apparatus is assembled, installed, maintained or used in order to ensure that when put into service the apparatus complies with the protection requirements;
- a clear indication (easily understood by the intended user) of restriction of use for apparatus for which compliance with the protection requirements is not ensured in residential areas;
- information required to enable the apparatus to be used in accordance with its intended purpose.

It is assumed that if no information is given with the apparatus, then users can install, use and maintain it without any special considerations regarding its EMC aspects, and it will still comply with the protection requirements. This is naturally the preferred default position, but the possibility exists that you could, for instance, require the use of, say, particular cables or earthing regimes if these were necessary for compliance, provided that this was made clear to the user.

The residential area limitation has been a knotty problem for the Commission for some time. It has its origins in the text of a warning notice which appears in CISPR 22/EN 55022, the IT equipment emissions standard, referring to Class A equipment:

> *Such equipment should not be restricted in its sale but the following warning shall be included in the instructions for use:*
>
> ***Warning***
>
> *This is a class A product. In a domestic environment this product may cause radio interference in which case the user may be required to take adequate measures.*

This text has caused headaches for the lawyers because, properly speaking, Class A equipment should never be marketed for or used in the domestic environment. The text is in the international (CISPR) version of the standard and other non-European countries, which have varying legal requirements, do not want to see it removed. The adopted compromise has been to delete the phrase "Such equipment should not be restricted in its sale" from the EN modified version of CISPR 22, and

make it also a requirement in the new EMCD that "apparatus for which compliance with the protection requirements in residential areas is not ensured by the manufacturer has to be accompanied by a clear indication of this restriction of use".

9.4.2.4 Manufacturing quality assessment

The Directive covers every individual, physically existing finished product, but it would be impractical to test every item in series production fully for all the EMC characteristics that it must exhibit. The conformity assessment procedures for all the technical harmonization Directives are contained in Council Decision 90/683/EEC. This document contains a range of modules which may be applied in the case of each specific Directive. However, the EMC Directive does not specifically refer to this Decision, and therefore conformity assessment requirements have been left somewhat open.

9.4.2.4.1 Production control

The second edition Directive requires in annex II that "the manufacturer must take all measures necessary to ensure that the products are manufactured in accordance with the technical documentation … and with the provisions of this Directive that apply to them". No specific means of determining what these measures might be are mentioned in either the Directive or the Council Decision.

9.4.2.4.1.1 CISPR sampling schemes

For many years before the adoption of the EMC Directive, the standards committee CISPR had recognized the need for some form of production quality testing, and had incorporated sampling schemes into the RF emission standards which form the basis of EN 55011, EN 55014 and EN 55022. The purpose of these schemes is to ensure that at least 80% of series production complies with the limits with an 80% confidence level, the so-called 80/80 rule. Practically, to comply fully with the 80/80 rule the manufacturer has to aim at about 95% of the products being in compliance with the specified limit.

The first scheme requires measurements of the actual emission levels from between 3 and 12 identical items, from which the mean and standard deviation are derived. The limit levels are then expressed in the form:

$$L \geq \overline{X} + k \cdot S_n \qquad (9.4.1)$$

where $\overline{X}$ is the arithmetic mean and S_n the standard deviation of the measured emission levels, and k is a constant derived from the non-central t-distribution between 2.04 and 1.2 depending on sample size.

Table 9.4.1 Acceptable number of failures versus sample size n

n (sample size)	7	14	20	26	32
c (no. of failures)	0	1	2	3	4

If the emission levels are similar between items (a low value of S_n) then a small margin below the limit is needed; if they are highly variable, then a large margin is needed. This sampling method can only be applied to emissions measurements and cannot be used for immunity.

A second scheme which is applicable to both emissions and immunity is based on recording test failures over a sample of units. Compliance is judged from the condition that the number of units with an immunity level below the specified limit, or that exceed the emissions limits, may not exceed c in a sample of size n, as per Table 9.4.1: this test is based on the binomial distribution and produces the same result as the first, in accordance with the 80/80 rule.

As well as the above sampling schemes, published EN standards also allow a single test to be made on one item only, but then advise that subsequent tests are necessary from time to time on samples taken at random from production. The "one item only" dispensation is, of course, used in the vast majority of cases, although manufacturers with a high production volume and, especially, their own in-house test facilities, may well carry out some sample testing to reinforce and protect their compliance position.

Significantly, according to the CISPR standards the banning of sales is to occur only after tests have been carried out in accordance with one or other sampling scheme. This appears to put an onus on the enforcement authorities always to carry out testing of multiple samples in their market surveillance campaigns.

9.4.2.5 Fixed installations

The question of how the EMCD applies to installations has been an issue ever since the original Directive was under discussion:

> *… Finally, there is something of a problem concerning systems comprising assemblies of apparatus. In some cases, the EMC of a system can differ from the EMC of its constituent parts, and a system which is marketed as such requires separate certification under the Directive. To a large extent, the problems can be resolved by ensuring that the standards contain adequate practical test methods for systems, but some systems can only be tested "in situ" – ie long after the marketing stage. A similar problem applies to "one off" items of large*

capital equipment. We are trying to obtain clarification of the legal implications and will report back.

John Ketchell, DTI, 28 February 1989, in a circular to interested parties regarding progress on the negotiations prior to introducing the EMC Directive

The application of the EMC Directive to installations is, from the experience obtained over the last four years, a very controversial issue.

EC Guidelines on the application of the EMC Directive, July 1997, section 6.5.2.1

Throughout the life of the first EMCD, systems and installations have been a bone of contention. So it is hardly a surprise that the biggest effect of the second edition is to substantially change the compliance regime for installations.

9.4.2.5.1 Changes in the second EMCD

The second EMC Directive has a separate section on fixed installations, defined as follows:

Fixed installation means a particular combination of several types of apparatus and, where applicable, other devices, which are assembled, installed and intended to be used permanently at a predefined location

This distinguishes a fixed installation from a mobile installation, which has its own definition as "a combination of apparatus and, where applicable, other devices, intended to be moved and operated in a range of locations", and which is subject to the same compliance regime as other apparatus. A fixed installation has to be installed applying "good engineering practices" and respecting the information on the intended use of its components, with a view to meeting the protection requirements. Those good engineering practices are to be documented, and the documentation held by the person(s) responsible for as long as the fixed installation is in operation. However, fixed installations do not need to be CE marked or subject to a declaration of conformity. Also, apparatus that is intended for incorporation into a specified fixed installation does not have to comply with certain aspects of the Directive.

This represents a significant change in the way that fixed installations and the apparatus that goes into them are regulated. Not all of the implications are yet clear, but one is that operators of installations will have to designate a "responsible person" who has responsibility for the establishment of compliance of a fixed installation with the relevant essential requirements. This will be new to most if not all such operators. Another is that the vague term "good engineering practices" will have to be fleshed out; at present there are no harmonized standards for installations, but the Commission clearly envisages that some will have to be created.

The principal dilemma of applying the Directive to complete installations is that to make legally relevant tests is difficult, but the nature of EMC phenomena is such that to test only the constituent parts without reference to their interconnection is meaningless. Two main possibilities have been explored in TCFs under the first EMCD. The first is to ensure that the system is built out of individual items which are themselves compliant, and that the method of installation follows suppliers' instructions such that this compliance is not breached. The TCF would reference installation drawings and work instructions that ensure this. This approach is continued and expanded in the new regime.

The alternative has been to perform limited site testing once the installation has been assembled to show that it is compliant. Since this would be less comprehensive, it would need to be balanced by a greater amount of documentation in the form of a matrix defining the EMC threats and a rationale for the claim to compliance, including a justification for the tests that were done. The new regime does not mention testing in any specific way.

9.4.2.5.2 Scope

The definition shown above for a fixed installation distinguishes it from a large system (discussed below in Section 9.4.2.6) in that it is not a single functional unit, that is, it is "a particular combination of several types of apparatus", and it is not placed on the market and cannot enjoy "free movement" around the EU: it is "assembled, installed and intended to be used permanently at a predefined location". The phrase "intended to be used permanently" has caused problems in interpretation; some people have thought that it meant being used 24 h a day. Actually, the likely interpretation (still to be fully accepted at the time of writing) is related to duration. The requirement would be fulfilled if it were intended at the time of putting into service that the constituent parts were to be used in the defined location for their expected lifetime. If they were expected to be moved for use at another location, it would not.

It is recognized that modifications will be made to fixed installations throughout their operational life, and such modifications would not invalidate the original intent. Indeed, many such installations (such as telecommunications centres) are in a continuous state of modification, and the legislation cannot be drafted in such a way as to prevent this. But the implication of the requirements are that the EMC dimension of such modifications must always be taken into account and properly documented.

One problem that arises when considering fixed installations is what the boundaries of the installation are, and therefore what parts have to be considered – and indeed, what entity should take responsibility for it. For

instance, in a campus with a network of power supplies and telecommunication lines, where is the physical "edge" of the installation? Sometimes the service entrance of the conductors can be easily located, but not always, and this certainly does not define the boundary for radiated phenomena. Knowing the boundary should enable the electromagnetic environment to be defined, but there will be many examples where the boundary is fuzzy at best. The question also introduces the idea of "nested" installations, that is, smaller installations within a larger one, with different boundaries and responsibilities. The DTI's draft guidance wryly points out that in the case of interference between installations in such a situation, "it is likely that only by further co-operation between the responsible persons will a solution be able to be found".

There is another consequence of the generalized definition given above, which if pursued to its logical end could have dramatic consequences. This is that there is no effective lower limit on the size or extent of the installation. It is clear, historically, that the definition was meant to encompass large premises such as telecommunication centres, factories, power stations or substations, and so forth. But actually, any location which has "a particular combination of several types of apparatus" that are intended only for that location would fall within the scope, and therefore need a responsible person and a documentation of good practice. It is possible that this definition could extend down to the level of individual shops, offices and even homes, if electrical or electronic equipment is installed in them which is specific to that one location. Such a possibility is not excluded by the EC's draft guidance, which suggests "The definition covers all installations from the smallest residential electrical installation through to national electrical and telephone networks, including all commercial and industrial installations."

9.4.2.5.3 The responsible person

Annex I of the EMCD requires that

> *good engineering practices shall be documented and the documentation shall be held by the person(s) responsible at the disposal of the relevant national authorities*

and article 13.3 says

> *Member States shall set out the necessary provisions for the identifying the person or persons responsible for the establishment of compliance of a fixed installation*

In the UK, the DTI has drafted a definition of the responsible person as the person who, "by virtue of their ownership or control of the relevant fixed installation is able to determine that the configuration of the installation is such that when used it complies with the protection requirements". It is accepted that this may be a different entity according to circumstances. For example in some cases it could be the site owner, in others the operator of the installation, in others the maintainer of the installation, and so on. Operators will need to identify the responsible person before the installation is taken into service, and often this will have to be determined contractually.

The implication of the DTI's definition above is that whoever the entity turns out to be for a given case, they will need to have a sufficient understanding of the EMC issues raised by the installation to be able to make the required judgement, as well as having sufficient seniority to be able to take on the required responsibility. As a matter of simple observation, such individuals, groups or departments rarely exist in typical organizations that administer installations, although there are exceptions: for instance, large telecom organizations will usually have the technical capability in-house, as will groups in some sectors that already deal with installation EMC, such as the railway industry.

Consequently, we can envisage either of two outcomes:

- either organizations will take their responsibilities seriously, and appoint or employ an individual, a consultant or an in-house group to deal with the likely proliferation of documentation that the new regulations imply; or
- comprehension of the scope and implications of the new regulations will be so minimal, and their enforcement so weak, that little if anything will be done to bring installations into compliance with the new regulations.

9.4.2.5.4 Good engineering practices

There is no definition in the EMCD of what is meant by the phrase "good engineering practices", except that by implication they are capable of being documented. There are various good practices that have been established for electrical installations, for instance in the UK the IEE's Wiring Regulations, now enshrined in BS 7671; but these have nothing significant to say about EMC and certainly wouldn't be useful in complying with the essential requirements. It is an emerging consensus that what is meant is "good *EMC* engineering practices", but this leaves the field wide open for interpretation. To date, such interpretation has been highly generalized, pointing at best to a few IEC documents such as those in the IEC 61000-5 series and to the existence of some books on the subject, a canon to which this author amongst others has already contributed. Much of the more detailed material in fact applies to military systems and would be irrelevant or hard to apply in the context of the EMC Directive.

Certainly, it seems very unlikely that any responsible person would be prosecuted for non-compliance on the basis that their documented engineering practices were actively bad for EMC; even if this were to be the case, actually proving it in a court of law would be expensive and uncertain in outcome at best. For instance, many system designers and installers use traditional engineering practices that they assume are good, such as single-point earthing and terminating cable screens at one end only using long lengths of wire, when in fact these are inimical to good EMC at the frequencies of interest for the EMCD's protection requirements. To prove the case, though, would require long, involved and probably inconclusive technical arguments that would not be compatible with an effective enforcement action. What is left is the requirement to "respect the information on the intended use of components", and this boils down to ensuring that if any such components include strictures on their installation and operation that have relevance for EMC, these should be observed and the fact that they have been observed should be documented.

The level of detail that is needed in the documentation will vary according to the complexity of the fixed installation. It should be sufficient to enable an enforcement authority to determine whether good practices have been followed. When interpreting this requirement, you should bear in mind that enforcement authorities (certainly in the UK) are not experts in EMC.

One point to note about the good practices that might be adopted is the distinction between intra-system and inter-system EMC. The EMCD is only concerned with interference between the fixed installation and the outside world, that is, *inter*-system EMC. So, although of great interest to the owner and operator of the installation, good practices for *intra*-system EMC, which are intended only to ensure mutual compatibility of the various parts of the installation, are of themselves not relevant for the EMCD.

9.4.2.5.5 Specific apparatus

Apparatus that is intended for incorporation into a specified (named) fixed installation – and not otherwise commercially available to an end user as a single functional unit – does not have to comply with:

- the essential requirements;
- the conformity assessment procedure;
- the technical documentation and EC declaration of conformity – but some technical documentation is required;
- CE marking.

The basis for this difference is that it is not necessary, and sometimes not possible, to carry out a conformity assessment for apparatus that is intended solely for a specific fixed installation, in isolation from that installation. This exemption cannot be used for apparatus which may become available in a general sense. Apparatus can only benefit from it if there is a direct link between the manufacturer of that specific apparatus and those responsible for the fixed installation for which that apparatus is intended. A relationship, probably contractual, between the provider and the customer is required. The information requirements remain, that is, the apparatus must be fully identified and include the name and address of its manufacturer; and it needs to document exactly which fixed installation it is intended for "and its EMC characteristics" and must indicate "the precautions to be taken for the incorporation of the apparatus into the fixed installation in order not to compromise the conformity of that installation".

The reference to EMC characteristics of the installation implies that the manufacturer of such apparatus must be familiar with the installation's electromagnetic environment. The level of detail needed, and the form that the information should take, is not given: it could mean as little as classifying the fixed installation as having a domestic, commercial, light-industrial, or industrial electromagnetic environment as per the generic standards; but this would hardly be enough to design a product that would not compromise the compliance of the installation. Similarly, indicating the precautions to be taken for EMC may simply mean listing well-known EM mitigation measures and letting the installer decide what to do; or it could mean giving the installer clear and detailed instructions for every aspect of the installation. Both of these issues are so novel, and the available guidance so limited, that for the present it is hard to see how an acceptable general practice could be developed to deal with the requirements.

The Commission's draft guidance envisages a number of instances where the exemption for specific apparatus might be used:

- specific apparatus made according to particular specifications and intended for a fixed installation or installations considered as equivalent by the manufacturer;
- apparatus designed according to a specification given by a customer, intended for a given fixed installation;
- apparatus derived from a generic model adapted to the specific need of the customer or to the specificity of any particular location, in a fixed installation;
- apparatus made in small series and delivered for incorporation into a well-defined type of installation, each item necessitating appropriate EMC adjustments at the final location.

9.4.2.6 Systems

A fixed installation cannot "enjoy free movement" within the EEA, in contrast to a system (or a moveable installation) which can. A typical system might be a personal computer workstation comprising the PC, monitor, keyboard, printer and any other peripherals. If the units were to be sold separately they would have to be tested and certified separately; if they were to be sold as a single package then they would have to be tested and certified as a package.

The definition unfortunately does not help system builders who will be "placing on the market" – i.e. supplying to their customer on contract – a single system, made up of separate items of apparatus but actually sold as one functional unit, not necessarily intended for a predefined location. Many industrial, commercial and public utility contracts fall into this category. According to the published interpretation, the overall assembly should be regarded as a system and therefore should comply as a package. As it stands at present, there are no standards which specifically cover large systems, i.e. ones for which testing on a test site is impractical, although some emissions standards do allow measurements in situ. These measurements are themselves questionable because of the difficulty of distinguishing external interference at the measurement position from that due to the installation, and because the variability of the physical installation conditions introduces reflections and standing waves which distort the measurement. There are no provisions for large systems in the immunity standards. Therefore the EMC assessment has to be performed in the absence of standards, but there is little guidance as to how to interpret the Directive's essential requirements in these cases.

9.4.2.7 Implementation, enforcement and sanctions

Member States cannot impede for EMC reasons the free circulation of apparatus covered by the Directive which meets its requirements when properly installed and maintained, and used for its intended purpose. They must presume that apparatus which bears the CE mark, and for which the manufacturer has performed and documented an EMC assessment, does in fact comply with the protection requirements unless there is evidence to the contrary.

On the other hand, Member States are required to ensure that equipment which is found not to comply is not placed on the market or taken into service, and to take appropriate measures to withdraw non-compliant apparatus from the market. Legislation which translates the Directive's requirements into national law in each Member State was required to be in place by 20 January 2007.

In the UK the enforcement regime includes the issue of "suspension notices" which prohibit the supply or use of specified equipment that the enforcement authorities believe does not comply with the EMC requirements. The notice may or may not have immediate effect, depending on the urgency of the situation; an appeal procedure allows persons on whom a notice is served to make representations for it to be revoked. Enforcement authorities can also apply to a court for forfeiture of apparatus, with its consequent destruction, modification or disposal, and officers of the enforcement authorities may be empowered to enter premises and inspect or seize apparatus and documents.

9.4.2.7.1 Offences

The EMC legislation does include criminal sanctions. But because of the difficulty of judging whether or not apparatus actually complies with the requirements, the UK legislators have not created an absolute criminal offence of supplying or using non-compliant equipment. Users and retailers cannot normally be expected to know whether or not the apparatus in question is non-compliant. Criminal offences on other fronts are necessary, for instance to guard against misuse of the CE mark or the provision of false or misleading information, and to penalize breaches of prohibition notices.

9.4.2.7.2 Practice

Two important questions are: how is enforcement operated in practice, and is the Directive enforced equally in all Member States (the so-called "level playing field"). These questions are directly related to the resources that national governments are prepared to devote to the task. The UK DTI has previously indicated that its enforcement efforts are complaint-driven. As well as investigating interference complaints arising from actual use of apparatus, it is open to complaints that apparatus does not conform to the Directive's requirements regardless of whether or not there is a problem in its use. A possible source of complaint will therefore be from companies testing samples of their competitors' equipment and, if they find that it does not comply, "shopping" them to the authorities. It is also likely, though, that such complaints will need to be backed by serious evidence of non-compliance before the authority will take them seriously.

On the other hand, the German authorities have stated that it is necessary to gain information from the market in the form of random spot checks in order to react to violations. Germany already had a strong regime for the control of RF emissions before the first EMCD in the form of the mandatory VDE standards, and these were stricter than the EN standards which are now used to demonstrate compliance with the Directive.

The Germans have been concerned that the Directive might dilute the effectiveness of their previous regime, and have therefore insisted that it is thoroughly enforced.

It is apparent that differences in enforcement practices within the various Member States work contrary to the stated intent of the Directive, which is to reduce technical barriers to trade. Article 10 of the Directive requires that "where a Member State ascertains that apparatus bearing the CE marking does not comply with the [protection] requirements, it shall take all appropriate measures to withdraw the apparatus from the market, prohibit its placing on the market or restrict its free movement", and shall immediately inform the Commission of any such measure. If the Commission finds, after consultation, that the action is justified, it will inform all other Member States. The competent Member State shall then take appropriate action against the author of the attestation. Therefore any Member State can take immediate action to prohibit an offending apparatus from its own market, but sanctions against the company that put the apparatus on the market in another Member States are dependent on the deliberations of the Commission and on the enforcement practices of the latter Member State.

9.4.2.7.3 Interpretation

As has already been indicated, the Directive is so widely drawn that many of its provisions have to be the subject of interpretation. You might expect that this interpretation would be the function of the EC or of the national implementing authorities, but it has been remarkably difficult to obtain answers to detailed questions of interpretation from such authorities, especially so when there is a technical dimension to the question. The 1997 Guidelines gave considerable interpretive help, but do not cover many more abstruse technical issues.

The new Directive also has a guidance document, a draft of which is referred to at various points throughout this chapter. The intent of this is that it should have been available in advance of Member States' transposition of the Directive into their own laws, so that such transposition took place more uniformly than was the case for the first Directive. Because of the delays and contention in creating the document, this has happened only imperfectly. Furthermore, it is intended as an internet-based document, partly for ease of access and linking to other sources of information, but also so that it can be readily updated if the need arises. This latter aim may not be seen entirely as an advantage by its users, who generally would like such guidance to be consistent rather than fluid, but the Commission is distinctly in favour of it.

In the UK, a group known as the EMC Test Laboratories Association (EMCTLA) was formed primarily to ensure a reasonable uniformity of approach to the assessment of technical construction files under the first EMCD by different competent bodies. Within that association, a working group has been set up which responds to queries regarding the implementation of the Directive and issues technical guidance notes which have a reasonably wide circulation, and which can form the basis for a uniform interpretation. These guidance notes are available via the EMCTLA's website; many are still relevant, despite the removal of the TCF route. The EMCTLA has been instrumental in founding a pan-European Association of Competent Bodies which is intended to spread this uniform approach within the EU.

9.4.3 Compliance of apparatus with the Directive

Of themselves, the essential protection requirements are too generalized to enable manufacturers to declare that their product has met them directly. So annexes II and III of the Directive provide a more detailed route (Figure 9.4.2) for manufacturers to achieve compliance with them.

9.4.3.1 Self certification and internal production control

The route which is expected to be followed by most manufacturers is that given in annex II, known as self certification with internal production control. This requires you as the manufacturer firstly to perform

> *an electromagnetic compatibility assessment of the apparatus, on the basis of the relevant phenomena, with a view to meeting the protection requirements … The*

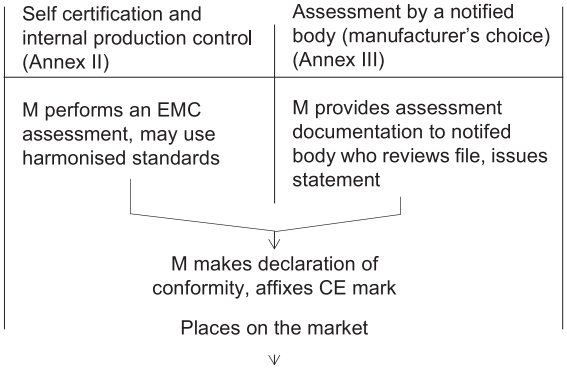


Fig. 9.4.2 Routes to compliance. M: manufacturer, authorized representative within the EU, or person who places apparatus on the EU market

electromagnetic compatibility assessment shall take into account all normal intended operating conditions. Where the apparatus is capable of taking different configurations, the electromagnetic compatibility assessment shall confirm whether the apparatus meets the protection requirements … in all the possible configurations identified by the manufacturer as representative of its intended use.

On the assumption that the assessment does indeed confirm that the apparatus meets the protection requirements, you then have to

draw up technical documentation providing evidence of the conformity of the apparatus with the essential requirements of this Directive, [and] hold the technical documentation at the disposal of the competent authorities for at least ten years after the date on which such apparatus was last manufactured.

This documentation requirement is no more than you would have done under the first edition of the Directive, since you would have kept your own CE marking documentation on file for your own purposes, whether or not you ever involved a third party such as a Competent Body. In other words, there is little change so far from the first Directive. Having developed the documentation, you then make the Declaration of Conformity and prepare to retain that also for 10 years. You are then free to apply the CE mark and market the apparatus, ensuring that each product is manufactured in accordance with the technical documentation and in compliance with the essential requirements (hence, "internal production control").

9.4.3.1.1 The EMC assessment

We now have to interpret what is meant by the phrase "EMC assessment". Annex II.1 specifically refers to "on the basis of the relevant phenomena, with a view to meeting the protection requirements". It also makes clear that the assessment must cover all configurations and normal operating conditions. In the first edition of the Directive, either you had to apply harmonized standards or you had to generate a Technical Construction File and run it past a Competent Body. In this new regime, all the responsibility of what to do is in your hands. The Commission's view is, fairly clearly, that having let manufacturers in gently with the first EMCD, the time has come for them to grow up and deal with the EMC of their products on their own.

Use of harmonized standards in the assessment is not mandatory, but if they are not used or are used only in part, you then have to decide how you are going to show that the protection requirements have been met. This means that the assessment will have to look at least at:

- a description of the apparatus' operating conditions and its intended use;
- the environments in which the apparatus will be used;
- the electromagnetic phenomena covered and compatibility levels applied;
- the performance criteria of the apparatus;
- test plan, including techniques, emissions limits and immunity stress levels;
- EMC design and installation techniques used;
- the user's information requirements, including any restriction (or not) on use in residential areas, and any precautions on installation;
- definition of the test selection matrix for variants, using worst case criteria.

Clearly, in the absence of other guidance, this could be a substantial exercise in its own right. So at this point, the Directive refers you to standards.

9.4.3.1.2 Application of harmonized standards

Annex II.1 says

The correct application of all the relevant harmonised standards whose references have been published in the Official Journal of the European Union shall be equivalent to the carrying out of the electromagnetic compatibility assessment.

Article 6.2 goes on to say

The compliance of equipment with the relevant harmonised standards whose references have been published in the Official Journal of the European Union shall raise a presumption, on the part of the Member States, of conformity with the essential requirements referred to in Annex I to which such standards relate.

But

This presumption of conformity is limited to the scope of the harmonised standard(s) applied and the relevant essential requirements covered by such harmonised standard(s).

Harmonized standards are those CENELEC, CEN or ETSI standards which have been announced in the *Official Journal of the European Union* (OJEU). In the UK these are published as BS EN standards, and each Member State has a similar system, administered by its own national standards body, for publishing them. They are not available from CENELEC itself,

although ETSI ones can be obtained (for free) direct from ETSI.

The potential advantage of using standards in the EMC assessment from the manufacturer's point of view is that the assessment process is simplified. There is no mandatory requirement for testing by an independent test house. The only requirement is that the manufacturer makes a declaration of conformity (see Section 9.4.2.3.2) which references the standards against which compliance is claimed. Of course the manufacturer will normally need to test the product to assure himself that it actually does meet the requirements of the standards, but this could be done in-house. Many firms will not have sufficient expertise or facilities in house to do this testing, and will therefore have no choice but to take the product to an independent test house. This is discussed further in Section 9.4.3.3. But the long-term aim ought to be to integrate the EMC design and test expertise within the rest of the development or quality department, and to decide which standards apply to the product range, so that the prospect of self certification for EMC is no more daunting than the responsibility of functionally testing a product before shipping it.

9.4.3.1.3 The technical documentation

The documentation needs to support the case that you have made using the EMC assessment for meeting the protection requirements for any particular product. One requirement for placing on the market in the draft UK regulations is that "the technical documentation has been prepared and is available".

9.4.3.1.3.1 Contents

Annex IV of the Directive specifies the general content of the documentation, whose overall purpose is "to enable the conformity of the apparatus with the essential requirements to be assessed". The content includes:

- a general description of the apparatus (one purpose of this is to allow an unambiguous link between the actual apparatus and the documentation);
- evidence of compliance with the harmonized standards, if any, applied in full or in part;
- where the manufacturer has not applied harmonized standards, or has applied them only in part, a description and explanation of the steps taken to meet the essential requirements of the Directive.

The third of these of course is the record of the EMC assessment described in Section 9.4.3.1.1. Notice that the second requirement is for "evidence" of compliance with standards; this may well not mean a full test report. What constitutes a minimum of "evidence" has yet to be ascertained.

As we have already mentioned, the products that are placed on the market must be "manufactured in accordance with the documentation". Since products change and develop over their lifetime, this means that the documentation should keep pace: as is necessary in any case, you will need to assess engineering changes for their EMC implications. Any that ripple through to affect the content of the technical documentation should result in appropriate changes to it.

Whilst the purpose of the documentation is clear enough, what is not clear is who will be doing the assessment of conformity with the essential requirements – in other words, who is the documentation being written for? Will it be an enforcement officer who has just a sketchy knowledge of EMC technicalities, or will it be an EMC consultant who is prepared and able to pick holes in even the densest material? Presumably, this will depend on the context of whatever enforcement action is being taken; so it is up to you as technical author to anticipate this context and set the technical level of the documentation appropriately.

9.4.3.2 The notified body

A substantial change between the first and second editions of the Directive is the removal of the old Technical Construction File route to compliance, and the concomitant disappearance of Competent Bodies. Instead, we now have Notified Bodies (NBs) and their use by the manufacturer is entirely voluntary.

Annex III of the Directive describes the use of Notified Bodies (annex II is the self certification/internal production control procedure discussed above):

1. *This procedure consists of applying annex II, completed as follows:*
2. *The manufacturer or his authorized representative in the Community shall present the technical documentation to the notified body referred to in article 12 and request the notified body for an assessment thereof. The manufacturer or his authorized representative in the Community shall specify to the notified body which aspects of the essential requirements must be assessed by the notified body.*
3. *The notified body shall review the technical documentation and assess whether the technical documentation properly demonstrates that the requirements of the Directive that it is to assess have been met. If the compliance of the apparatus is confirmed, the notified body shall issue a statement to the manufacturer or his authorized representative in the Community confirming the compliance of the*

apparatus. That statement shall be limited to those aspects of the essential requirements which have been assessed by the notified body.

4. *The manufacturer shall add the statement of the notified body to the technical documentation.*

There are a number of features to point out here.

9.4.3.2.1 The primacy of the manufacturer

Not only is it entirely at the manufacturer's discretion whether or not to use an NB, but the manufacturer tells the NB what to do, not vice versa (annex III.2). You may choose to ask an NB to assess the documentation as a whole, covering the totality of the essential requirements, or only in part. Of course, the statement that the NB issues will only cover those aspects that it has reviewed. So, for instance, you may be entirely comfortable with your compliance with emissions limits but you may want an opinion on how the product complies with the immunity requirements.

The NB only provides a "statement", and only if the compliance is confirmed. The statement has to be included with the technical documentation, but there is no indication that there is any responsibility on or ability of the NB to prevent the manufacturer placing the product on the market if it believes that compliance is not confirmed; that remains firmly the responsibility of the manufacturer. Manufacturers are free to choose any NB. There is no need to choose an NB located in the country where the apparatus is manufactured, nor in the country to which the apparatus will be shipped, marketed or taken into service. If the manufacturer has used the services of an NB for one of his products there is no obligation to use the same NB for another one, or for modifications of the originally assessed product.

A manufacturer may consult more than one NB simultaneously in respect of any apparatus: there is no prohibition on this, and the NB's opinion is not binding. In other words, it is apparently legitimate to "shop around", should you so desire, until you get a favourable opinion.

The Directive at article 12 specifies how NBs are to be appointed. The requirements are similar to those for Competent Bodies under the first Directive, but sufficiently different that it is unclear as to whether all such existing Competent Bodies will choose to or be able to convert to NB status. An NB can only render services within its scope of designation which may be limited to certain categories of apparatus, certain essential requirements or certain other specific aspects. But for the time being there are no agreed criteria within the EU for the definition of "scope of designation"; different countries are likely to have different ideas.

9.4.3.2.2 When would a manufacturer use an NB?

At first sight, there seems to be little point in consulting an NB. It adds expense, but has no teeth and is not a requirement of the process.

A common use of the TCF route and the involvement of a Competent Body under the first edition Directive was to certify the compliance of a fixed installation. As we have seen, the requirements for installations have changed drastically, and Notified Bodies have no role in assessing the good practice documentation that is required, even though that would be one of their most useful functions, and one in which (as Competent Bodies) they actually have gained some experience. Their role is limited entirely to assessing the compliance of *apparatus*.

The NB cannot perform the product's EMC assessment or generate the technical documentation itself: in no way does it supplant the annex II procedure, which remains firmly in the hands of the manufacturer. The NB is, in fact, required to be independent in the process of "preparing the reports and performing the verification function". (This may not stop you contracting one NB to create the documentation and another to assess it, if you are prepared to sign up to the result.)

The main driving force for involving an NB will be where the manufacturer requires reassurance by a third party to increase his confidence in the compliance process, or where the equipment is so complex that its assessment requires expert guidance. In most cases, this could be done by any competent EMC consultant; the use of an NB merely gives the imprimatur of appointment by the statutory authority. However, the NB statement can be seen as a shield against future enforcement action, and some manufacturers will be prepared to pay for such insurance. It may also be viewed in a few cases as an extra marketing tool.

9.4.3.3 Testing

Except in the case of benign products, which it is clear will intrinsically not cause interference or be susceptible to it, such as an electric fire or pocket torch, each manufacturer will need to submit products to some degree of EMC testing to be sure that they comply with the Directive. To cover the eventual requirements of the standards, the scope of the tests will need to include mains harmonic, conducted and radiated RF emissions, plus immunity to RF, transients, electrostatic discharge and supply disturbances. A test facility to address all these phenomena at compliance level is beyond the budget of all but the largest companies. Not only are a screened room and/or an open area test site and all the test equipment needed, but also the staff to run the

facility – which itself requires a level of skill, experience and competence not usually found in most development or test departments. A large company may have the product volume and available capital which allows investment (of the order of £1 m) in an in-house facility of this nature, and there are several such companies throughout Europe who have taken this step. The EMCD places no external constraints on the operation of these in-house test facilities.

9.4.3.4 Using standards

Compliance with harmonized standards is "equivalent to" the performance of the EMC assessment (see Section 9.4.3.1.2). This mechanism depends on the availability of standards which can be applied to the product in question. This section will discuss their general availability and applicability.

Prior to the adoption of the first EMC Directive, the EMC standards regime had developed in a somewhat piecemeal fashion. The existing standards fell into a number of categories:

- RFI: intended to protect the radio spectrum from specific interference sources, such as information technology equipment, motor vehicle ignition, household appliances or fluorescent lights
- mains emissions: specifically harmonic currents and short-term variations, to protect the low-voltage power distribution network
- product- and industry-specific: to ensure the immunity from interference of particular types of product, such as process instrumentation or legal metrology, or to regulate emissions from equipment that will be used in a specific environment, such as marine equipment

These standards are not over-ridden by the Directive; those which have been harmonized by CENELEC may be applied to products within their scope and are regarded as adequate to demonstrate compliance, at least as far as their scope and coverage of the essential requirements allows.

9.4.3.4.1 The generic standards

In the early days of the EMC Directive, there were many industry sectors for which no product-specific standards had been developed. This was especially so for immunity, which was a new concept for many products. In order to fill this gap wherever possible, CENELEC gave a high priority to developing the Generic Standards. These are standards with a wide application, not related to any particular product or product family, and are intended to represent the essential requirements of the Directive. They are divided into two groups, one for immunity and one for emissions, each of which has separate parts for different environment classes (Table 9.4.2).

Table 9.4.2 The generic standards

Environment	Residential, commercial and light industry	Industrial
Emissions	EN 61000-6-3	EN 61000-6-4
Immunity	EN 61000-6-1	EN 61000-6-2

Where a relevant product-specific standard does exist, this takes precedence over the generic standard. It is quite common, though, for a particular product to need a variety of standards. For instance, a fire alarm may be covered by one product standard for mains harmonic emissions, another for immunity, and the generic standard for emissions. All these standards must be satisfied before compliance with the Directive can be claimed. Although a comprehensive range of product standards has been developed, other mixed combinations still occur and there will always be unusual products that "fall through the cracks".

9.4.3.4.1.1 Environment classes

The distinction between environmental classes is based on the electromagnetic conditions that obtain in general throughout the specified environments. The inclusion of the "light industrial" environment (workshops, laboratories and service centres) in class 1 has been the subject of some controversy, but studies have shown that there is no significant difference between the electromagnetic conditions at residential, commercial and light industrial locations. Equipment for the class 2 "industrial" environment is considered to be connected to a dedicated transformer or special power source, in contrast to the class 1 environment which is considered to be supplied from the public mains network.

9.4.3.4.2 Performance criteria

A particular problem with immunity is that the equipment under test may exhibit a wide variety of responses to the test stimulus. This can range from a complete lack of response, through a degradation in the accuracy of measured variables to total corruption of its operation. The same problem does not exist for emissions, where comparison with a defined test limit is possible. To account for this variety, the generic immunity standards include three generalized performance criteria for the purpose of evaluating test results. In the test report, you must include a functional description and a specific definition of performance criteria based on these, during or as a consequence of the EMC testing. Most

noteworthy is that the criteria are grounded on what performance the user may reasonably expect *or is told to expect*. In other words, if you specify a given performance loss during application of the immunity test and write this into the user documentation, then provided the equipment does not actually become unsafe as a result of the test, you have met the requirements of the generic standards.

9.4.3.4.3 Basic and product standards

The tests defined in the generic standards are based only on internationally approved, already existing standards. For each electromagnetic phenomenon the test procedure given by such a standard is referenced, and a single test level or limit is laid down. No new tests are defined in the body of any generic standard.

Those standards which are referenced in the generic standards, for example the various parts of EN 61000 along with some of the CISPR standards, are known as "basic" standards. This means that such standards are entirely devoted to aspects of EMC that will prove to be of general interest and use to all committees developing other standards – for instance, product-specific standards. Generally, a product-specific standard will take a form similar to the generic standard, with similar limits, but will be more specific as regards operational modes and configurations, and about performance criteria that are considered acceptable. It will refer to the basic standards for the test methods wherever possible.

9.4.3.4.4 Harmonization

The discussion in this chapter has repeatedly referred to standards as being "harmonized". This has a special meaning under the EC's New Approach, which covers the EMC and R&TTE Directives as well as many others. The meaning and structure is comprehensively outlined in the Commission's "Blue Guide", which covers the implementation of New Approach Directives. In summary:

> *Harmonised standards are European standards, which are adopted by European standards organisations, prepared in accordance with the General Guidelines agreed between the Commission and the European standards organisations, and follow a mandate issued by the Commission after consultation with the Member States. European standards are technical specifications adopted by European standards organisations for repeated or continuous application, with which compliance is not compulsory. According to the internal rules of these organisations, European standards must be transposed at national level. This transposition means that the European standards in question must be made available as national standards in an identical way, and that all conflicting national standards must be withdrawn in a given period.*
>
> *Harmonised standards in the meaning of the New Approach are deemed to exist when the European standards organisations formally present to the Commission the European standards elaborated or identified in conformity with the mandate. Harmonised standards are not a specific category amongst European standards. The terminology used in New Approach directives is a legal qualification of technical specifications existing as European standards, but to which a special meaning has been given by these directives.*

From this, you can see that international (IEC) standards by themselves cannot be harmonized; only their European editions can be. Harmonized standards provide a *presumption of conformity* with the essential requirements, once their reference has been published in the *Official Journal of the EU*. The objective of publishing the reference is to set the earliest date for the presumption of conformity to take effect, and the entry in the OJ includes this information. The OJ listing is updated at occasional intervals, and the most recent can be found via links on the EC's Europa website.

In fact, because of this system, publication in the OJ assumes a considerable importance especially when difficulties have arisen with the application of a particular standard, and it has been amended or revised to take these into account. You cannot legitimately use a new version of a standard until it has been harmonized by an OJ listing. In 2005 and 2006 a complex situation arose with the application of EN 55022 which was exacerbated by the Commission's failure to publish necessary updates in any reasonable timescale, and indeed to introduce errors when it did. It's probably fair to say that the resulting quagmire has severely tested many participants' faith in the compliance system as it now exists.

9.4.3.4.5 Presumption of conformity

The meaning of this term is important: it says that if you correctly apply all relevant harmonized standards, it is to be *presumed* that you comply with the Directive's essential requirements; but it doesn't *guarantee* that you do. There may be deficiencies in a standard even if you apply it correctly, which mean that it is possible for a product still to cause or suffer from interference, despite complying with its requirements. Some standards have known loopholes which still exist, although their existence has been well documented for many years. Since you are required to certify compliance with the

essential requirements, there will always be some element of risk which is unavoidable. The best that can be said is that an enforcement authority is likely to have a harder time proving non-compliance if you can show correct application of the standards.

The presumption of conformity is limited to the scope of the standard, and to the relevant essential requirements that are covered by the standard, and to *compliance with* (not merely application of) the standard (see Section 9.4.3.1.2). Not many standards explicitly state which essential requirements they cover, although it is slowly becoming more common as standards are updated. Just because it is listed in the OJEU, with a title that applies to your apparatus, this does not mean it covers exactly what you want it to, wholly and completely. So you still need to carry out (and document) some analysis yourself, in order to tell whether you are using an adequate standard or group of standards, or whether you need to expand the EMC assessment to cover those aspects that the chosen standard doesn't. A test plan is still needed to interpret even a comprehensive standard in the light of your particular product, so you would have had to do some such analysis anyway.

There is a further question as to what is meant by "correct application" of a harmonized standard. As discussed in Section 9.4.2.3.5, the Directive states that this means doing everything by the book, that is, in a manner fully compliant with the methods as given in the standard. You may be prepared to do this, or to get a test house to do it for you; the cost may be justified. But if it isn't, and you use only "precompliance" methods of test, then your EMC assessment should include an analysis of the limitations of such methods and how you have dealt with them, since you will then only have "partially" applied a harmonized standard.

In respect of the protection requirements, the EMC Directive, at recital 13, indicates that harmonized standards are understood to "reflect the generally acknowledged state of the art as regards EMC matters in the European Union". Where harmonized standards are not employed in full, the manufacturer should have regard to the state of the art in terms of the services to be protected, and the electromagnetic disturbances to which his apparatus may be subjected.

9.4.4 Action for compliance for a product manufacturer

The second edition EMC Directive is all about risk. The dimensions of the risk are either that you will take insufficient measures to prove compliance with the Directive, in terms of identifying, performing the EMC assessment and testing to the correct harmonized standards, and meeting their requirements; or, you will spend so much time and effort in doing this that your product will not make it to market on time or to budget.

With this in mind, the steps to take for a new product to achieve compliance with the EMC Directive and bear the CE mark can be detailed as follows:

A. Self certification

1. From the marketing specification, determine what type of product it will be and what environment it will be sold for use within, and hence which if any product-specific or product-family standards published in the OJEU apply to it. If your company only ever makes or imports products for one particular application then you will be able to use the same product standard(s) for all products.
2. If no product standards apply, or they don't cover all the essential requirements (generally: LF disturbance emissions and immunity, transient immunity, and RF emissions and immunity), check the generic standards to see if the tests specified in them are applicable. The environmental classification will depend on the intended power supply connection.
3. If you cannot apply any harmonized standards, or can only partly apply them, then you will need to do a detailed EMC assessment of the equipment to cover the essential requirements, which may or may not involve other test methods.
4. Having determined what standards you will use, decide on the test levels and to what ports of the equipment (enclosure, power leads, signal/control leads) they will apply. In some cases there will be no choice, but in others the test applicability will depend on factors such as length of cable, EUT configuration and class of environment.
5. From this information you will be able to draw up a test plan, which specifies in detail the version and configuration of the EUT and any associated apparatus, the tests that will be applied to it and the pass/fail criteria. You can discuss this with your selected test house or your in-house test facility staff, and it will form the basis for your agreement with them and also for the technical documentation required by the provisions of the Directive.
6. Knowing the requirements of the test plan will enable you to some degree to incorporate cost-effective EMC measures into the product design, since the test limits and the points to which they will be applied will have been specified.
7. As the design progresses through prototype and pre-production stages you can make

pre-compliance confidence tests to check the performance of the product and also the validity of the test plan. It is normal for both design and test plan to undergo iterative modifications during this stage.

8. Once the design has been finalized and shortly before the product launch you can then perform, or get a test house to perform, full compliance tests against harmonized standards or against your own assessment as appropriate, on one or more production samples, the results of which are recorded in the technical documentation. Provided that confidence tests were satisfactory this should be no more than a formality.
9. You are then at liberty to mark the product, and/or its packaging or documentation with the CE mark (if there is no other Directive to satisfy) and your empowered signatory can sign the Declaration of Conformity, to be kept for 10 years. If other New Approach Directives apply, a parallel process needs to be followed for each. The product can be placed on the market.
10. Once the product is in series production you must take steps to ensure that any changes to its build state allow it to continue to comply with the protection requirements.

B. Involvement of a Notified Body

This approach is in addition to the steps shown above.

11. In theory, once you have produced the full package of documentation at step 8 above, you select a Notified Body capable of assessing your type of product based on their published scope of designation, hand over the documentation to them, and expect a statement confirming compliance within a few weeks.
12. In practice, you will have chosen an NB at a much earlier stage, quite possibly from step 1 above, and confirmed with them what standards and/or what test plan content would be acceptable to them in order for them to provide a positive statement; also you will have agreed what the scope of their involvement will be, i.e. what aspects of the essential requirements they will be asked to review.

This page is intentionally left blank

Section Ten

Design techniques: mechanics

This page is intentionally left blank

Chapter 10.1

10.1

Introduction to mechanical design

Peter Childs

The aim of this book is to present an overview of the design process and design methodology, and to introduce the technology and selection of a number of specific machine elements that are fundamental to a wide range of engineering. This chapter introduces the design process from an inventor's perspective to a more formal model called 'total design'.

Learning objectives: at the end of this chapter you should

- be able to recognize the iterative steps associated with the design process;
- be aware of the process involved in developing a product design specification from a market brief;
- be able to develop a morphological design chart to aid the development of conceptual solutions;
- be able to participate in brain-storming activities;
- recognize the scope of available technologies to the designer.

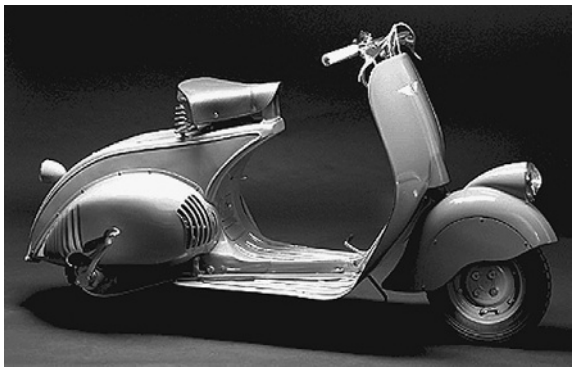


Fig. 10.1.1 Piaggio's Vespa launched in 1946 (Piaggio Museo). The Vespa was an early example of monocoque construction where the skin and frame are combined as a single construction to provide appropriate rigidity and mounting for the vehicle's components and riders.

10.1.1 Introduction

The term 'design' is popularly used to refer to an object's aesthetic appearance with specific reference to its form or outward appearance, as well as its function. For example, we often refer to designer clothes, design icons and beautiful cars, examples of which are given in Figures 10.1.1 and 10.1.2. In these examples it is both visual impact, appealing to our visual perception, and the concept of function, that the product will fulfil a range of requirements, which are important in defining so-called 'good design'.

The word 'design' comes from the Latin 'designare', which means to designate or mark out. Design can be

Fig. 10.1.2 The Audi TT, originally launched in 1998, and a contender for the most attractive sports car of the 20th century. (Courtesy of Audi.)

Mechanical Design; ISBN: 9780750657716
Copyright © 2004 Elsevier Ltd. All rights of reproduction, in any form, reserved.

taken to mean all the processes of conception, invention, visualization, calculation, refinement and specification of details that determine the form of a product. Design generally begins with either a need or requirement or, alternatively, an idea. It ends with a set of drawings or computer representations and other information that enables a product to be manufactured and utilized.

In this book, design is defined as 'the total activity necessary to provide a product or process to meet a market need'. This definition comes from the SEED (Sharing Experience in Engineering Design, now DESIG the Design Education Special Interest Group of the Design Society) model, see Pugh (1990).

10.1.2 The design process

Probably from your own experience you will know that design can consist of examining a design need and working on the problem by means of sketches, models, brain-storming, calculations as necessary, development of styling as appropriate, making sure the product fits together and can be manufactured, and calculation of the costs. The process of design can be represented schematically to levels of increasing formality and complexity. Figure 10.1.3 represents the traditional approach associated with lone inventors. This model comprises the generation of the 'bright idea', drawings and calculations giving form or shape to the idea, judgement of the design and re-evaluation if necessary, resulting in the generation of the end product. The process of evaluation and reworking an idea is common in design and is represented in the model by the iteration arrow taking the design activity back a step so that the design can be improved. Figure 10.1.4 illustrates the possible results from this process for a helmet providing peripheral and reverse vision.

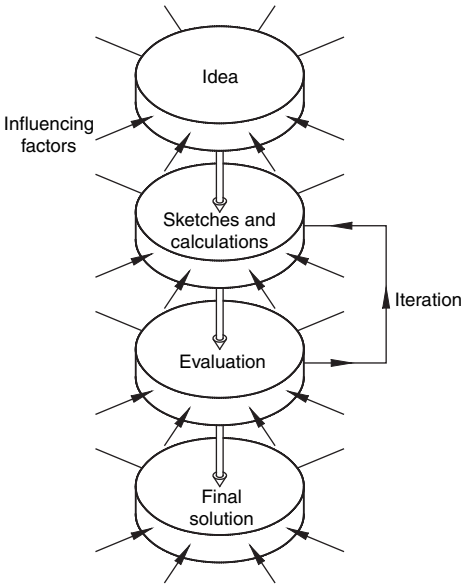


Fig. 10.1.3 The traditional and familiar'inventor's' approach to design.

Figure 10.1.5 shows a more formal description of the design process that might be associated with engineers operating within a formal company management structure. The various terms used in Figure 10.1.5 are described in the following text.

- *Recognition of need*. Often design begins when an individual or company recognizes a need, or identifies a potential market, for a product, device or process. Alternatively 'need' can be defined as when a company decides to re-engineer one of its existing products (for example, producing a new car model). The statement of need is sometimes referred to as the brief or market brief.
- *Definition of problem*. This involves all the specifications of the product or process to be designed. For example, this could include inputs and outputs, characteristics, dimensions and limitations on quantities.
- *Synthesis*. This is the process of combining the ideas developed into a form or concept, which offers a potential solution to the design requirement. The term synthesis may be familiar from its use in chemistry where it is used to describe the process of producing a compound by a series of reactions of other substances.
- *Analysis*. This involves the application of engineering science: subjects explored extensively in traditional engineering courses, such as statics and dynamics, mechanics of materials, fluid flow and heat transfer. These engineering 'tools' and techniques can be used to examine the design to give quantitative information such as whether it is strong enough or will operate at an acceptable temperature. Analysis and synthesis invariably go together. Synthesis means putting something together and analysis means resolving something into its constituent parts or taking it to pieces. Designers have to synthesize something before it can be analysed. The famous chicken and the egg scenario! When a product is analysed some kind of deficiency or inadequacy may be identified requiring the synthesis of a new solution prior to reanalysis and repetition of the process until an adequate solution is obtained.
- *Optimization*. This is the process of repetitively refining a set of often-conflicting criteria to achieve the best compromise.
- *Evaluation*. This is the process of identifying whether the design satisfies the original requirements. It may involve assessment of the analysis, prototype testing and market research.

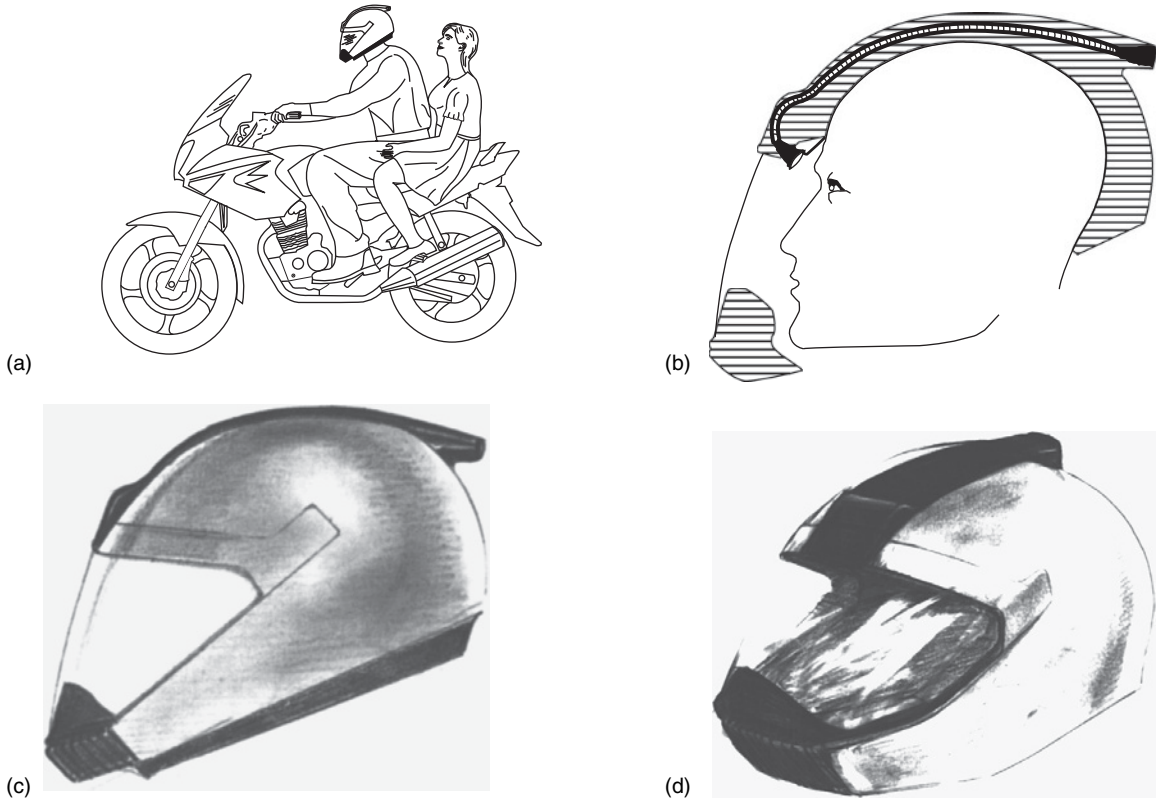


Fig. 10.1.4 Panoramic helmet by Alberto Meda and Denis Santachiara (Manzini, 1989). (a) The need: to be able to view behind you. (b) The idea: an optical link using fibre optics and lenses. (c) & (d) Practical sketches showing the concept.

Although Figures 10.1.3 and 10.1.5 at first sight show design occurring in a sequential fashion, with one task following another, the design process may actually occur in a step forward, step back fashion. For instance, you may propose a solution to the design need and then perform some calculations or judgements, which indicate that the proposal is inappropriate. A new solution will need to be put forward and further assessments made. This is known as the iterative process of design and forms an essential part of refining and improving the product proposal.

Note that the flow charts shown in Figures 10.1.3 and 10.1.5 do not represent a method of design, but rather a description of what actually occurs within the process of design. The method of design used is often unique to the engineer or design team. Design methodology is not an exact science and there are indeed no guaranteed methods of design. Some designers work in a progressive fashion, others work on several aspects simultaneously.

10.1.2.1 Case study

The process identified in Figure 10.1.5 can be illustrated by example. Following some initial market assessments the board of a plant machinery company has decided to proceed with the design of a new product for transporting pallets around factories. The board has in mind a forklift truck, but do not wish to constrain the design team to this concept alone. The process of the design can be viewed in terms of the labels used in Figure 10.1.5 as listed below.

- *Recognition of need (or market brief)*. The company has identified a potential market for a new pallet-moving device.
- *Definition of problem*. A full specification of the product desired by the company should be written. This allows the design team to identify whether their design proposals meet the original request. Here a long list of information needs to be developed and clarified before design can proceed. For example, for the pallet-moving device being explored here this would likely include aspects for consideration such as:
  - What sizes of pallet are to be moved?
  - What is the maximum mass on the pallet?
  - What is the maximum size of the load on the pallet?
  - What range of materials are to be moved and are they packaged?
  - What maximum height must the pallet be lifted?

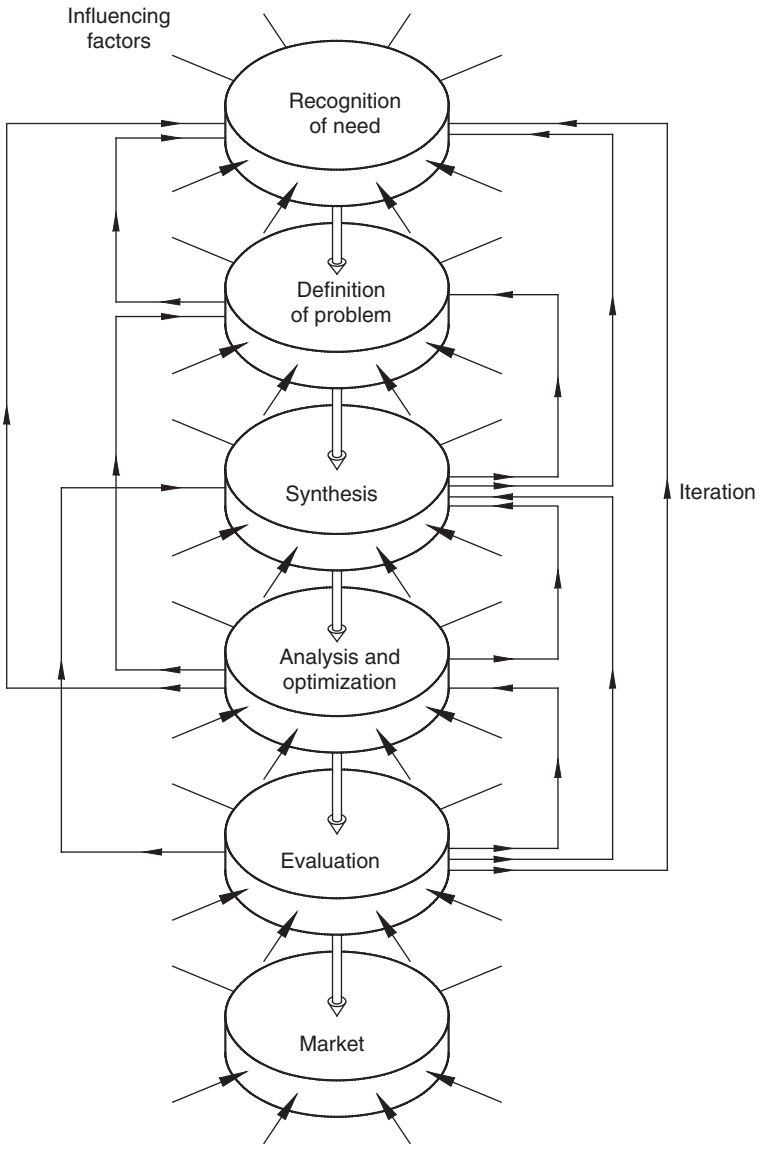


Fig. 10.1.5 The design process illustrating some of the iterative steps associated with the process.

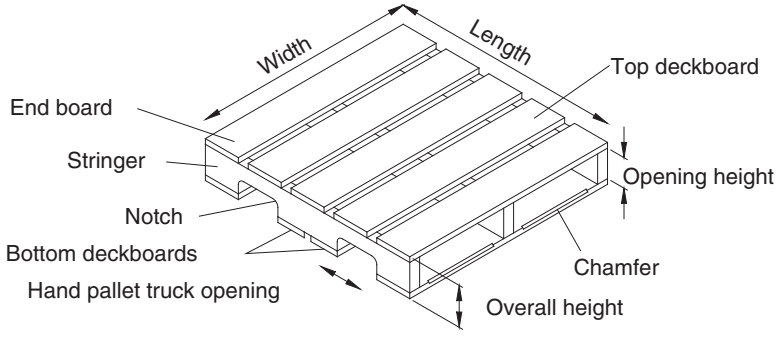


Fig. 10.1.6 Pallet dimensions and terminology (see BS ISO 509, 99/712554 DC and 99/712555 DC).

- What terrain must the pallet-moving device operate on?
- What range is required for the pallet-moving device?
- Is a particular energy source/fuel to be used?
- What lifetime is required?
- Are there manufacturing constraints to be considered?
- What is the target sales price?
- How many units can the market sustain?
- Is the device to be automatic or manned?
- What legal constraints need to be considered?

This list is not exhaustive and would require further consideration. The next step is to quantify each of the criteria. For instance the specification may yield that standard size pallets, see Figure 10.1.6, are involved, the maximum load to be moved is 1000 kg, the maximum volume of load is 2 m^3, the reach must be up to 3 m, use on factory floor and asphalt surfaces, the pallet-moving device must be capable of moving a single pallet 100 m and must be able to repeat this task at least 300 times before refuelling if necessary, electricity, gas or diesel fuel, 7-year lifetime, production in an East European country, target selling price 9000 Euros, 12,000 units per year, manned use, design to ISO (International Organisation for Standardisation) and target country national standards (e.g. see BS ISO 509, BS ISO 6780, BS EN ISO 445, BS EN 1726-1, BS EN 13545, 99/705213 DC, ISO 18334, 99/712554 DC, BS 3726, BS 5639-1 and BS ISO 2330).

- *Synthesis*. This is often identified as the formative and creative stage of design. Some initial ideas must be proposed or generated in order for them to be assessed and improved. Concepts can be generated by imagination, experience or by the use of design techniques, such as morphological charts. Some evaluation should be made at this stage to reduce the number of concepts requiring further work. Various techniques are available for this including merit and adequacy assessments.
- *Analysis*. Once a concept has been proposed it can then be analysed to determine whether constituent components can meet the demands placed on them in terms of performance, manufacture, cost and any other specified criteria. Alternatively, analysis techniques can be used to determine what size components need to be to meet the required functions.
- *Optimization*. Inevitably there are conflicts between requirements. In the case of the fork– lift truck, size, manoeuvrability, cost, aesthetic appeal, ease of use, stability and speed are not necessarily all in accordance with each other. Cost minimization may call for compromises on material usage and manufacturing methods. These considerations form part of the optimization of the product producing the best or most acceptable compromise between the desired criteria.
- *Evaluation*. Once a concept has been proposed and selected and the details of component sizes, materials, manufacture, costs and performance worked out, it is then necessary to evaluate it. Does the proposed design fulfil the specification? If it appears to, then further evaluation by potential customers and use of prototype demonstrators may be appropriate to confirm the functionality of the design, judge customer reaction and provide information of whether any aspects of the design need to be reworked or refined.

10.1.3 Total design

The process of design has been the focus of research for many years and a number of design models and methodologies are available. Design methodology is a framework

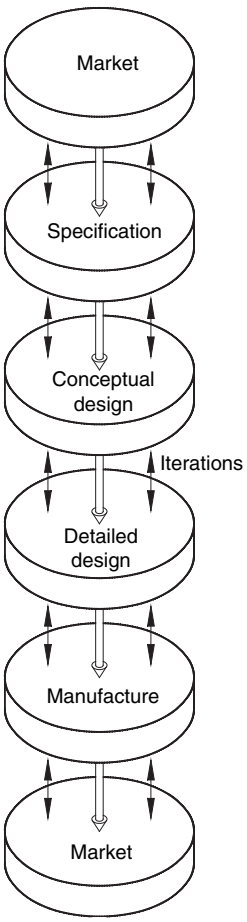


Fig. 10.1.7 The total design core (after Pugh, 1990).

within which the designer can practise with thoroughness. One such approach called 'total design' has been proposed by the SEED programme (1985) and Pugh (1990) and is illustrated schematically in Figure 10.1.7. This shows the core activities of design: marketing, specification, conceptual design, detailed design and marketing/selling. As in Figures 10.1.3 and 10.1.5 the iterative nature of design is accounted for where work on a design results in the need to go back and redo previous work in order to produce a better overall design to meet the requirements. Indeed it is sometimes necessary to go back several levels. An example might be the discovery at manufacture that an item cannot be made as envisaged and a new concept altogether is required. Ideally such a discovery should not occur, as every other level of the design process illustrated in Figure 10.1.7 should be considered at each stage. Each of the design activities illustrated in Figure 10.1.7 is described in more detail in the following text. As it is the same process being described these descriptions are similar to those described for Figure 10.1.5.

- *Market*. The market phase refers to the assessment of sales opportunities or perceived need to update an existing product resulting in a statement sometimes called the market brief, design brief, brief or statement of need.
- *Specification*. Specification involves the formal statement of the required functions, features and performance of the product or process to be designed. Recommended practice from the outset of design work is to produce a product design specification that should be formulated from the statement of need. The product design specification is the formal specification of the product to be designed. It acts as the control for the total design activity because it sets the boundaries for the subsequent design. Further details of the product design specification are described in Section 10.1.4.
- *Conceptual design*. The early stages of design where the major decisions are to be made is sometimes called 'conceptual design'. During this phase a rough idea is developed as to how a product will function and what it will look like. The process of conceptual design can also be described as the definition of the product's morphology, how it is made up and its layout. Some methods such as brain-storming, morphological analysis and function trees used to aid the generation of concepts are described in Section 10.1.5.
- *Detailed design*. The detailed design phase consists of the determination of the specific shape and size of individual components, what materials should be used, how they fit together and the method of

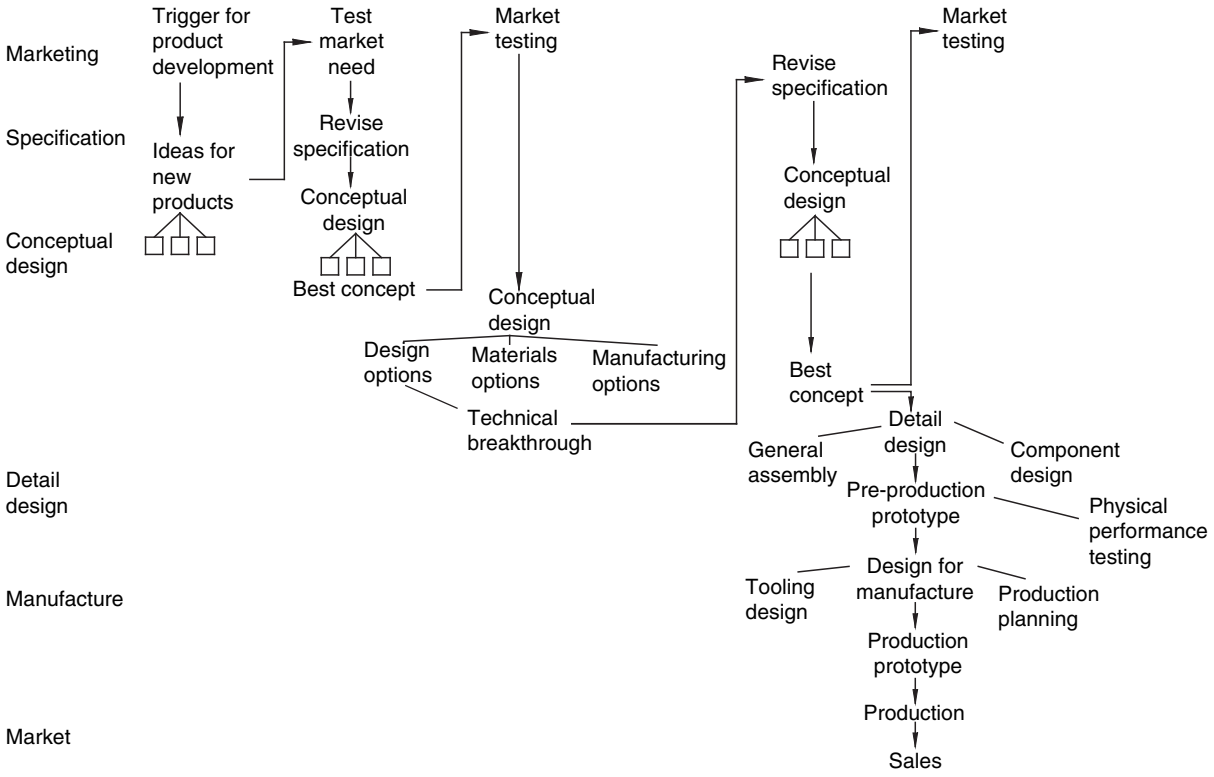


Fig. 10.1.8 Design activities at different stages in product development (adapted from Baxter, 1995).

manufacture. Detailed design makes use of the many skills acquired and developed by engineers in the areas of analysis. It is the detailed design phase that can take up the bulk of the time spent on a design. However, as implied earlier it is wise to only spend time on details once a sensible concept has been selected.

- *Manufacturing*. The manufacture phase, although identified as distinct within the structure, is typical of other phases in that it influences all the others. The design of any item must be such that it is feasible to manufacture it! The materials selected must be compatible with the manufacturing facilities and skills available and at acceptable costs to match marketing requirements. Manufacturing is so important that design strategies to reinforce its importance have been developed, such as design for assembly DFA and design for manufacture DFM (see for instance Boothroyd (1997)). More recently, the concept of concurrent engineering has been popular which is a systematic approach that encourages the developer from the outset to consider all the elements of a product life cycle or process from concept to disposal, including quality control, scheduling and user requirements.
- *Marketing/sales*. The last phase, selling, is of course essential and should match the expectations of the initial market phase. This phase too has an impact on other phases within the design core. Information such as customer reaction to the product, any component failures or wear, damage during packaging and transportation should be fed back to influence the design of product revisions and future designs.

The double arrows shown in Figure 10.1.7 represent the flow of information and control from one activity to another, as well as the iterative nature of the process. For instance detailed design activity may indicate that an aspect of the conceptual design is not feasible and must be reconsidered. Alternatively conceptual work may

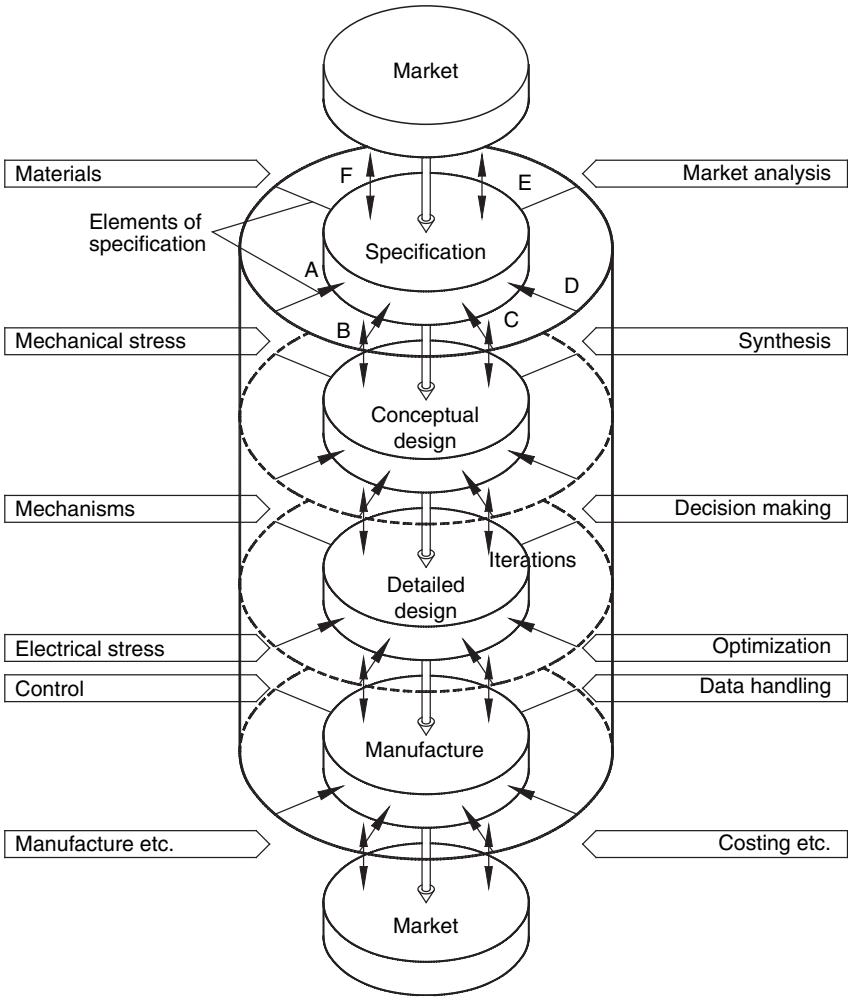


Fig. 10.1.9 The total design process (after Pugh, 1990).

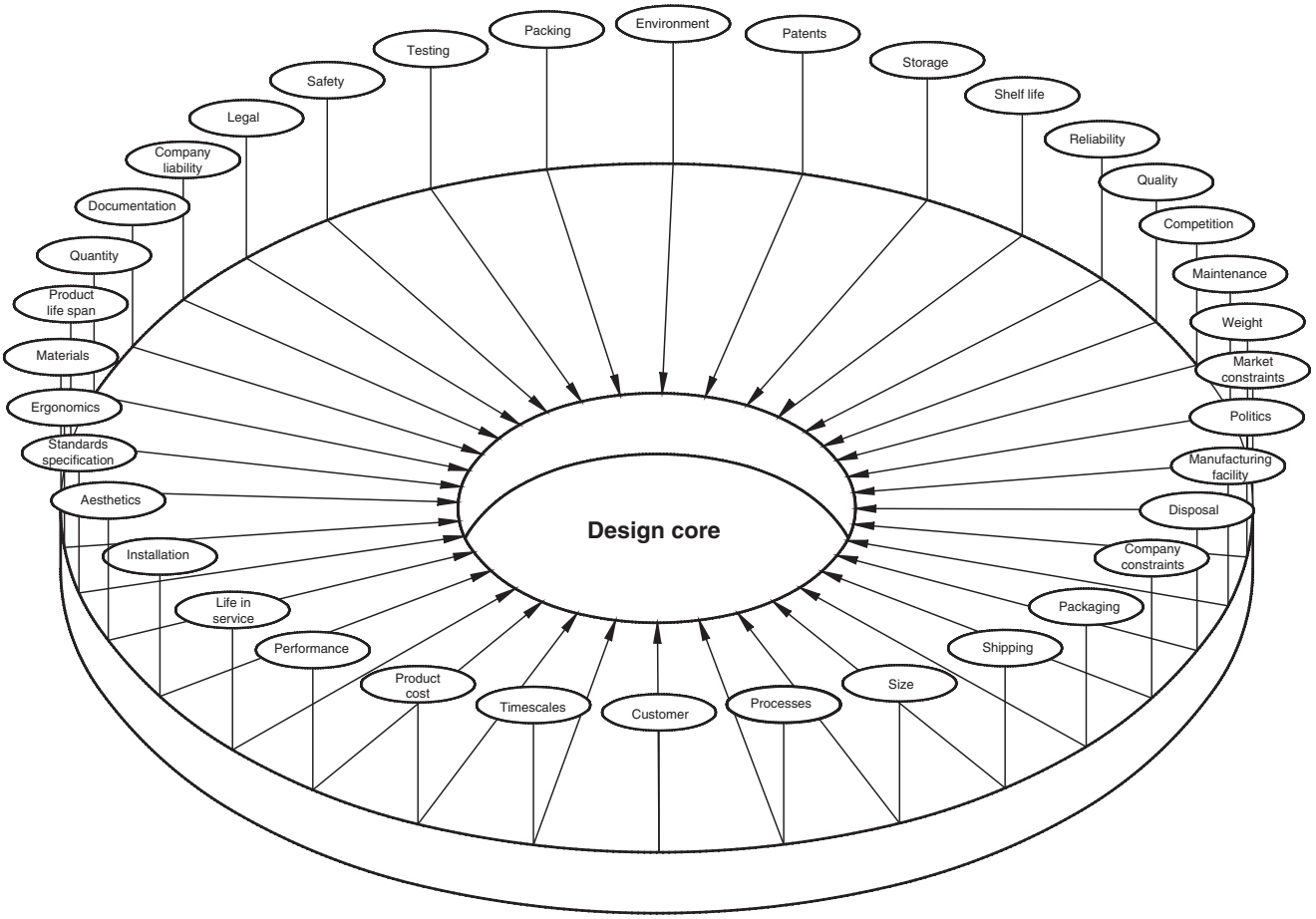


Fig. 10.1.10 Information content of atypical product design specification (after Pugh, 1990).

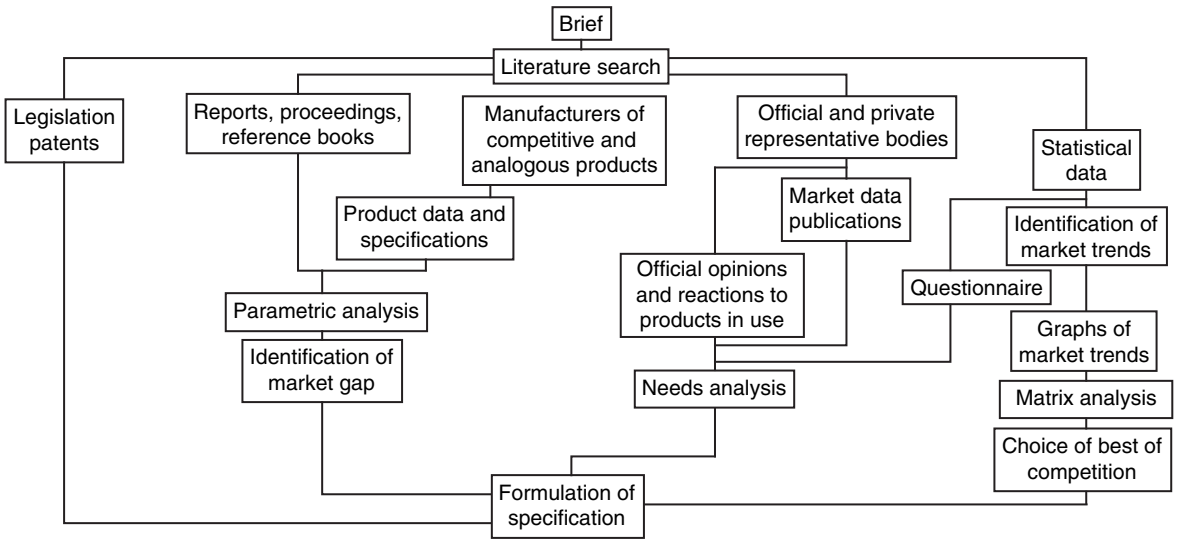


Fig. 10.1.11 Information required for the production of the product design specification (after Pugh, 1990).

yield features, which have potential for additional marketing opportunities. In other words the activity on one level can and does interact dynamically with activities on other levels. Figure 10.1.8 illustrates the possible flow of this process during the development of a product.

Almost any product such as a vacuum cleaner, kettle, automobile or cordless hand-tool requires input from people of many disciplines, including engineering, legal and marketing and this requires considerable co-ordination. In industrial terms, the integration comes about as a result of the partial design inputs from each discipline. In Figure 10.1.9 additional activities, such as market analysis, stressing and optimization, have been added to the design core as inputs. The effective and efficient design of any product invariably requires the use of different techniques and skills.The disciplines indicated are the designer's tool-kit and indicate the multi-disciplinary nature of design. The forklift truck example mentioned in Section 10.1.2.1 will require engine management and control systems, as well as the design of mechanical components. Although this text concentrates on mechanical design, this is just one, albeit an important, interesting and necessary aspect of the holistic or total design activity.

A number of circumferential inputs have been shown as arrows in Figures 10.1.3, 10.1.5, 10.1.7 and 10.1.9. These represent elements of the specification listed in order of importance for each phase of design. The

Table 10.1.1 Format for the documentation of the product design specification (after Pugh, 1990)

Date:	**Product: .....**			**Issue: .....**
Parameters	**Competition on best**	**Current model (ours)**	**This design (intent)**	**World class (target)**
Performance description				
Safety description				

Table 10.1.2 Example table for developing a product design specification for one particular aspect. All relevant aspects identified in Figure 10.1.10 should be considered in a similar fashion

Aspect	**Objective**	**Criteria**	**Test conditions**
Performance	Low wear of moving parts	Switch will operate 10,000 times without fault Electric motor will operate for 250 h without fault Cutting blade will cut effectively for 100 h without attention	Test to ISO Standard Continuous-run laboratory test until failure

Table 10.1.3 Performance and weight aspects of a product design specification for a lawn mower

Aspect	**Objective**	**Criteria**	**Test conditions**
Performance	Low wear of moving parts	Switch will operate 10,000 times without fault Electric motor will operate for 250 h without fault Cutting blade will cut effectively for 1000 h without attention	Test to British Standards Continuous run laboratory test
	Low probability of damage	Body of lawnmower can withstand a side impact without cracking as a result of being dropped from a height of 1 m	Practical test
	Low level of vibration	Static out of balance of cutting blades not to exceed 2 g	To British Standards
Weight	Low weight	Total weight of machine not to exceed 50 kg	Weigh it!

Adapted from Walker et al. (1991)

priority order of these specifications may alter for different phases of the design activity. The exact number will depend on the actual case under consideration.

Industry is usually concerned with total design. Total design is the systematic activity necessary from the identification of a market need to the commercialization of the product to satisfy the market need. Total design can be regarded as having a central core of activities consisting of the market potential, product specification, conceptual design, detailed design, manufacture and marketing.

10.1.4 Product design specification

Specification has different meanings depending on the context and the individuals concerned. Here specification refers to an activity that assists in providing a framework within which design can take place. Work on a design may commence with an identification of a need, which is sometimes summarized in the form of a statement commonly known as the 'design brief' or 'design intent'. A design brief may provide some direction to initial design activity, but rarely sets sufficient bounds on it to allow a design team to know when their efforts are adequate, or not, to fulfil the requirements. In order to overcome this shortcoming and compel the individuals concerned to explore what is actually required of the product or process concerned a 'product design specification' (PDS) should be developed.

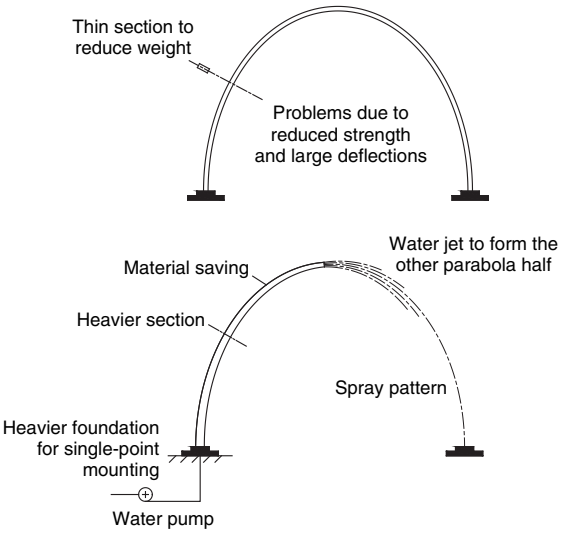
In developing a product design specification, bounds should be set on a wide range of relevant parameters for the product or process concerned.

The spider chart shown in Figure 10.1.10 provides a helpful checklist of aspects for consideration. A useful rule associated with specification, which will be stated frequently, is the mantra:

quantify wherever possible.

Having identified a general need or requirement and documented this in the form of a brief, the specification of the product to be developed should be generated. This is called the 'product design specification' or PDS. The PDS acts as the controlling mechanism, mantle or envelope for the total design activity. Whatever you are concerned with, the PDS gives you the terms of reference. The starting point for design activity is market research, analysis of the competition, literature and patent searching. Once this has been undertaken the PDS can be developed. Figure 10.1.11 shows the areas of information and research required to produce the PDS and Table 10.1.1 shows the information content of a typical PDS.

In a new design it is preferable to consider each aspect identified in Figure 10.1.10, however for the re-engineering of a current product, engineering expediency dictates that some aspects will be passed over. The PDS thus consists of a document covering a wide range of considerations. A possible format for the PDS is shown in Table 10.1.2 and explored in Example 10.1, although other types of pro-forma exist and are widely



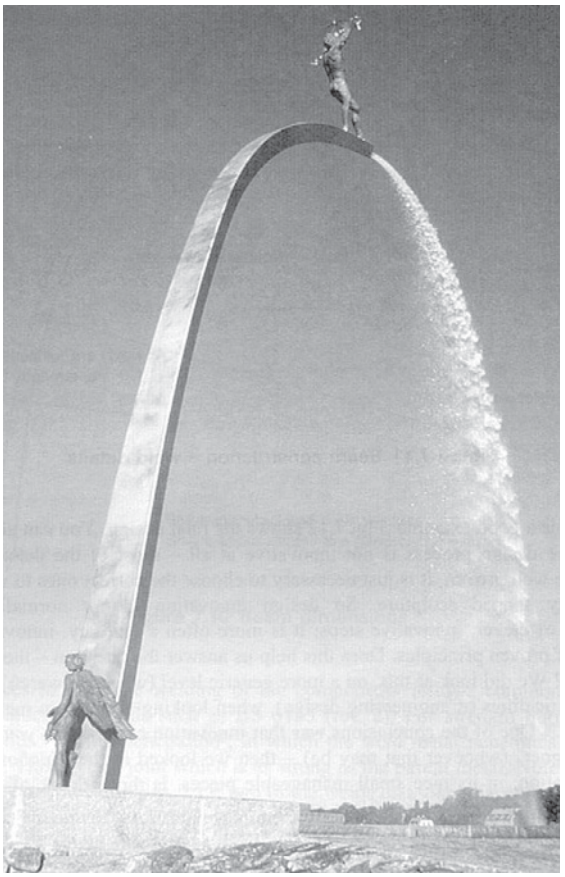


Fig. 10.1.12 Conceptual and final solutions (Matthews, 1998).

used. Table 10.1.3 shows just two of the 30+ items listed in Figure 10.1.10, but gives a sensible format for documentation of the PDS. The PDS is dynamic. If during the design of a product there is a good reason for changing the basic PDS this can be done. It should be noted, however, that the PDS may actually form part of contractual obligations and the legal implications must be addressed.

Undertaking a PDS will typically be an iterative process where initial parameters are identified and limits set, but more information may be required in order to confirm these. The process of developing a PDS will require input from people with a variety of skills and knowledge. It will very rarely be simply one day's work and may extend to many man months. Recall that conceptual design, let alone detailed design, cannot sensibly proceed towards any firm status until the specification has been completed.

Example 10.1.1: The managing director of a company specializing in gardening equipment has called a meeting to explain a marketing opportunity for a power-assisted grass-cutting device. The director explains that the key objectives of this device are safety and reliability. The director has asked you to develop part of the product design specification for this device considering low wear of moving parts, low probability of damage, low level of vibration and low weight.

Solution: There are no right answers to this activity. The suggestions given in Table 10.1.3 are a possibility. Note the inclusion of numbers. Think about the values given. Why 250 h operation for the motor?

Table 10.1.4 A general morphological analysis table

Subsystem	Means			
1	Method 1 of fulfilling subsystem 1	Method 2 of fulfilling subsystem 1	Method 3 of fulfilling subsystem 1	Method *n* of fulfilling subsystem 1
2	Method 1 of fulfilling subsystem 2	Method 2 of fulfilling subsystem 2	Method 3 of fulfilling subsystem 2	Method *n* of fulfilling subsystem 2
3	Method 1 of fulfilling subsystem 3	Method 2 of fulfilling subsystem 3	Method 3 of fulfilling subsystem 3	Method *n* of fulfilling subsystem 3
4	Method 1 of fulfilling subsystem 4	Method 2 of fulfilling subsystem 4	Method 3 of fulfilling subsystem 4	Method *n* of fulfilling subsystem 4
5	Method 1 of fulfilling subsystem 5	Method 2 of fulfilling subsystem 5	Method 3 of fulfilling subsystem 5	Method *n* of fulfilling subsystem 5

Table 10.1.5 Morphological chart for a pallet moving device

Feature	Means				
Support	Track	Wheels	Air cushion	Slides	Pedipulators
Propulsion	Driven wheels	Air thrust	Moving cable	Linear induction	
Power	Electric	Diesel	Petrol	Bottled gas	Steam
Transmission	Belts	Chains	Gears and shafts	Hydraulics	Flexible cable
Steering	Turning wheels	Air thrust	Rails	Magnetism	
Stopping	Brakes	Reverse thrust	Ratchet	Magnetism	Anchor
Lifting	Hydraulic ram	Rack and pinion	Screw	Chain or rope hoist	Linkage
Operator	Standing	Walking	Seated at front	Seated at rear	Remote control

Reproduced with modifications from Cross (2000).

10.1.5 Conceptual design

Conceptual design is the generation of solutions to meet specified requirements. Conceptual design can represent the sum of all subsystems and component parts that go on to make up the whole system. Ion and Smith (1996) describe conceptual design as an iterative process comprising of a series of generative and evaluative stages that converge to the preferred solution. At each stage of iteration, the concepts are defined in greater detail allowing more thorough evaluation. It is important to generate as many concepts and ideas as possible or economically expedient. There is a temptation to accept the first promising concept and proceed towards detailed design and the final product. This should be resisted as such results can invariably be bettered. It is worth noting that sooner or later your design will have to compete against those from other manufacturers, so the generation of developed concepts is prudent.

Some of the popular methods for assisting concept generation are described in Sections 10.1.5.1–10.1.5.3.

Fig. 10.1.13 Forklift truck.

10.1.5.1 Boundary shifting

Boundary shifting involves challenging the constraints defined in the PDS to identify whether they are necessary. For example, the PDS may define that steel should be used for a component. Boundary shifting would challenge this specification to see whether it is appropriate and, if not, other materials could be considered.

Example 10.1.2: The mayor of Stockholm has outlined funding for a sculpture, based upon an architect's sketches, to span the entrance of Stockholm harbour and has presented you as a leading engineering company with the contract to design and construct this (see Matthews (1998)).

Solution: There are often problems with designing something new. This sculpture is not an exception. Consideration of fatigue due to wind loading is critical for a structure of this sort. In this case the designer challenged the original specification for a sculpture that spanned the harbour and in place proposed a half arch structure with the remainder of the geometry completed by a stream of water, Figure 10.1.12.

Table 10.1.6 Morphological chart for a pallet moving device with choices identified

Feature	Means				
Support	Track	**Wheels**	Air cushion	Slides	Pedipulators
Propulsion	**Driven wheels**	Air thrust	Moving cable	Linear induction	
Power	Electric	Diesel	**Petrol**	Bottled gas	Steam
Transmission	Belts	Chains	**Gears and shafts**	Hydraulics	Flexible cable
Steering	**Turning wheels**	Air thrust	Rails	Magnetism	
Stopping	**Brakes**	Reverse thrust	Ratchet	Magnetism	Anchor
Lifting	Hydraulic ram	Rack and pinion	Screw	Chain or rope hoist	**Linkage**
Operator	Standing	Walking	Seated at front	**Seated at rear**	Remote control

Reproduced with modifications from Cross (2000).

10.1.5.2 Brain-storming

Brain-storming is a group-based activity for the generation of multiple ideas and concepts to fulfil a given design requirement. The emphasis in brain-storming is on quantity rather than quality. Following a brain-storming session a review or debrief should be undertaken to identify concepts or aspects of the concepts generated that have merit and should be explored further. A typical outcome is that aspects of a few of the concepts generated should be combined to form a solution to the design need that gives a better overall solution than any one of the individual concepts from the original brain-storming session.

There are a number of rules that have been found helpful for brain-storming sessions.

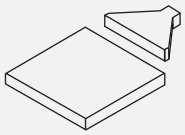
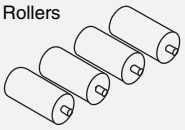
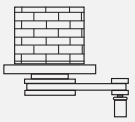
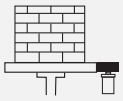
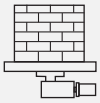
- Brain-storming should be a group activity undertaken by an interdisciplinary team. For design-based applications, this might typically involve individuals with marketing, styling, mechanical and electronic engineering, packaging, materials, software and manufacture skills.
- There should be no more than ten people present in order to preserve an effective group dynamic.
- No criticism or mockery of any idea, statement or individual is allowed in the brain- storming session. Any infringement on this should be dealt with by exclusion of the individual concerned from the session for say a period of 15 min.
- Flip charts, marker pens, pencils and paper should be provided to facilitate communication and recording of ideas.
- The design brief should be clearly stated at the start of the session.
- A set period of time should be allocated to the brain-storming session with the length of time typically being less than 2 h.

Fig. 10.1.14 The Teletruk concept by JCB. Teletruk photograph courtesy of J.C. Bamford Excavators Ltd.

- An individual should be identified to introduce the brief, enforce exclusion of any misbehaving individual from the session and to close the session and gather relevant notes and sketches arising from the exercise. The session should, however, not be chaired. Brain-storming activities tend to be most productive when participants are able to freely share their concepts and participate.

The activities outlined here involve an element of role-play and group work. It is envisaged that this will be undertaken in an appropriate environment, such as a design studio or meeting room. It is also envisaged that

Table 10.1.7 A morphological chart for some elements of a machine to stretch wrap a pallet and its load (after Wright, 1998)

Feature	Means
Pallet support drive	Ram  Rollers 
Pallet rotation	Pulley  Pneumatic roller  Geared motor 
Film carriage guidance	Guide pillar

only a selection of the questions will be attempted as advised by your course leader.

1. Working in groups of five, numbers allowing, develop as many ideas as you can in 10 min to ensure that the driver of a vehicle is able to see adequately during wet weather.
2. Working in a group of seven, numbers allow ing, develop as many ideas as you can in 20 min to move 150 kg of gravel piled at floor level in a garden centre to the boot of a car parked at a distance of up to 70 m away.

10.1.5.3 Morphological analysis

Morphological analysis is a technique that can be used to generate additional ideas for products that would not normally spring to mind. The technique involves considering the function of a generic solution to a problem and breaking it down into a number of systems or subfunctions. The next step is to generate a variety of means to fulfil each of these systems or subfunctions. The subfunctions and potential means of fulfilling each of these subfunctions are arranged in a grid. An overall solution is then formulated by selecting one means for each subfunction and the combination of these forms the overall solution.

A general morphological analysis table is illustrated in Table 10.1.4. The grid can be filled by text or by sketches depicting the potential means of fulfilling the subsystem requirement.

The use of morphological charts requires the designer to consider the function of components rather than their specific details. The pallet-moving device considered in Section 10.1.2.1 would probably comprise of support, propulsion, power, transmission, steering, stopping and lifting systems. Within each of these criteria all plausible options should be listed as shown in Table 10.1.5. The task of the designer is to look at this chart and select a justifiable option for each of the criteria producing the make-up of the pallet-moving device. In this case there would not necessarily be any surprises in the final make-up of the product and selection of wheels, driven wheels, petrol, gears and shafts, turning wheels, brakes, hydraulic ram, and seated at the front would result in the traditional forklift truck illustrated in Figure 10.1.13. However, the use of this method can assist in the production of alternatives as shown in Table 10.1.6 and Figure 10.1.14 where the option for a linkage based lifting method has been considered.

Morphological analysis can be made up using text, icons or sketches. Table 10.1.7 gives an example of a morphological analysis for a pallet-wrapping device.

10.1.5.4 Function trees

The function or objective tree method (Cross, 2000) involves decomposing the function of a product into different subfunctions. Using the function tree method, conceptual design takes place at the subsystem or component level. For example, it may not be possible to produce new concepts for the power plant of an automobile, but it may be possible to generate new concepts for the subsystems or for individual components.

For example, in the design of a lawnmower, the market brief may include key objectives such as product

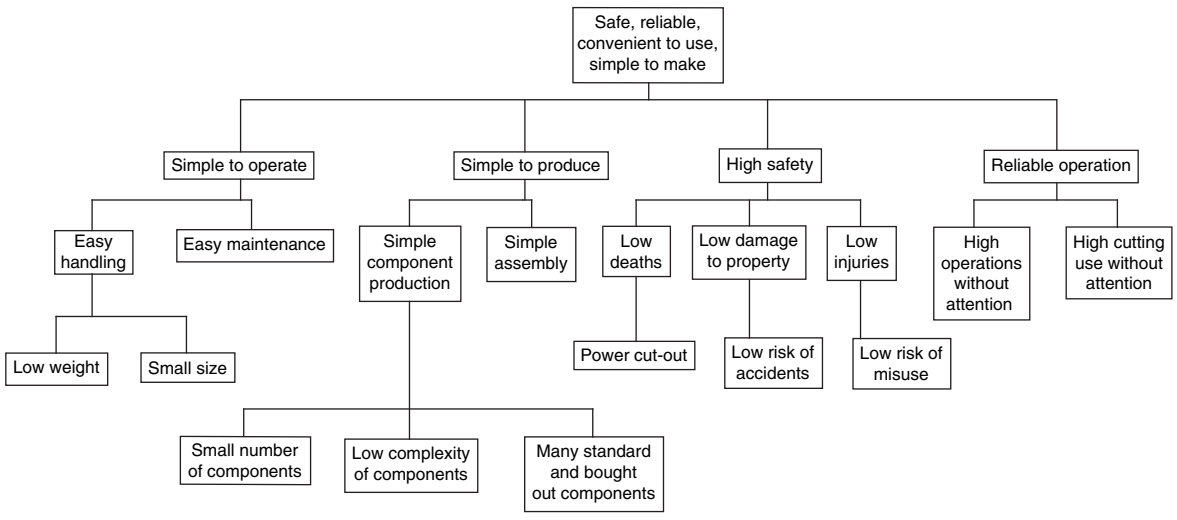


Fig. 10.1.15 An objectives tree for a lawnmower that is to be safe, reliable, convenient to use and simple to make (after Walker et al., 1991).

Table 10.1.8 An overview of the scope of machine elements

Energy conversion	Energy transmission	Energy storage	Locating	Friction reduction	Switching	Sealing	Sensors	Miscellaneous mechanisms
Turbomachinery	Gears	Flywheel	Threaded fasteners	Rolling element bearings	Clutches	Dynamic seals	Motion	Hinges, pivots
Gas turbine engines	Spur	Springs	Bolts	Deep groove	Square jaw	Mechanical face	Dimensional	Linkages
Rotodynamic pumps and compressors	Helical	Helical	Nuts and lock	Cylindrical roller	Multiple serration	Lip ring	Mass	Levers
Fans	Bevel	Leaf	Nuts	Needle roller	Sprag	Bush	Force	Tools
Propellers	Worm	Belleville	Grub screws	Taper roller	Roller	Labyrinth	Torque	Cutters
Turbines	Conformal	Roller	Studs	Angular contact	Disk	Brush	Power	Shears
Internal combustion engines	Belts	Spiral	Screws	Self aligning	Drum	Ferrofluidic	Pressure	Drills
Rotary	Flat	Garter	Washers	Thrust	Cone	Rim seals	Pitot tubes	Formers
Reciprocating	Vee	Fluid accumulator	Nails	Recirculating ball	Magnetic	O rings	Static tappings	Grips
Boilers and combustors	Wedge	Gas spring	Pins	Sliding bearings	Synchromesh	Packings	Manometers	Guides
Electric motors	Round	Reservoir	Cylindrical	Plain rubbing	Ratchet and pawl	Piston rings	Piezoelectric	Followers
Alternators and generators	Synchronous	Pressure vessel	Taper	Hydrodynamic	Geneva Mechanisms	Static seals	Sound	Housings
Solenoids	Chains	Solid mass	Spring	Hydrostatic	Valves	Gaskets	Flow	Frames
Pneumatic and hydraulic actuators	Roller	Chemical	Rivets	Hydrodynamic	Latches	O rings	Laser doppler	Casings
Brakes	Leaf	Charge	Tolerance rings	Slideway	Triggers	Gaskets	Hot wire	Enclosures
Disk	Conveyor	Torsion bar	Expanding bolts	Wheels and rollers	Bimetallic strips	Sealants	Ultrasonic	Sprayers
Drum and band	Silent		Keys	Brushes			Level	Shutters
Pumps	Cables and ropes		Flat, round				Humidity	Hooks
Rockets	Couplings		Profiled				Temperature	Pulleys
Heat exchangers	Rigid		Gib head				Thermocouples	Handles
Guns	Flexible		Woodruff				Resistance	Rollers and drums
Dampers and shock absorbers	Universal		Splines				Thermometers	Centrifuges
Fuel cells	Cranks		Circlips				Pyrometers	Filters
	Cams		Snap rings				Heat flux	
	Power screws and threads		Clamps				Thermopile	
	Levers and linkages		Retaining rings				Gardon gauges	
	Pipes, hoses and Ducts		Shoulders				Strain and stress	
	Ball screws		Spacers				Time	
			Grooves				Chemical composition	
			Fits					
			Clearance					
			Transition					
			Interference					
			Adhesives					
			Welds					

After private communication, C. McMahon, Bristol University, 1997.

safety and reliability. The function tree method helps expand on these objectives by use of the questioning adverbs: why, what, where and how? For instance, what is meant by reliable? As a result of this question this statement can be developed into: high number of operations without attention, high cutting use without attention, low wear of moving parts, low level of vibration, low risk of misuse.

The process of developing a list of functions or objectives will usually involve several different levels. For example the means of achieving an objective will be an objective in itself, albeit a lower level of objective. The process of developing functions in this way results in a branching of functions from the original key objectives. The objectives tree for a lawnmower that is to be safe, reliable, convenient to use and simple to make has been developed in Figure 10.1.15.

10.1.6 The technology base

The range of existing and well-tried technology is extensive. Many items are available as standard components and can be purchased from specialist suppliers. In addition, standard practice and design methodology have been specified for many others. The designer is able to exploit available technology as it is or by combination and adaptation. The scope of mechanical machine elements available to the designer is outlined in Table 10.1.8. This list is not exhaustive, but does give an idea of some of the 'building blocks' available. Many of these machine elements are identified by shading in the table.

Developing an overview of the available technology base is useful to the designer as it avoids time spent repeating the design process on a technology that has already been developed. In addition, knowledge of science and engineering enables a designer to judge what is technically and scientifically feasible so that 'far-fetched' ideas can be objectively ruled out.

10.1.7 Conclusions

The total design model illustrated in Figures 10.1.7 and 10.1.9 was originally proposed some years ago (1985). Since then the ethos of design has developed to include the design of products for total life recycling. This means that the end of a product should be considered at the design stage. When the product finally fails or is deliberately scheduled for withdrawal, consideration should be given to its disposal or recycling. For instance, are there any environmentally threatening aspects that must be dealt with? Some automotive companies, for example, have developed recycling policies whereby approaching 100 per cent of an old exchanged vehicle is recycled. This has the added benefit of giving the company the opportunity to sell a new product to the customer.

Although the total design model has been adopted for this course, other models have been developed. The reader is recommended to review the models proposed by Pahl and Beitz (1996), March (1976) and Cross (2000). For example, in the total design model the conceptual design phase includes the development of bulk concepts as well as the process of providing form and layout to the design. The model presented by Pahl and Beitz (1996) models this process as distinct conceptual and embodiment phases. In the embodiment phase the layout and form of the design concept are quantified with consideration of the overall objectives of the specification. The subject of design has been practised for thousands of years but understanding of the process of design is in its infancy and all these models will no doubt be developed further in the future.

References

Baxter, M. (1995). *Product Design.* Chapman and Hall.

Boothroyd, G. (1997). Design for manufacture and assembly, in *ASM Handbook, vol. 20. Material Selection and Design. ASM* International, pp. 676—686.

Cross, N. (2000). *Engineering Design Methods. Strategies for Product Design,* 3rd edn. Wiley.

Ion, B. and Smith, D. (1996). *The Conceptual Design Phase.* SEED.

Manzini, E. (1989). *The Material of Invention.* Design Council.

March, L.J. (1976). *The Architecture of Form.* Cambridge University Press.

Matthews, C. (1998). *Case Studies in Engineering Design.* Arnold.

McMahon, C. (1997). Private communication. University of Bristol.

Pahl, G. and Beitz, W. (1996). *Engineering Design a Systematic Approach,* 2nd edn. Springer.

Pugh, S. (1990). *Total Design. Addison* Wesley.

Sharing Experience in Engineering Design. (1985). *Curriculum for Design. Engineering Undergraduate Courses.* SEED.

Walker, D.J., Dagger, B.K.J. and Roy, R. (1991). *Creative Techniques in Product and Engineering Design.* Woodhead Publishing Ltd.

Wright, I.C. (1998). *Design Methods in Engineering and Product Design.* McGraw Hill.

Standards

99/705213 DC. ISO 18334. *Timber pallets for materials handling. Quality of new wood pallet assembly.*

99/712554 DC.PrEN 13698-1. *Pallet product specification. Part 1. Construction specification for 800 mm* X *1200 mm wooden pallet.*

99/712555 DC. PrEN 13698-2. *Pallet product specification. Part 2. Construction specification for 1000 mm* X *1200 mm wooden pallet.*

British Standards Institution. (1978). BS 3726:1978, ISO 1074-1975. *Specification for counterbalanced lift trucks — stability — basic tests.*

British Standards Institution. (1978). BS 5639-1:1978, ISO 2331-1974. *Fork arms for fork lift trucks. Vocabulary for hook-on type form arms.*

British Standards Institution. (1988). BS ISO 6780: 1988. *General purpose flat pallets for through transit of goods. Principal dimensions and tolerances.*

British Standards Institution. (1996). BS ISO 509: 1996. *Pallet trucks. Principal dimensions.*

British Standards Institution. (1999). BS EN 1726- 1:1999. *Safety of industrial trucks. Self-propelled trucks up to and including 10000 kg capacity and industrial tractors with a drawbar pull up to and including 20000 N. General requirements.*

British Standards Institution. (1999). BS EN ISO 445:1999. *Pallets for materials handling Vocabulary.*

British Standards Institution. (2002). BS EN 13545:2002. *Pallet superstructures. Pallet collars. Test methods and performance requirements.*

British Standards Institution. (2002). BS ISO 2330: 2002. *Fork-lift trucks. Fork arms.Technical characteristics and testing.*

Websites

At the time of going to press the World Wide Web contains useful information relating to this chapter at the following sites.

www.bsi.org.uk
www.iso.org
www.hyster.com/products/
www.nissanforklift.com
www.ski-doo.com
www.jcb.co.uk

Nomenclature

BS British Standard
DFA Design for assembly
DFM Design for manufacture
ISO International Organisation for Standardisation
PDS Product design specification

This page is intentionally left blank

Chapter 10.2

10.2

Mechanical elements

Peter Childs

The aim of this chapter is to introduce machine design and the individual machine elements that go together to form an overall machine design. The chapter introduces fundamental design philosophies.

10.2.1 Introduction

The design activity involves endowing form and ensuring functionality of a given product. The function of many products depends on devices that modify force or motion and consist of a number of interrelated units. An example is a gearbox that modifies the torque speed characteristic of a prime mover to match that of the load and comprises gears, shafts, bearings and seals. Such devices that modify force or motion are called machines and the interrelated units are sometimes known as 'machine elements'. The technology base has developed to a mature status in this area providing a number of such devices that can in some cases be used as they are or designed fit-for-purpose.

In the design of a machine and its machine elements it is the designer's task to determine the motion, forces and changes in energy involved so that sizes, shapes and materials for each machine element making up the whole machine can be determined. Although design work may involve concentrating on one component at a given time, it is essential to consider its interrelationship with the whole product following the total design philosophy, taking into account the market requirement, detailed specifications, the conceptual design, design of other components and manufacture requirements.

The goal in machine design is to determine the size, shape and selection of materials and manufacturing processes for each of the parts of the machine so that the machine will perform its intended function without failure. This requires the designer to model and predict the performance and operation of each component and overall assemblies studying the mode of failure for each and ensuring that conditions expected in service for the product are met. This may involve determining stresses, deflections, temperatures and material degradation for each and every component and overall assembly, as well as its influence within the total design. As stresses are a function of forces, moments and torques, consideration of the dynamics of the system will often be necessary before detailed stress analysis can be undertaken. Furthermore, as the strength of materials is a function of temperature, a thermal analysis may also be necessary in order to determine operating temperatures throughout the performance cycle of the machine.

If a machine has no moving parts then the task for the designer is simpler since only static force analysis and thermal analysis is necessary. However, even structures are subject to loading from external loads, such as weather conditions of wind and rain, earth tremors and traffic. As such a consideration of the environment into which the machine is to be installed should be made and appropriate failure modes mitigated against. Mitigation involves considering what may go wrong and determining what action to take to either prevent it from going wrong in the first case or if it does go wrong what is going to be done about it.

If the motions involved in a machine are slow then it is likely that a static force analysis will be sufficient. In a static structure such as the floor of a casing designed to support a particular weight, the payload that the structure can support can be increased by adding

Mechanical Design; ISBN: 9780750657716
Copyright © 2004 Elsevier Ltd. All rights of reproduction, in any form, reserved.

appropriately distributed weight (dead weight) to its structure. If, however, the components of a machine have significant accelerations, a dynamic force analysis will be necessary. Adding additional mass to a moving component increases the forces involved and it is possible for such components to become victims of their own mass. This is because some of the loading causing stress in a moving component is due to inertial forces modelled by Newton's second law of motion

$$F = ma \tag{10.2.1}$$

Although addition of mass may increase the strength of a component, the benefit may be reduced by the increases in inertial forces.

To meet a design requirement, the designer needs to conceive a form and express the connectivity of its parts in order to meet the desired goal. A proportion of the mental sculpting involved is geometric in nature because shape often permits the achievement of a given function. The design procedures presented here follow this approach with an initial proposal for the geometric configuration followed by analysis to justify the proposal or provide guidance for its modification. Boden (1990) discusses creativity in terms of the connection of conceptual spaces in the mind. In order to create a design it is often necessary to combine ideas from different fields or disciplines. This may or may not take place as part of a structured design procedure or in a prescribed time period and over-reliance on procedure should in some cases actually be guarded against as this may stymie creativity.

The variety of machine elements and technologies already available to the designer as a result of years of development is extensive as indicated in Table 10.1.8 in Chapter 10.1. Machine elements considered in this chapter are those more commonly found in machines transmitting power by means of moving components, such as motors and engines. Figure 10.2.1 illustrates the range of machine elements involved in the design of a gearbox for a single cylinder air compressor and comprises: shafts, bearings, gears, seals and enclosure. The machine elements considered in this chapter provide skills in the design of such applications and cover bearings, shafts, gears, belts and chains, seals, clutches and brakes, springs, fasteners and enclosures. These subjects are introduced in Sections 10.2.3–10.2.9. A common feature to many machine elements is the wear of surfaces and this subject is introduced in Section 10.2.2.

10.2.2 Tribology

Tribology is the study of friction, lubrication and wear of surfaces in relative motion. Because of the desire to produce efficient machines where the loss of energy due to friction is minimal, extensive research and effort has been put into studying tribology. As a subject it has considerable interrelationships and crossover with materials and fluid mechanics. The study and reduction of friction has been of concern to humans for as long as machines have been in use and the subject of tribology includes consideration of machine elements such as bearings, gears, belts, chains, clutches, brakes and seals.

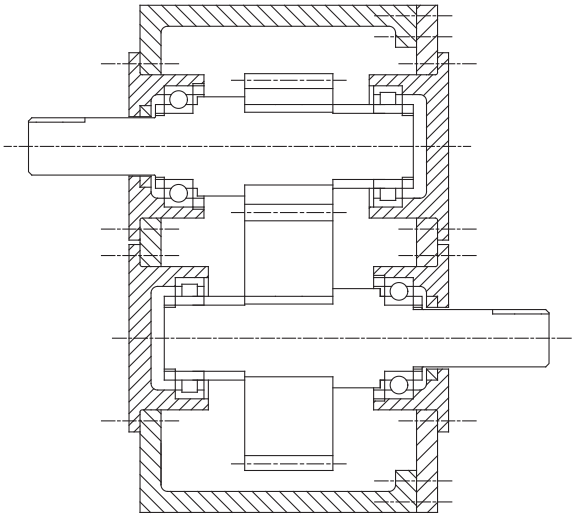


Fig. 10.2.1 Transmission for an engine driven compressor.

10.2.3 Bearings

The term 'bearing', in its general sense, is used to refer to the assembly formed by two surfaces that have the capacity for relative motion. A wide range of bearings has been developed some of which involve rolling motion, some sliding motion and some both. Lubrication of a bearing is frequently used to reduce friction and therefore the wear and power absorbed by the bearing, and to remove heat to ensure operation of the bearing assembly at temperatures compatible with the materials and lubricant used.

Typical examples of bearings are shown in Figures 10.2.2 and 10.2.3. Figure 10.2.2 illustrates rolling bearings. These come in many forms each designed for a given set of performance characteristics, such as speed and load capability. Rolling element bearings usually need to be lubricated by grease or oil with the choice depending on the speed and temperature of operation. Rolling element bearings are typically available as stock items. This is of significant potential advantage to a machine designer, as the task becomes that of bearing selection rather than bearing design. Bearing design is a significant undertaking involving consideration of very high point

Fig. 10.2.2 Examples of rolling element bearings available from existing suppliers.

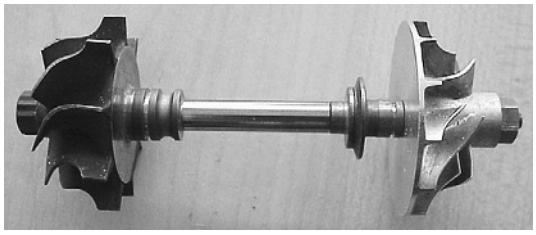


Fig. 10.2.3 Journal bearing for a turbocharger rotor.

contact stresses and use of special materials. The technology involved relies heavily on the science of tribology and there are a number of worldwide companies already operating in this market.

A designer needs to carefully consider the issue of whether time is better spent producing a new design for a bearing that is already available from an existing supplier or whether it is better spent on designing and adding value to the product in hand.

Plain bearings rely on either rubbing contact between the bearing surfaces or pressure to separate the two surfaces. The pressure required can be induced in a lubricant by movement of one component relative to another or by the external supply of the lubricant under the required pressure. Plain bearings are also available from specialist bearing suppliers. In principle they can be significantly cheaper than rolling element bearings as they involve fewer moving components. The selection of an appropriate bearing for a given task, however, is an involved activity, which needs to take into account, amongst other factors:

- load
- speed
- location
- size
- cost
- starting torque
- noise
- lubrication supply.

A consequence of bearings being supplied by a select number of specialist manufacturers is that they have been subject to standardization and are usually only available in a limited number of standard sizes. Standardization has the advantage that parts from one original equipment manufacturer (OEM) will be interchangeable in terms of geometry with another. This does not, however, mean that the performance of two similarly sized components from different manufacturers will be the same.

10.2.4 Gears, belts and chains

Gears, belts and chains can be used to transmit power from one shaft to another. Typical examples are illustrated in Figures 10.2.4–10.2.6.

Many power producing machines, or prime movers, such as internal combustion engines, gas turbines and electric motors produce power in the form of rotary motion. The operating characteristics of prime movers vary according to their type and size. Typical examples are illustrated in Figure 10.2.7 for an internal combustion engine, an alternating current electric motor and a series wound direct current electric motor. Unfortunately these characteristics may not match those of a given application. For example, the internal combustion engine

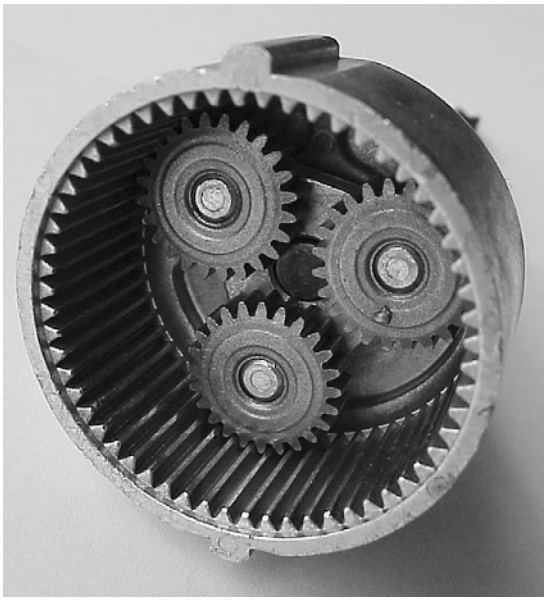


Fig. 10.2.4 Spur gears.

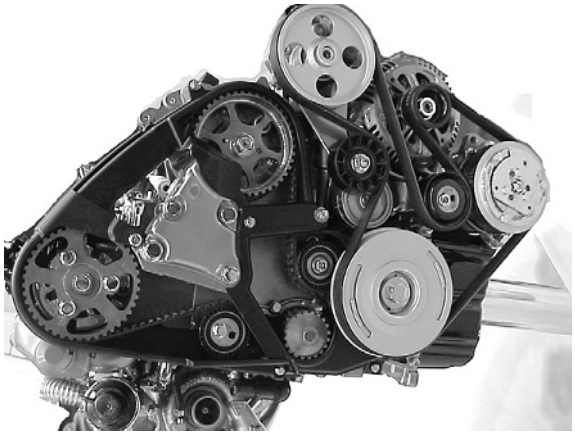


Fig. 10.2.5 Flat belt and synchronous belt drives.

Fig. 10.2.6 Roller chain.

gives little torque at the low speeds associated with starting. It is for this reason that a car engine requires a clutch to decouple the engine from the load during starting and a gearbox to transform the characteristic of the prime mover to match that of the load.

A change of speed can be used to transform the torque speed characteristic of a prime mover to a useful output characteristic. An example is the use of a reduction gearbox to reduce the speed and increase the torque seen by a load as illustrated in Figure 10.2.8. Recall the relationship between torque, angular velocity and power,

$$\text{torque} = \text{power}/\text{speed} \tag{10.2.2}$$

Examination of this equation shows that as the speed reduces the torque rises for a given power. So if increased torque is required then a reduction gearbox, which decreases the speed of the output shaft relative to the input shaft can be used to increase the torque delivered by the output shaft of the gearbox.

When transmitting power from a source to the required point of application a series of devices is available including belts, pulleys, chains, hydraulic and electrical

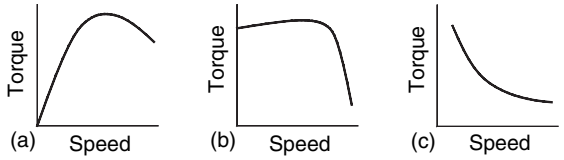


Fig. 10.2.7 Typical torque speed characteristics. (a) Internal combustion engine. (b) Three phase induction alternating current motor. (c) Series wound direct current motor.

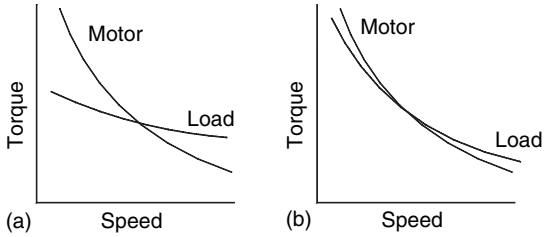


Fig. 10.2.8 Transformation of the torque speed characteristic using a reduction gear ratio. (a) Mismatched system. (b) Transformed system.

systems, and gears. Generally, if the distances of power transmission are large, gears are not suitable and chains and belts can be considered. However, when a compact, efficient or high-speed drive is required gear trains offer a competitive and suitable solution.

Gears, belts and chains are all available from specialist suppliers. The manufacture of gears requires specialist machinery and the stresses in the region of contact between gears in mesh can be exceptionally high requiring the use of specialist materials. Similarly the design of roller chain and associated sprockets involves high contact stresses, and in addition the design of an articulated joint and associated bearings. For these reasons, combined with financial expediency, leaving the cost of specialist manufacture infrastructure to other companies, gear, chain and belt manufacture is usually best left to the specialist and the task of the machine designer becomes that of selecting and specifying appropriate components from the specialist stock suppliers and OEMs.

10.2.5 Seals

Seals are devices that are used to reduce or eliminate the flow of fluid or particulates between two locations. Seals are an important part of a machine design where fluids under relatively high differential pressures must be contained within a particular region. An example is the fluid within the cylinder of an internal combustion engine. Here the working fluid needs to be contained in the cylinder for part of the cycle and excluded, in as much as possible, from contaminating the lubricating oil for the

Fig. 10.2.9 Sealing of an internal combustion engine cylinder.

crankshaft bearings. This is conventionally achieved by means of piston rings, illustrated in Figure 10.2.9.

There are many types of seal, but they can generally be classified as either static or dynamic depending on whether relative movement is involved. Many static seals involve use of a gasket or sealing material, such as an elastomer. These are widely available from specialist suppliers and the task facing the designer is the selection of appropriate materials for the application. Some dynamic seals are also commercially available such as radial lip seals and mechanical face seals. However, there is not the level of standardization in shaft design and the range of possible applications is seemingly infinite with differing pressure ratios and fluids being sealed. As a result the range of seal geometries required is too extensive to warrant an OEM designing or stocking a wide number of seal sizes and types. Instead many dynamic seals are designed fit-for-purpose. Typical considerations in the selection and design of seals include the nature of the fluid or particulates to be contained, the differential pressure, the nature of any relative motion between components, the level of sealing required, operating conditions, life expectancy and total costs.

10.2.6 Clutches and brakes

Clutches and brakes provide frictional, magnetic, hydraulic or mechanical connection between two machine elements, usually shafts. There are significant similarities between clutches and brakes. If both shafts rotate then the machine element will be classed as a clutch and the usual function is to connect or disconnect a driven load from a driving shaft. If one shaft rotates and the other is stationary then the machine element is classed as a brake and the likely function is to decelerate a shaft. In reality,

Fig. 10.2.10 An automotive clutch.

however, the same device can function as a brake or clutch by fixing its output element to a shaft that can rotate or to ground, respectively.

Brakes and clutches are familiar devices from their use in automotive applications as illustrated in Figures 10.2.10 and 10.2.11. They are, however, also used extensively in production machinery. Clutches allow a high inertia load to be rotated with a smaller electric motor than would otherwise be required if it was directly connected. Clutches are also used to maintain a constant torque on a shaft for tensioning of webs and filaments. They can be used, in emergencies, to disconnect a driven machine from a motor in the event of jamming of the machinery. In such cases a brake may also be installed in order to bring the machinery to a rapid stop.

Complete clutches and brakes are available from specialist manufacturers. In addition, key components such as discs and hydraulic and pneumatic actuators are also available enabling the designer to opt, if appropriate, to design the overall configuration of the clutch or brake and buy in say the frictional lining surface.

10.2.7 Springs

Springs are flexible elements used to exert force or torque and store energy. The force exerted by a spring can be a linear push or pull or a radial force similar to that of a rubber band around a roll of paper. Alternatively, a spring can be configured to produce a torque with applications including door closers. Springs store energy when they are deformed and return energy

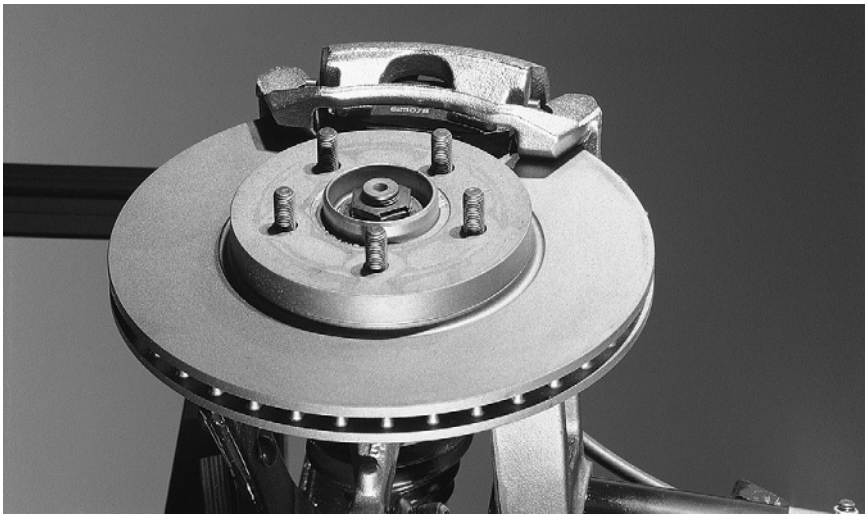


Fig. 10.2.11 An automotive brake.

when the force that is causing the deflection is removed. A further type of spring is the power springs or spring motor, which once wound up can dissipate energy at a steady pace.

Springs can be made from round or rectangular wire bent into a suitable form, such as a coil, or made from flat stock and loaded as a cantilevered beam. Many standard size spring configurations are available as stock items from specialist manufacturers and suppliers and it is usually expedient to use these where possible. As well as paying attention to the performance characteristics of the spring, attention should also be focussed on the spring ends or terminations. Many options are available such as twist loops or hooks, side loops, extended hooks and so on. Some applications require a custom spring design. Spring design involves consideration of many variables such as length, wire diameter, forces, spring rate, spring index, number of coils, pitch, pitch angle, installation considerations, coil clearance, materials, types of loading and allowable stresses.

10.2.8 Fasteners

Fasteners connect or join two or more components. It has been estimated that there are over 2,500,000 fasteners in a Boeing jumbo jet. For more modern airliners, such as the Airbus Industrie A340, this figure has been dramatically reduced, but the total is still significant. This type of example illustrates the importance of fasteners to the designer.

There are thousands of types of fastener commercially available from specialist suppliers. The most common include threaded fasteners such as bolts, screws, nuts, studs and set screws, rivets, fast operating fasteners requiring say a quarter turn to connect or disconnect and snap joints commonly used in plastic component assembly. In addition, there is the range of permanent fastening techniques by means of welding, brazing, soldering and use of adhesive.

Given the wide range of fastening techniques available and the costs involved in sourcing, stocking and assembly, the correct choice of fastener is very important.

10.2.9 Enclosures

Individual machine elements, such as shafts or bearings, must themselves be supported and this is the function of an enclosure or frame. The total design model requires attention to be applied to all other phases of the design process, whilst undertaking any one task. As such the design of any given machine element should not occur in isolation, but instead give due consideration to how it will fit in and interact with the overall assembly. The enclosure for a machine provides not only the support for the machine, but also provides form to a product. Attention to visual attractiveness is not always given the attention it warrants by designers of machine elements. However, we produce our components to be purchased and current thinking indicates that visual perception is the dominant amongst the senses with over half the cerebral cortex dedicated to visual processing. The shape of our products is therefore important and we should design accordingly.

Enclosures are usually unique to a given class of machine, as each machine tends to be different in terms of the number, size and function of its constituent components. It is not practical, therefore, to have a completely general approach to enclosure and frame

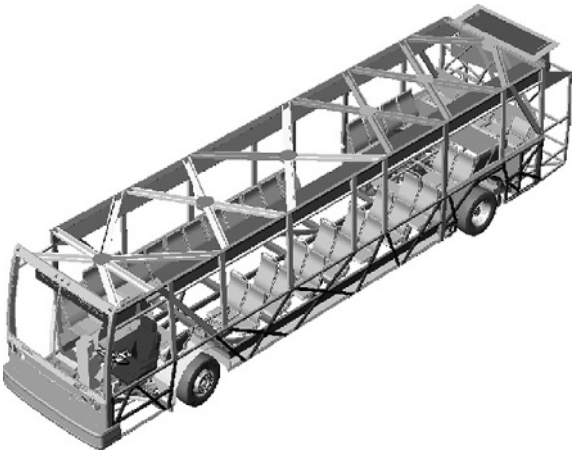


Fig. 10.2.12 Bus frame design (Allin and Simon, 2002)

design. Common examples of the variation of frames and enclosures are those for vehicles and toys. Toys must be safe, functional and fun, whilst minimizing the material and personnel effort in producing the toy, in order to minimize the production cost and maximize profit.

The design of an enclosure or frame is to some extent an art in that the constituent machine components must be accommodated. There may be constraints on where supports can be placed in order not to interfere with the function and operation of the machine or in order to ensure access to particular components. The frame for the bus shown in Figure 10.2.12 illustrates this point in that access for doors, windows and occupants must be provided.

Some of the more important design parameters to consider for an enclosure or frame include strength, size, assembly, appearance, corrosion resistance, stiffness, vibration, weight, noise, cost to produce and maintain, and life. A list such as this may be too all-embracing to be of much use at the start of a design process and a more helpful list of factors to consider as the starting point for a frame or enclosure include:

- forces exerted by the machine components through mounting points, such as bearings, pivots, brackets and feet;
- how the frame itself will be supported;
- allowable deflections of constituent machine elements;
- cooling requirements;
- the environments into which the machine will be transported and installed and the relationship to other machines and infrastructure;
- the quantity required and production facilities available;
- the expertise available for the design task.

The complexity involved when considering more that one machine element is evident and because of the need to keep the bigger picture in mind, models for the design process such as total design are useful.

10.2.10 Conclusions

The technology base available to us includes a wide range of machine elements, such as bearings, shafts, gears, chains, belts, seals, clutches, brakes and fasteners. This chapter has served to introduce these items, many of which are available as stock items from specialist manufacturers. This means that the designer's task becomes that of selecting and specifying the appropriate item and relieves him or her of the task of the detailed design of, for instance, a roller chain so the design of the overall machine can be concentrated on.

It is appropriate to take time in overviewing the technologies available to the designer in order to be able to consider and take into account the interrelationships between individual machine elements. This is in keeping with the total design model where each phase takes into account all other phases and aspects of the design process.

References

Allin, K. and Simon, D. (2002). A new approach to vehicle architecture design. In R.K. Stobart and P.R.N. Childs (eds) *Total Vehicle Technology. How Do We Get the Innovation Back into Vehicle Design?* Professional Engineering Publishing, pp. 49–56.

Boden, M.A. (1990). *The Creative Mind: Myths and Mechanisms* Weidenfeld and Nicolson.

Nomenclature

Generally, preferred SI units have been stated.

a	acceleration (m/s^2)
F	force (N)
m	mass (kg)
OEM	original equipment manufacturer

This page is intentionally left blank

Section **Eleven**

Design techniques: plastics

This page is intentionally left blank

Chapter 11.1

11.1

Overview

Dominick V. Rosato and Donald V. Rosato

11.1.1 Introduction

Section 'Design techniques: plastics' of this book provides information on the behavior of plastics that influence the application of practical and complex engineering equations and analysis in the design of products. For over a century plastics with its versatility and vast array of inherent plastic properties as well as high-speed/low-energy processing techniques have resulted in designing and producing many millions of cost-effective products used worldwide. The profound worldwide benefits of plastics in economics and modern living standards have been brought about by the intelligent application of logic with modern chemistry and engineering principle.

Today's plastics industry is comprised of both mature practical and theoretical technology. Improved understanding and control of materials and manufacturing processes have significantly increased product performances and reduced their variability. Performance requirements for these products can be characterized in many different ways. Examples meeting different commercial and industrial market requirements worldwide include:

1. lightweight,
2. flexible to high strength,
3. provide packaging aesthetics and performances,
4. excellent appearance and surface characteristics without using secondary operations,
5. degradation resistance in different environments,
6. performance in all kinds of environments,
7. adapt well to mass production methods,
8. wide range of color and appearance,
9. high impact to tear resistance,
10. decorative to industrial load bearing structures,
11. short to very long service life, degradable to non-degradable,
12. process virgin with recycled plastics or recycled alone,
13. simple to complex shapes including many that are difficult or impossible to form with other materials,
14. breathable film for use in horticulture,
15. heat and ablative resistance,
16. and so on.

There is a plastic for practically any product requirements, particularly when not including cost for a few products. One can say that if plastics were not to be used it would be catastrophic worldwide for people, products, communications, and so on with a major economic crisis because much more expensive materials and processes would be used.

Materials can be blended or compounded to achieve practically any desired property or combination of properties. The final product performance is affected by interrelating the plastic with its design and processing method. The designer's knowledge of all these variables is required otherwise it can profoundly affect the ultimate success or failure of a consumer or industrial product. When required the designer makes use of others to ensure product success.

Plastic plays a crucial and important role in the development of our society worldwide. With properties ranges that can be widely adjusted and ease of processing, plastics can be designed to produce simple to highly integrated conventional and customized products. While it is mature, the plastics industry is far from having

Plastics Engineered Product Design; ISBN: 9781856174169
Copyright © 2003 Elsevier Ltd. All rights of reproduction, in any form, reserved.

exhausted its product design potential. The worldwide plastics industry offers continuous innovations in plastic materials, process engineering, and mechanical engineering design approaches that will make it possible to respond to ever more demanding product applications (Fig. 11.1.1).

Innovation trends emerging in plastics engineering designs are essentially combinations and improvements of different processes, combinations and improvements of different materials, integration of a wide range of functions within a single product, reduced material consumption, and recyclability of the materials employed. At the same time, rising requirements are being placed on design efficiency, product quality, production quality, and part precision, while costs are expected to be reduced wherever possible. This combination of objectives is achievable by factors such as process-engineering innovations that reduce the number of process steps.

The basic and essential design exercise in product innovation lies in predicting performances. This includes the process of devising a product that fulfills the total requirements of the end user and satisfies the needs of the producer in terms of a good return on investment. The product designer must be knowledgeable about all aspects of plastics such as behavioral responses, processing, and mechanical and environmental load stresses. Product loads range from short-time static, such as tensile, flexural, torsion, etc., to long time dynamic, such as creep, fatigue, high-speed loading, motion control, and so on. Plastics design concepts that can be applied to designing products for a range of behaviors are presented here.

An inspired idea alone will not result in a successful design. Designing is, to a high degree, intuitive and creative, but at the same time empirical and technically influenced. Experience plays an important part that requires keeping up to date on the endless new developments in materials and processes. An understanding of one's materials and a ready acquaintance with the relevant processing technologies are essential for converting an idea to an actual product. In addition, certain basic tools are needed, such as those for computation and measurement and for testing of prototypes and/or fabricated products to ensure that product performance requirement are met. A single individual designer may not have all of these capabilities so inputs from many reliable people and/or sources are required.

Inputs from many disciplines, both engineering and non-engineering, may be required when designing a product such as a toy, flexible package, rigid container, medical device, car, boat, underwater device, spring, pipe, building, aircraft, missile, or spacecraft. The conception of such products usually requires coordinated inputs from different specialists. Input may involve concepts of man-machine interfaces (ergonomics), shape, texture, and color (aesthetics). Unless these are in balance, the product may fail in the market place. The

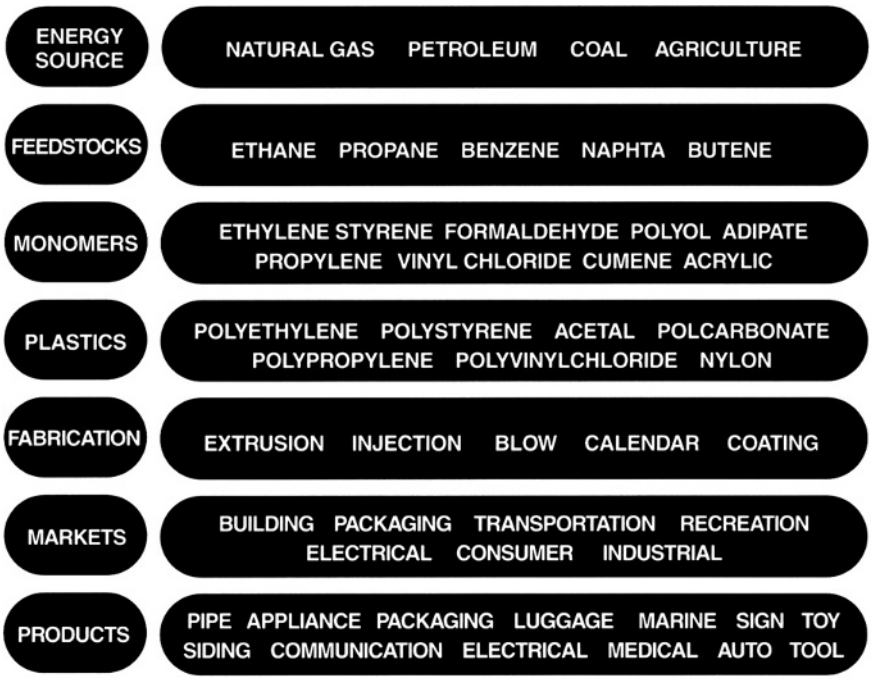


Fig. 11.1.1 Flow chart from raw materials to products. (Courtesy of Plastics FALLO.)

successful integrated product is the result of properly collecting all of the required design inputs.

While plastic product design can be challenging, many products seen in everyday life may require only a practical, rather than rigorous approach. They are not required to undergo sophisticated design analysis because they are not required to withstand high static and dynamic loads. Their design may require only the materials information in conventional data sheets from plastic material suppliers. Examples include containers, cups, toys, boxes, housings for computers, radios, televisions, electric irons, recreational products, and nonstructural or secondary structural products of various kinds like the interiors in buildings, automobiles, and aircraft. The design engineer will need to know when to use the practical approach, the rigorous approach, or a combination approach.

Plastics do not only have advantages but also have disadvantages or limitations. Other materials (steel, wood, etc.) also suffer with disadvantages or limitations. Unfortunately there is no one material (plastic, steel, etc.) that can meet all requirements thus these limitations or faults are sometimes referred to incorrectly as disadvantages. Note that the faults of materials known and utilized for hundreds of years are often overlooked; the faults of the new materials are often overemphasized.

Iron and steel are attacked by the elements of weather and fire [815 °C (1500 °F)] but the common practice includes applying protective coatings (plastic, cement, etc.) and then forgetting their susceptibility to attack is all too prevalent. Wood is a useful material yet who has not seen a rotted board, wood on fire, etc. There is cracked concrete and so on. Regardless of these and many other disadvantages, lack of perfection does not mean that any steel, wood, or concrete should not be used. The same reasoning should apply to plastics. In many respects, the gains made with plastics in a short span of time far outdistance the advances made in these other materials.

Recognize that modern design engineering has links with virtually every technical area; material, mechanical, electrical, thermal, processing, and packaging to name a few. Any attempt to explain engineering by referring to the special rules for each area would mean that the engineer would need to have a thorough knowledge of each special field. This is not possible in the current state of technology. It is the case, however, that there are certain common concepts behind these specialized areas. Similar features exist among many consumer and industrial products. These features can be described by using a standard procedure, and the fundamental laws of engineering apply to all products, irrespective of the different forms of materials and equipment involved. Proper applications are required.

11.1.2 Materials of construction

Plastics comprise many different materials based on their polymer structure, additives, and so on. Practically all plastics at some stage in their fabrication can be formed into simple to extremely complex shapes that can range from being extremely flexible to extremely strong. Polymers, the basic ingredients in plastics, are high molecular weight (MW) organic chemical compounds, synthetic or natural substances consisting of molecules. Practically all of these polymers use virtually an endless array of additives, fillers, and reinforcements to perform properly during product fabrication and/or in service. There are many compounded base polymer combinations so that new materials are always on the horizon to meet new industry requirements that now total over 35,000 plastics worldwide. Table 11.1.1 provides examples of their manufacturing stages from raw materials to products. Examples of the diversified use of different plastics are shown in Figures 11.1.2 and 11.1.3.

Plastic, polymer, resin, elastomer, and reinforced plastic (RP) are somewhat synonymous terms. The most popular term worldwide is plastics. Polymer denotes the

Table 11.1.1 Examples of stages in plastic manufacturing

Basic chemicals Petroleum is converted to petrochemicals such as ethylene, benzene, propylene and acetylene.
Monomers Petrochemicals plus other chemicals are converted into monomers such as styrene, ethylene, propylene, vinyl chloride, and acrylontrile.
Polymerization One or more monomers are polymerized to form polymers or copolymers such as polyethylene, polystyrene, polyvinyl chloride, and polypropylene.
Compounding Additives, fillers, and/or reinforcements are mixed with polymers (referred to as plastics) providing different properties and/or different fabricating methods for plastics. Hundreds of different materials are used such as heat stabilizers, color pigments, antioxidants, inhibitors, and fire retardants.
Processing Plastics are formed into different shapes such as sheets, films, pipes, buckets, primary and secondary structures (boats, cars, airplanes, bridges, etc.), toys, housings, and many thousand more products. Basically heat and pressure are used to shape these products that usually are in finished form. Processes used include extrusion, injection molding, blow molding, thermoforming, compression molding, spraying, rotational moldings, reaction injection molding, and filament winding.
Finishing In certain applications a finishing step is required on the fabricated part such as printing, bonding, machining, etc.

basic material. Whereas plastic pertains to polymers or resins (as well as elastomers, RPs, etc.) containing additives, fillers, and/or reinforcements. An elastomer is a rubber-like material (natural or synthetic). Reinforced plastics (also called plastic composites) are plastics with reinforcing additives such as fibers and whiskers, added principally to increase the product's mechanical properties but also provides other benefits such as increased heat resistance and improved tolerance control.

There are thermoplastics (TPs) that melt (also called curing) during processing. Cure occurs only with thermoset (TS) plastics or when a TP is converted to a TS plastic and in turn processed. The term curing TPs occurred since at the beginning of the twentieth century the term curing was accurately used for TSs. At that time TSs represented practically all the plastic used worldwide. Thus TPs took on the incorrect term curing even though there is no chemical reaction or curing action.

Fig. 11.1.2 Use of plastics in recreational products range from unsophisticated types to high-performance types such as boats. (Courtesy of Plastics FALLO.)

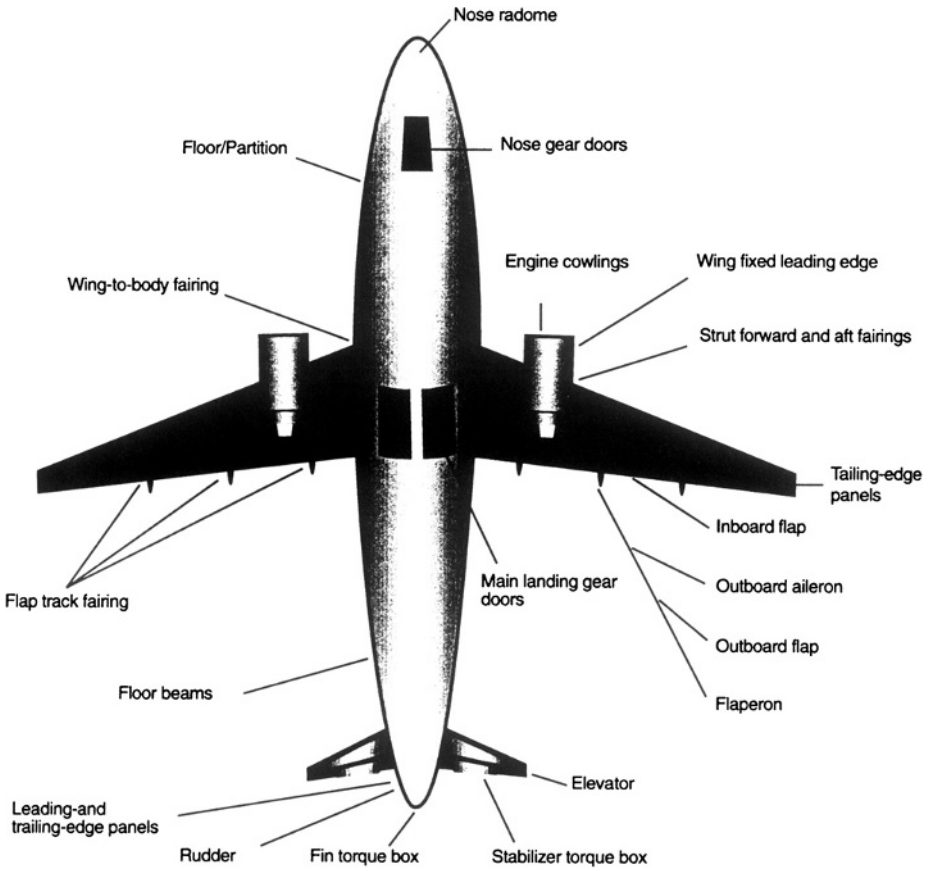


Fig. 11.1.3 Boeing 777 uses different types of plastics that include high-performance RPs.

Appreciate the polymer chemist's ability to literally rearrange the molecular structure of the polymer to provide an almost infinite variety of compositions that differ in form, appearance, properties, cost, and other characteristics. One must also approach the subject with a completely open mind that will accept all the contradictions that could make it difficult to pin common labels on the different families of plastics or even on the many various types within a single family. Each plastic (of the 35,000 available) has specific performance and processing capabilities.

There are many different routes that the starting materials for plastics can take on the way to the user. Here, we are concerned with those plastics that are supplied to the processor in the form of granules, powder, pellets, flake, or liquids and in turn they are transformed into plastic products. However, the same starting materials used to make these plastics can take other routes and end up in the textile industry (nylon fibers share common roots with a molded nylon gear; acrylic fibers share common roots with acrylic sheet for glazing; etc.), paint industry, adhesives industry, and other industries meeting their special requirements.

Worldwide total plastic consumption is over 154 million ton (340 billion lb) with about 90 wt% TP and 10% TS plastics. USA and Europe consumption is about one-third each of the world total.

These two major classifications of TPs and TSs in turn have different classifications such as virgin or recycled plastics. Virgin plastics have not been subjected to any fabricating process. NEAT plastics identify plastics with *N*othing *E*lse *A*dded *T*o. They are true virgin polymers since they do not contain additives, fillers, etc. However, they are rarely used since they do not provide the best performances. Thus the technically correct term to identify the materials is plastics. Of the 35,000 types available worldwide there are about 200 basic types or families drat are commercially recognized with less than 20 that are popularly used. Examples of these plastics are shown in Table 11.1.2.

Within these 20 popular plastics there are five major TP types that consume about two-thirds of all TPs. Approximately 20 wt% are low-density polyethylenes (LDPEs), 15% polyvinyl chlorides (PVCs), 10% high-density polyethylenes (HDPEs), 15% polypropylenes (PPs), 8% polystyrenes (PSs). Each has literally many thousands of different formulated compounds and

Table 11.1.2 Summation of the plastic families with their abbreviations

- Acetal (POM)
- Acrylics
  - Polyacrylonitrile (PAN)
  - Polymethylmethacrylate (PMMA)
- Acrylonitrile butadiene styrene (ABS)
- Allyl diglycol carbonate
- Alkyd
  - Diallyl isophthalate (DAIP)
  - Diallyl phthalate (DAP)
- Aminos
  - Melamine formaldehyde (MF)
  - Urea formaldehyde (UF)
- Cellulosics
  - Cellulose acetate (CA)
  - Cellulose acetate butyrate (CAB)
  - Cellulose acetate propionate (CAP)
  - Cellulose nitrate
  - Ethyl cellulose (EC)
- Chlorinated polyether
- Epoxy (EP)
- Ethylene vinyl acetate (EVA)
- Ethylene vinyl alcohol (EVOH)
- Fluorocarbons
  - Fluorinated ethylene propylene (FEP)
  - Polytetrafluoroethylene (FTFE)
  - Polyvinyl fluoride (PVF)
  - Polyvinylidene fluoride (PVDF)
- Ionomer
- Liquid crystal polymer (LCP)
  - Aromatic copolyester (TP polyester)
- Melamine formaldehyde (MF)
- Nylon (or polyamide) (PA)
- Parylene phenolic
- Phenolic
  - Phenol formaldehyde (PF)
- Polyamide-imide (PAI)
- Polyarylether
  - Polyaryletherketone (PAEK)
  - Polyaryl sulfone (PAS)
- Polyarylate (PAR)
- Polycarbonate
- Polyester
  - Saturated polyester (TS polyester)
  - Thermoplastic polyesters (TP polyester)
    - Polybutylene terephthalate (PBT)
    - Polyethylene terephthalate (PET)
- Polyetherketone (PEK)
- Polyetheretherketone (PEEK)
- Polyetherimide (PEI)
- Polyimide (PI)
  - Thermoplastic PI
  - Thermoset PI
- Polymethylmethacrylate (or acrylic) (PMMA)
- Polyolefin (PO)
  - Chlorinated PE (CPE)
  - Cross-linked PE (XLPE)
  - High-density PE (HDPE)
  - Linear LDPE (LLDPE)
  - Low-density PE (LDPE)
  - Polyallomer
  - Polybutylene (PE)
  - Polyethylene (PE)
  - Polypropylene (PP)
  - Ultra-high-molecular-weight PE (UHMWPE)
- Polyurethane (PUR)
- Silicone (SI)
- Styrene
  - Acrylic styrene acrylonitrile (ASA)
  - Acrylonitrile butadiene styrene (ABS)
  - General-purpose PS (GPPS)
  - High-irnpact PS (HIPS)
  - Polystyrene (PS)
  - Styrene acrylonitrile (SAN)
  - Styrene butadiene (SB)
- Sulfone
  - Polyether sulfone (PES)
  - Polyphenyl sulfone (PPS)
  - Polysulfone (PSU)
- Urea formaldehyde (UF)
- Vinyl
  - Chlorinated PVC (CPVC)
  - Polyvinyl acetate (PVAc)
  - Polyvinyl alcohol (PVA)
  - Polyvinyl butyrate (PVB)
  - Polyvinyl chloride (PVC)
  - Polyvinylidene chloride (PVDC)
  - Polyvinylidene fluoride (PVF)

different processing and performance behaviors. These basic types, with their many modifications of different additives/fillers/reinforcements, catalyst systems, grafting, and/or alloying provide different processing capabilities and/or product performances. As examples there are the relatively new generation of high-performance metallocene and elastomeric plastics providing different modifications.

11.1.2.1 Thermoplastics

TPs are plastics that soften when heated and upon cooling harden into products. TPs can be repeatedly softened by reheating. Their morphology, molecular structure, is crystalline or amorphous. Softening temperatures vary. The usual analogy is a block of ice that can be softened (turned back to a liquid), poured into any shape mold or die, then cooled to become a solid again. This cycle repeats. During the heating cycle care must be taken to avoid degrading or decomposition of the plastic. TPs generally offer easier processing and better adaptability to complex designs than do TS plastics.

There are practical limits to the number of heating and cooling cycles before appearance and/or mechanical properties are drastically affected. Certain TPs have no immediate changes while others have immediate changes after the first heating/cooling cycle.

11.1.2.1.1 Crystalline and amorphous polymers

The overall molecular physical structure of a polymer identifies its morphology. Crystalline molecular structures tend to have their molecules arranged in a relatively regular repeating structure such as acetal (POM), polyethylene (PE), PP, nylon (PA), and polytetrafluoroethylene plastics. The structures tend to form like cooked spaghetti. These crystallized plastics have excellent chemical resistance. They are usually translucent or opaque but they can be made transparent with chemical modification. They generally have higher strength and

softening points and require closer temperature/time processing control than the amorphous TPs.

Polymer molecules that can be packed closely together can more easily form crystalline structures in which the molecules align themselves in some orderly pattern. Commercially crystalline polymers have up to 80% crystalline structure and the rest is amorphous. They are identified technically as semicrystalline TPs. Polymers with 100% crystalline structures are not commercially produced.

Amorphous TPs have no crystalline structure. Their molecules form no patterns. These TPs have no sharp melting points. They are usually glassy and transparent, such as acrylontrile-butadiene-styrene (ABS), acrylic (PMMA), polycarbonate (PC), PS, and PVC. Amorphous plastics soften gradually as they are heated during processing. If they are rigid, they may become brittle unless modified with certain additives.

During processing, all plastics are normally in the amorphous state with no definite order of molecular chains. TPs that normally crystallize need to be properly quenched; that is, the hot melt is cooled to solidify the plastic. If not properly quenched, they become amorphous or partially amorphous solids, usually resulting in inferior properties. Compared to crystalline types, amorphous polymers undergo only small volumetric changes when melting or solidifying during processing. This action influences the dimensional tolerances that can be met after accounting for the heating/cooling process and the design of molds or dies.

Crystalline plastics require tighter process control during fabrication. They tend to shrink and warp more than amorphous types due to their higher melting temperatures, with their relatively sharp melting point, they do not soften gradually with increasing temperature but remain hard until a given quantity of heat has been absorbed, then change rapidly into a low-viscosity liquid. If the correct amount of heat is not applied properly during processing, product performance can be drastically reduced and/or an increase in processing cost occurs. With proper process control this is not a problem.

During the melting process as the symmetrical molecules approach each other within a critical distance, crystals begin to form. They form first in the areas where they are the most densely packed. This crystallized area becomes stiff and strong. The noncrystallized, amorphous, area is tougher and more flexible. With increased crystallinity, other effects occur such as with PE (crystalline plastic) there is increased resistance to creep.

11.1.2.1.2 Liquid crystalline polymers

A special classification of TPs are liquid crystalline polymers. They are self-reinforcing because of densely packed fibrous polymer chains. Their molecules are stiff, rod-like structures organized in large parallel arrays in both the melt and solid states. They resist most chemicals, weathers oxidation, and can provide flame resistance, making them excellent replacements for metals, ceramics, and other plastics in many product designs. They are exceptionally inert and resist stress cracking in the presence of most chemicals at elevated temperatures, including the aromatic and halogenated hydrocarbons as well as strong acids, bases, ketones, and other aggressive industrial products. Regarding flammability, LCPs have an oxygen index ranging from 35% to 50% (ASTM). When exposed to an open flame, they form an intumescent char that prevents dripping.

When injection molded or extruded the molecules align into long, rigid chains that in turn align in the direction of flow. Thus the molecules act like reinforcing fibers giving LCPs both very high strength and stiffness. LPCs with their high strength-to-weight ratios are particularly useful for weight-sensitive products (Table 11.1.3). They have outstanding strength at extreme temperatures, excellent mechanical property retention after exposure to weathering and radiation, good dielectric strength as well as arc resistance and dimensional stability, low coefficient of thermal expansion, excellent flame resistance, and easy processability.

Table 11.1.3 Liquid crystal polymer properties compared to other TPs

Property	Crystalline	Amorphous	Liquid crystalline
Specific gravity	Higher	Lower	Higher
Tensile strength	Higher	Lower	Highest
Tensile modulus	Higher	Lower	Highest
Ductility, elongation	Lower	Higher	Lowest
Resistance to creep	Higher	Lower	High
Maximum usage temperature	Higher	Lower	High
Shrinkage and warpage	Higher	Lower	Lowest
Chemical resistance	Higher	Lower	Highest

Their UL (Underwriters Laboratory) continuous-use rating for electrical properties is as high as 240 °C (464 °F), and for mechanical properties it is 220 °C (428 °F) permitting products to be exposed to intermittent temperatures as high as 315 °C (600 °F) without affecting performance properties. Their resistance to high-temperature flexural creep is excellent, as are their fracture–toughness characteristics.

Because of their structure they provide special properties such as greater resistance to most solvents and heat. They have the lowest warpage and shrinkage of all the TPs. Unlike many high-temperature plastics, LCPs have a low melt viscosity and are thus more easily processed resulting in faster cycle times than those with a high melt viscosity thus reducing processing costs.

11.1.2.2 Thermosets

Outstanding properties of TS plastic products are their substantially infusible and insoluble characteristic along with resistance to high temperatures, greater dimensional stability, and strength. TSs undergo a cross-linking chemical reaction by techniques such as the action of heat (exothermic reaction), oxidation, radiation, and/or other means often in the presence of curing agents and catalysts. However, if excessive heat is applied, degradation rather than melting will occur.

TSs are not recyclable because they do not melt when reheated, although they can be granulated and used as filler in other TSs as well as TPs. An analogy of TSs is that of a hard-boiled egg that has turned from a liquid to a solid and cannot be converted back to a liquid. As shown in Figure 11.1.4, TSs are identified by A–B–C stages during the curing process. A-stage is uncured, B-stage is partially cured, and C-stage is fully cured. Typical B-stage is TS molding compounds and prepregs, which in turn are processed to produce C-stage fully, cured plastic material products.

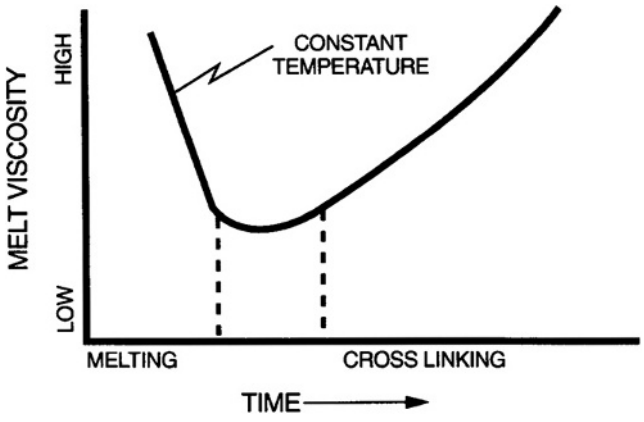


Fig. 11.1.4 TS A–B–C stages from melt to solidification.

TSs generally cannot be used alone in primary or secondary structural applications; they must be filled with additives and/or reinforcements such as glass or wood fibers, etc. These compounds provide dimensional product precision and certain other desirable properties for use in certain products. There are TSs particularly suitable as substitutes for metals in products that have to meet severe demands such as high temperature with the added advantage of offering a very good cost reduction. Applications include kitchen appliances, heat-shield for an electric iron, collectors and a wide variety of circuit breaker housings in electrical devices, and automotive parts including headlamp reflectors, brake servo units, brake pistons, pump housings, valve caps, pulleys, and so on. Compression and transfer molding (CM and TM) are the two main methods used to produce molded products from TSs.

Within the TS family there are natural and synthetic rubbers, elastomers, such as styrene–butadiene, nitrile rubber, millable polyurethanes, silicone, butyl, and neoprene. They attain their properties through the process of vulcanization. Vulcanization is the process by which a natural rubber or certain plastic elastomer undergoes a change in its chemical structure brought about by the irreversible process of reacting the materials with sulfur and/or other suitable agents. The cross-linking action results in property changes such as decreased plastic flow, reduced surface tackiness, increased elasticity, greater tensile strength, and considerably less solubility.

11.1.2.3 Cross-linked TPs

TPs can be converted to TSs to improve or change properties. TPs can be cross-linked by different processes such as chemical and irradiation. PE is a popular plastic that can be cross-linked; it is identified as XLPE or PEX. Cross-linking is an irreversible change that occurs through a chemical reaction, such as condensation, ring closure, addition, and so on. Cure is usually accomplished by the addition of curing (cross-linking) agents with or without heat and pressure.

For TP systems such as PE, chemical or irradiation techniques have been used as the cross-linking technology; this is the recognized standard for manufacturing industrial materials such as cable coverings, cellular materials (foams), rotationally molded articles, and piping.

Enhancement of properties is the underlying incentive for the commercial development of cross-linked TPs. Cross-linking improves resistance to thermal degradation, cracking by liquids and other harsh environments,

and creep and cold flow, among other improvements. The primary commercial interest has been in aliphatic polymers, which includes the main olefins PE and PP, also popular are PVC and acrylates. Cross-linked films with low shrinkage and high adhesion properties have been used in such applications as pressure-sensitive adhesives, glass coatings, and dental enamels.

11.1.2.4 Reinforced plastics

The term reinforced plastic (RP) refers to composite combinations of plastic, matrix, and reinforcing materials, which predominantly come in chopped and continuous fiber forms as in woven and nonwoven fabrics. Other terms used to identify an RP include: glass fiber reinforced plastic, aramid fiber reinforced plastic, boron fiber reinforced plastic, carbon fiber reinforced plastic, graphite fiber reinforced plastic, etc.

In addition to fabrics, reinforcements include other forms such as powders, beads, and flakes. Both TP and TS plastics are used in RPs. At least 90 wt% use glass fiber materials. At least 55 wt% use TPs. RPs using primarily TS polyester plastics provides significant property and/or cost improvements compared to other composites. Primary benefits of all RPs include high strength, directional strength, lightweight, high strength-to-weight ratio, creep and fatigue endurance, high dielectric strength, corrosion resistance, and long-term durability.

Both reinforced TSs (RTSs) and reinforced TPs (RTPs) can be characterized as engineering plastics, competing with engineering unreinforced TPs. When comparing processability of RTSs and RTPs, the RTPs are usually easier to process and permit faster molding cycles with efficient processing such as during injection molding (IM). Higher performing fibers that are used include high-performance glass (other than the usual E-glass), aramid, carbon, and graphite. Also available are whisker reinforcements with exceptional high performances (Figure 11.1.5).

Fiber orientations have improved to the extent that 2-D and 3-D RPs can be used to produce very high strength and stiff products with long service lives. RTPs even with their relatively lower properties compared to RTSs are used in about 55 wt% of all RP products. Practically all RTPs with short glass fibers are injection molded at very fast cycles, producing high-performance products in highly automated environments.

RPs can be characterized by their ability to be molded into either extremely small to extremely large structurally loaded shapes well beyond the basic capabilities of other materials or processes at little or no pressure. In addition to shape and size, RPs possess other characteristics that make them very desirable in design engineering. The other characteristics include cost reduction, ease of fabrication, simplified installation, weight reduction, aesthetic appeal, and the potential to be combined with many other useful qualities.

Their products have gone worldwide into the deep ocean waters, on land, and into the air including landing on the moon and in spacecraft. In USA annual consumption of all forms of RPs is over 3.9 billion lb (1.8 billion kg). Consumption by market in million lb is aerospace at 24, appliances/business machines at 210, building/construction at 775, consumer at 253, corrosion at 442, electrical/electronic at 390, marine at 422, transport at 1268, and others at 116.

The form the RP takes, as with non-RPs, is determined by the product requirements. It has no inherent

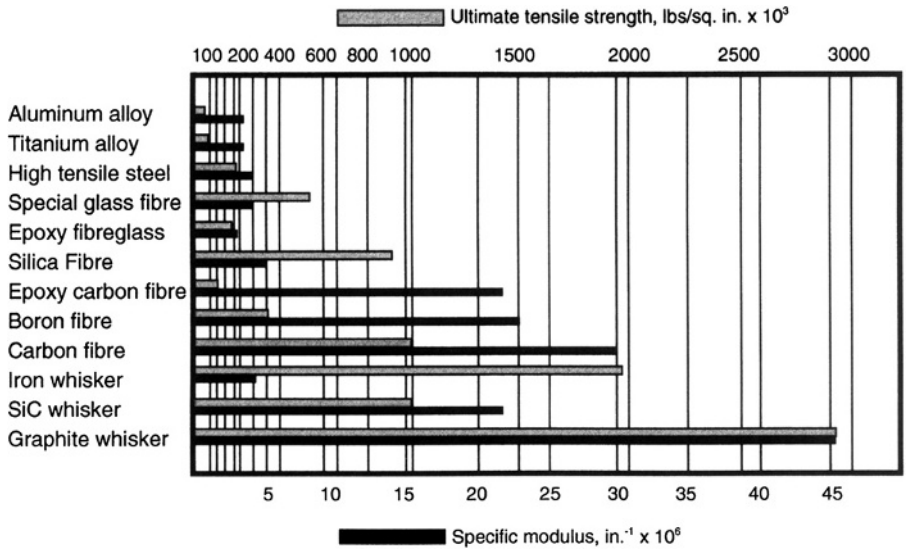


Fig. 11.1.5 High-performance whisker reinforcements compared to other materials.

form of its own; it must be shaped. This provides an opportunity to select the most efficient forms for the application. Shape can help to overcome limitations that may exist in using a lower-cost material with low stiffness. As an example underground fuel tanks can include ribs to provide added strength and stiffness to the RP orientation in order to meet required stresses at the lowest production cost.

The formability of these products usually leads to one-piece consolidation of construction products to eliminate joints, fasteners, seals, and other potential joining problems. As an example, formed building fascia panels eliminate many fastenings and seals. Examples of design characteristics gained by using RP materials are presented in the following.

11.1.2.4.1 Thermal expansions

Non-RPs generally have much higher coefficients of linear thermal expansion (CLTE) than conventional metal, wood, concrete, and other materials. CLTEs also vary significantly with temperature changes. There are RPs that do not have these characteristics. With certain types and forms of fillers, such as graphite, RPs can eliminate CLTE or actually shrink when the temperature increases.

11.1.2.4.2 Ductilities

Substantial yielding can occur in response to loading beyond the limit of approximate proportionality of stress to strain. This action is referred to as ductility. Most RPs do not exhibit such behavior. However, the absence of ductility does not necessarily result in brittleness or lack of flexibility. For example, glass fiber-TS polyester RPs do not exhibit ductility in their stress–strain behavior, yet they are not brittle, have good flexibility, and do not shatter upon impact. TS plastic matrix is brittle when unreinforced. However, with the addition of glass or other fibers in any orientation except parallel, unidirectional, the fibers arrest crack propagation. This RP construction results in toughness and the ability to absorb a high amount of energy. Because of the generally high ratio of strength to stiffness of RPs, energy absorption is accomplished by high elastic deflection prior to failure. Thus ductility has been a major factor promoting the use of RPs in many different applications.

11.1.2.4.3 Toughness

The generally low-specific gravity and high strength of reinforcement fibers such as glass, aramid, carbon, and graphite can provide additional benefits of toughness. For example, the toughness of these fibers allows them to be molded into very thin constructions. Each fiber has special characteristics. For instance, compared to other fiber reinforcements, aramid fibers can increase wear resistance with exceptionally high strength or modulus to weight.

11.1.2.4.4 Tolerances/shrinkages

TSs combined with all types of reinforcements and/or fillers are generally more suitable for meeting tight dimensional tolerances than are TPs. For injection molded products they can be held to extremely close tolerances of less than a thousandth of an inch (0.0025 cm) effectively down to zero (0.0%). Achievable tolerances range from 5% for 0.020 in. (0.05 cm), to 1% for 0.500 in. (1.27 cm), to 1/2% for 1.000 in. (2.54 cm), to 1/4% for 5.000 in. (12.70 cm), and so on.

Some unreinforced molded plastics change dimensions, shrink, immediately after molding or in a day or a month due to material relaxation and changes in temperature, humidity, and/or load application. RPs can significantly reduce or even eliminate this dimensional change after molding.

When comparing tolerances and shrinkage behaviors of RTSs and RTPs, there is a significant difference. Working with crystalline RTPs can be yet more complicated if the fabricator does not understand their behavior. Crystalline plastics generally have different rates of shrinkage in the longitudinal, melt flow direction, and transverse directions. In turn, these directional shrinkages can vary significantly due to changes in processes such as during IM. Tolerance and shrinkage behaviors are influenced by factors such as injection pressure, melt heat, mold heat, and part thickness and shape. The amorphous type materials can be easier to balance.

11.1.2.4.5 Compounds

Commercial RP compounds are available in several forms: pellets for IM or extrusion, unidirectional tape for filament winding and similar applications, sheets for stamping and compression molding, bulk compounds for compression molding, and so on. There are RTP elastomeric materials that provide special engineered products such as conveyor belts, mechanical belts, high temperature or chemical resistant suits, wire and cable insulation, and architectural designed shapes. Common categories of RP compounds are reviewed.

11.1.2.4.6 Prepregs

Preimpregnated materials usually are a compound of a reinforcement and a hot melt or solvent system. Prepreg also includes wet systems without solvent using TS polyester. They are stored for use at a latter time either in-house or to ship to a fabricator. The plastic is partially cured, B-stage, ready-to-mold material in web

form that may have a substrate of glass fiber mat, fabric, roving, paper, cotton cloth, and so forth. With proper temperature storage conditions, their shelf life can be controlled to last at least 6 months.

11.1.2.4.7 Sheet molding compounds

A ready-to-mold material, SMC represent a special form of a prepreg. It is usually a glass fiber RTS polyester resin compound in sheet form. The sheet can be rolled into coils during its continuous fabricating process. A plastic film covering, usually PE, separates the layers to enable coiling and to prevent contamination, sticking, and monomer evaporation. This film is removed before the SMC is charged into a mold, such as a matched-die or compression mold.

Depending on product performance requirements, the SMC consists of additional ingredients such as low-profile additives, cure initiators, thickeners, and mold-release agents. They are used to enhance the performance or processing of the material. Glass fibers are usually chopped into lengths of 12 mm (0.5 in.) to at least 50 mm (2 in.). The amount can vary from 25 to 50 wt%. The usual ratio is based on performance requirements, processability, and cost considerations.

11.1.2.4.8 Bulk molding compounds

Also called dough molding compounds, bulk molding compounds (BMCs) are mixtures usually of short 3 mm to 3 cm (1/8–1¼ in.) glass fibers, plastic, and additives similar to the SMC compound. This mixture, with the consistency of modeling clay, can be produced in bulk form or extruded in rope-like form for easy handling. The extrudate type is called a "log" that is cut to specific lengths such as 0.3 cm (1 ft).

BMC is commercially available in different combinations of resins, predominantly TS polyesters, additives, and reinforcements. They meet a wide variety of end-use requirements in high-volume applications where fine finish, good dimensional stability, part complexity, and good overall mechanical properties are important. The most popular method of molding BMCs is compression. They can also be injection molded in much the same way as other RTS compounds using ram, ram-screw, and, for certain BMC mixes, conventional reciprocating screw.

11.1.2.5 Commodity and engineering plastics

About 90 wt% of plastics can be classified as commodity plastics (CPs), the others being engineering plastics (EPs). The EPs such as PC representing at least 50 wt% of all EPs, nylon, acetal, etc. are characterized by improved performance in higher mechanical properties, better heat resistance, and so forth (Table 11.1.4).

Table 11.1.4 Thermoplastic engineering behaviors

Crystalline	Amorphous
Acetol	*Polycarbonate*
Best property balance	Good impact resistance
Stiffest unreinforced thermoplastic	Transparent
Low friction	Good electrical properties
Nylon	*Modified PPO*
High melting point	Hydrolytic stability
High elongation	Good impact resistance
Toughest thermoplastic	Good electrical properties
Absorbs moisture	
Polyester (glass-reinforced)	
High stiffness	
Lowest creep	
Excellent electrical properties	

The EPs demand a higher price. About a half century ago the price per pound was at 20¢; at the turn of the century it went to $1.00, and now higher. When CPs with certain reinforcements and/or alloys with other plastics are prepared they become EPs. Many TSs and RPs are EPs.

11.1.2.6 Elastomers/rubbers

In the past rubber meant a natural TS elastomeric (TSE) material obtained from a rubber tree, *Hevea braziliensis*. The term elastomer developed with the advent of rubber-like synthetic materials. Elastomers identify natural or synthetic TS elastomers (TSEs) and thermoplastic elastomers (TPEs). At room temperature all elastomers basically stretch under low stress to at least twice in length and snaps back to approximately the original length on release of the stress, pull, within a specified time period.

The term elastomer is often used interchangeably with the term plastic or rubber; however, certain industries use only one or the other terminology. Different properties identify them such as strength and stiffness, abrasion resistance, oil resistance, chemical resistance, shock and vibration control, electrical and thermal insulation, waterproofing, tear resistance, cost-to-performance, etc.

Natural rubber with over a century's use in many different products and markets will always be required to attain certain desired properties not equaled (to date) by synthetic elastomers. Examples include transportation tires, with their relative heat build-up resistance, and certain types of vibrators. However, both synthetic TSE and TPE have made major inroads in product markets previously held only by natural rubber. Worldwide, more synthetic types are used than natural. The basic processing types are conventional, vulcanizable, elastomer, reactive type, and TPE.

11.1.3 Plastic behaviors

A knowledge of the chemistry of plastics can be used to help with the understanding of the performance of designed products. Chemistry is the science that deals with the composition, structure, properties, and transformations of substances. It provides the theory of organic chemistry, in particular our understanding of the mechanisms of reactions of carbon (C) compounds.

The chemical composition of plastics is basically organic polymers. They have very large molecules composed of connecting chains of carbon (C), generally connected to hydrogen atoms (H) and often also oxygen (O), nitrogen (N), chlorine (Cl), fluorine (F), and sulfur (S). Thus, while polymers form the structural backbone of plastics, they are rarely used in pure form. In almost all plastics other useful and important materials are added to modify and optimize properties for each desired process and/or product performance application.

The chemical and physical characteristics of plastics are derived from the following factors of chemical structure: form, arrangement, and size of the polymer. As an example, the chemical structure influences density. Chemical structure refers to the types of atoms and the way they are joined to one another. The form of the molecules, their size and disposition within the material, influences mechanical behavior. It is possible to deliberately vary the crystal state in order to vary hardness or softness, toughness or brittleness, resistance to temperature, and so on. The chemical structure and nature of plastics have a significant relationship both to properties and the ways they can be processed, designed, or otherwise translated into a finished product.

11.1.3.1 Morphology/molecular structure/mechanical property

Morphology is the study of the physical form or chemical structure of a material; that is, the physical molecular structure. As a result of morphology differences among polymers, great differences exist in mechanical and other properties as well as processing plastics.

Knowledge of molecular size and flexibility explains how individual molecules behave when completely isolated. However, such isolated molecules are encountered only in theoretical studies of dilute solutions. In practice, molecules always occur in a mass, and the behavior of each individual molecule is very greatly affected by its intermolecular relationships to adjacent molecules in the mass. Three basic molecular properties affect processing performances, such as flow conditions, that in turn affect product performances, such as strength or dimensional stability. They are (1) mass or density, (2) MW, and (3) molecular weight distribution (MWD).

11.1.3.1.1 Densities

Absolute density (d) is the mass of any substance per unit volume of a material. It is usually expressed in grams per cubic centimeter (g/cm^3) or pounds per cubic inch (lb/in^3) (Table 11.1.5). Specific gravity (s.g.) is the ratio of the mass in air of a given volume compared to the mass of the same volume of water. Both d and s.g. are measured at room temperature [23 °C (73.4 °F)]. Since s.g. is a dimensionless quantity, it is convenient for comparing different materials. Like density, specific gravity is used extensively in determining product cost vs. average product thickness, product weight, quality control, and so on. It is frequently used as a means of setting plastic specifications and monitoring product consistency.

In crystalline plastics, density has a direct effect on properties such as stiffness and permeability to gases and liquids. Changes in density may also affect other mechanical properties.

The term *apparent density* of a material is sometimes used. It is the weight in air of a unit volume of material including voids usually inherent in the material. Also used is the term *bulk density* that is commonly used for compounds or materials such as molding powders, pellets, or flakes. Bulk density is the ratio of the weight of

Table 11.1.5 Comparing densities of different polyethylene thermoplastics

Type	Density, g/cm^3 ($/b/ft^3$)
LDPE	0.910–0.925 (56.8–57.7)
MDPE	0.926–0.940 (57.8–58.7)
HDPE	0.941–0.959 (58.7–59.9)
HMWPE	0.960 & above (59.9 & above)

the compound to its volume of a solid material including voids.

11.1.3.1.2 Molecular weights

MW is the sum of the atomic weights of all the atoms in a molecule. Atomic weight is the relative mass of an atom of any element based on a scale in which a specific carbon atom (carbon-12) is assigned a mass value of 12. For polymers, it represents a measure of the molecular chain length. MW of plastics influences their properties. With increasing MW, polymer properties increase for abrasion resistance, brittleness, chemical resistance, elongation, hardness, melt viscosity, tensile strength, modulus, toughness, and yield strength. Decreases occur for adhesion, melt index (MI), and solubility.

Adequate MW is a fundamental requirement to achieve desired properties of plastics. If the MW of incoming material varies, the fabricating and fabricated product performance can be altered. The greater the differences, the more dramatic the changes that occur during processing.

11.1.3.1.3 Molecular weight distributions

MWD is basically the amounts of component polymers that make up a polymer (Figure 11.1.6). Component polymers, in contrast, arc a convenient term that recognizes the fact that all polymeric materials comprise a mixture of different polymers of differing MWs. The ratio of the weight average MW to the number average MW gives an indication of the MWD.

One method of comparing the processability with product performances of plastics is to use their MWD. A narrow MWD enhances the performance of plastic products. Wide MWD permits easier processing. Melt flow rates are dependent on the MWD. With MWD differences of incoming material the fabricated performances can be altered requiring resetting process controls. The more the difference, the more dramatic changes that can occur in the products.

11.1.3.2 Viscosities and melt flows

Viscosity is a measure of resistance to plastic melt flow. It is the internal friction in a melt resulting when one layer of fluid is caused to move in relationship to another layer. Thus viscosity is the property of the resistance of flow exhibited within a body of material. It is the constant ratio of shearing stress to the rate of shear. Shearing is the motion of a fluid, layer by layer, like playing cards in a deck. When plastics flow through straight tubes or channels they are sheared: the viscosity expresses their resistance.

The MI or melt flow index (MFI) is an inverse measure of viscosity. High MI implies low viscosity and low MI means high viscosity. Plastics are shear thinning, which means that their resistance to flow decreases as the shear rate increases. This is due to molecular alignments in the direction of flow and disentanglements.

11.1.3.3 Newtonian/non-Newtonian

Viscosity is usually understood to mean Newtonian viscosity in which case the ratio of shearing stress to the shearing strain is constant. In non-Newtonian behavior, typical of plastics, the ratio varies with the shearing stress. Such ratios are often called the *apparent viscosities* at the corresponding shearing stresses. Viscosity is measured in terms of flow in Pa·s (P) with water as the base standard (value of 1.0). The higher the number, the less flow.

11.1.3.3.1 Melt index

The melt indexer (extrusion plastometer) is the most widely used rheological device for examining and studying plastics (principally TPs) in many different

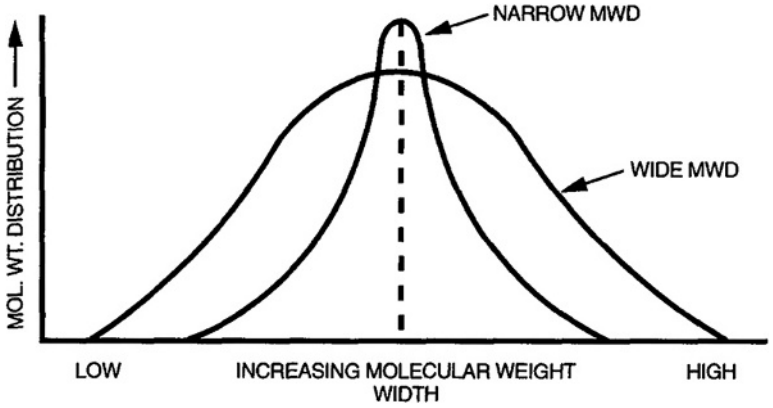


Fig. 11.1.6 Examples of narrow and wide MWDs.

Table 11.1.6 Examples of melt index for different processes

Process	MI range
Injection molding	5–100
Rotational molding	5–20
Coating extrusion	0.1–1
Film extrusion	0.5–6
Profile extrusion	0.1–1
Blow molding	0.1–1

fabricating processes. It is not a true viscometer in the sense that a reliable value of viscosity cannot be calculated from the measured flow index. However, the device does measure isothermal resistance to flow, using standard apparatus and test methods that are standard throughout the world. The standards used include ASTM D 1238 (U.S.A.), BS 2782-105 °C (U.K.), DIN 53735 (Germany), JIS K72 IO (Japan), IS0 RI 133/R292 (international), and others.

The standard apparatus is a ram type plasticator which at specified temperatures and pressure extrudes a plastic melt through the die exit opening. The standard procedure involves the determination of the amount of plastic extruded in 10 min. The flow rate, expressed in g/10 min., is reported. As the flow rate increases, viscosity decreases. Depending on the flow behavior, changes are made to standard conditions (die opening size, temperature, etc.) to obtain certain repeatable and meaningful data applicable to a specific processing operation. Table 11.1.6 lists typical MI ranges for the certain processes.

11.1.3.4 Rheology and mechanical analysis

Rheology and mechanical analysis are usually familiar techniques, yet the exact tools and the far-reaching capabilities may not be so familiar. Rheology is the study of how materials flow and deform, or when testing solids it is called dynamic mechanical thermal analysis.

During rheometer and dynamic mechanical analyses instruments impose a deformation on a material and measure the material's response that gives a wealth of very important information about structure and performance of the basic polymer. As an example stress rheometers are used for testing melts in various temperature ranges. Strain controlled rheology is the ultimate in materials characterization with the ability to handle anything from light fluids to solid bars, films, and fibers.

With dynamic testing, the processed plastic's elastic modulus (relating to energy storage) and loss modulus (relative measure of a damping ability) are determined. Steady testing provides information about creep and recovery, viscosity, rate dependence, etc.

11.1.3.4.1 Viscoelasticities

Understanding and properly applying the following information to product design equations is very important. A material having this property is considered to combine the features of a so-called perfect elastic solid and a perfect fluid. It represents the combination of elastic and viscous behavior of plastics that is a phenomenon of time-dependent, in addition to elastic deformation (or recovery) in response to load.

This property possessed by all fabricated plastics to some degree, indicates that while plastics have solid-like characteristics such as elasticity, strength, and form or shape stability, they also have liquid-like characteristics such as flow depending on time, temperature, rate, and amount of loading. The mechanical behavior of these viscoelastic plastics is dominated by such phenomena as tensile strength, elongation at break, stiffness, rupture energy, creep, and fatigue which are often the controlling factors in a design.

11.1.4 Processing-to-performance interface

Different plastic characteristics influence processing and properties of plastic products. Important are glass transition temperature and melt temperature.

11.1.4.1 Glass transition temperatures

The T_g relates to temperature characteristics of plastics (Table 11.1.7). It is the reversible change in phase of a plastic from a viscous or rubbery state to a brittle glassy state (Figure 11.1.7). T_g is the point below which plastic behaves like glass and is very strong and rigid. Above this temperature it is not as strong or rigid as glass, but neither is it brittle as glass. At and above T_g the plastic's volume or length increases more rapidly, and rigidity and strength decrease. As shown in Figure 11.1.8 the amorphous TPs have a more definite T_g when compared to crystalline TPs. Even with variation it is usually reported as a single value.

The thermal properties of plastics, particularly its T_g, influence the plastic's processability performance and cost in different ways. The operating temperature of a TP is usually limited to below its T_g. A more expensive plastic could cost less to process because of its T_g location

Table 11.1.7 Range of T_g for different thermoplastics

Plastic	°C	°F
Polyethylene	−120	−184
Polypropylene	−22	−6
Polybutylene	−25	−13
Polystyrene	95	203
Polycarbonate	150	302
Polyvinyl chloride	85	185
Polyvinyl fluoride	−20	−4
Polyvinylidene chloride	−20	−4
Polyacetal	−80	−112
Nylon 6	50	122
Polyester	110	230
Polytetrafluoroethylene	−115	−175
Silicone	−120	−184

that results in a shorter processing time, requiring less energy for a particular weight, etc. (Figure 11.1.9).

The T_g generally occurs over a relatively narrow temperature span. Not only do hardness and brittleness undergo rapid changes in this temperature region, but other properties such as the coefficient of thermal expansion and specific heat also change rapidly. This phenomenon has been called second-order transition, rubber transition, or rubbery transition. The word transformation has also been used instead of transition. When more than one amorphous transition occurs in a plastic, the one associated with segmental motions of the plastic backbone chain, or accompanied by the largest change in properties, is usually considered to be the T_g.

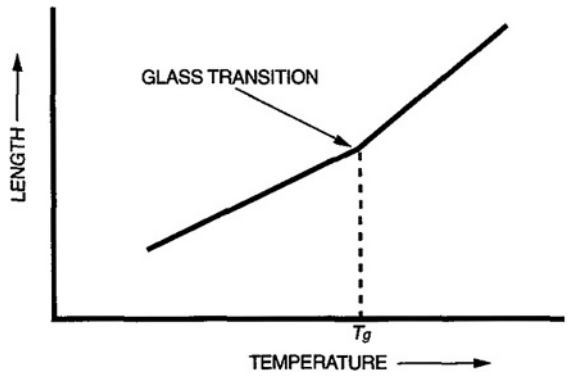


Fig. 11.1.7 TP volume or length changes at the glass transition temperature.

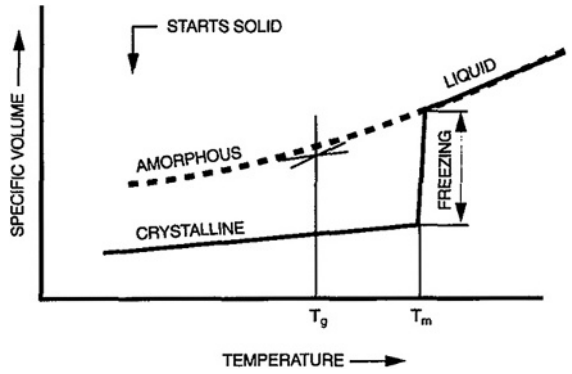


Fig. 11.1.8 Change of amorphous and crystalline TP's volume at T_g and T_m.

Important for designers to know that above T_g, many mechanical properties are reduced. Most noticeable is a reduction that can occur by a factor of 1000 in stiffness.

11.1.4.1.1 Melt temperatures

Crystalline plastics have specific melt temperatures (T_m) or melting points. Amorphous plastics do not. They have softening ranges that are small in volume when solidification of the melt occurs or when the solid softens and becomes a fluid type melt. They start softening as soon as the heat cycle begins. A melting temperature is reported usually representing the average in the softening range.

The T_m of crystalline plastics occurs at a relatively sharp point going from solid to melt. It is the temperature at which melts softens and begins to have flow tendency (Table 11.1.8). They have a true T_m with a latent heat of fusion associated with the melting and freezing process, and a relatively large-volume change during fabrication. Crystalline plastics have considerable order of the molecules in the solid state indicating that many of the atoms are regularly spaced. The melt strength of the plastic occurs while in the molten state. It

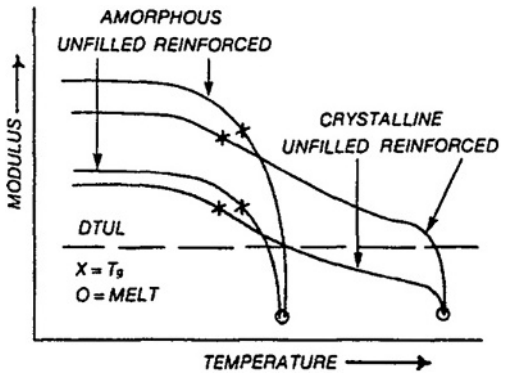


Fig. 11.1.9 Modules behavior with increase in temperature (DTUL, deflection temperature under load). (Courtesy of Bayer.)

Table 11.1.8 Crystalline thermoplastic melt temperatures

Plastic	°C	°F
Low-density polyethylene	116	240
High-density polyethylene	130	266
Polypropylene	175	347
Nylon 6	215	419
Nylon 66	260	500
Polyester	260	500
Polyarylamide	400	755
Polytetrafluoroethylene	330	626

is an engineering measure of the extensional viscosity and is defined as the maximum tension that can be applied to the melt without breaking.

The T_m is dependent on the processing pressure and the time under heat, particularly during a slow temperature change for relatively thick melts during processing. Also, if the melt temperature is too low, the melt's viscosity will be high and more costly power required processing it. If the viscosity is too high, degradation will occur. There is the correct processing window used for the different melting plastics.

11.1.4.2 Processing and moisture

Recognize that properties of designed products can vary, in fact can be destructive, with improper processing control such as melt temperature profile, pressure profile, and time in the melted stage. An important condition that influence properties is moisture contamination in the plastic to be processed. There are the hygroscopic plastics (PET, etc.) that are capable of retaining absorbed and adsorbed atmospheric moisture within the plastics. The non-hygroscopic plastics (PS, etc.) absorb moisture only on the surface. In the past when troubleshooting plastic's reduced performance was 90% of the time due to the damaging effect of moisture because it was improperly dried prior to processing. At the present time it could be at 50%.

All plastics, to some degree, are influenced by the amount of moisture or water they contain before processing. With minimal amounts in many plastics, mechanical, physical, electrical, aesthetic, and other properties may be affected, or may be of no consequence. However, there are certain plastics that, when compounded with certain additives such as color, could have devastating results. Day-to-night temperature changes is an example of how moisture contamination can be a source of problems if not adequately eliminated when plastic materials are exposed to the air. Moisture contamination can have an accumulative effect. The critical moisture content that is the average material moisture content at the end of the constant-rate drying period, is a function of material properties, the constant-rate of drying, and particle size.

Although it is sometimes possible to select a suitable drying method simply by evaluating variables such as humidities and temperatures when removing unbound moisture, many plastic drying processes do not involve removal of bound moisture retained in capillaries among fine particles or moisture actually dissolved in the plastic. Measuring drying-rate behavior under control conditions best identifies these mechanisms. A change in material handling method or any operating variable, such as heating rate, may effect mass transfer.

11.1.4.2.1 Drying operations

When drying at ambient temperature and 50% relative humidity, the vapor pressure of water outside a plastic is greater than within. Moisture migrates into the plastic, increasing its moisture content until a state of equilibrium exists inside and outside the plastic. But conditions are very different inside a drying hopper (etc.) with controlled environment. At a temperature of 170 °C (350 °F) and −40 °C (−40 °F) dew point, the vapor pressure of the water inside the plastic is much greater than the vapor pressure of the water in the surrounding area. Result is moisture migrates out of the plastic and into the surrounding air stream, where it is carried away to the desiccant bed of the dryer.

Target is to keep moisture content at a designated low level, particularly for hygroscopic plastics where moisture is collected internally. They have to be carefully dried prior to processing. Usually the moisture content is >0.02 wt%. In practice, a drying heat 30 °C below the softening heat has proved successful in preventing caking of the plastic in a dryer. Drying time varies in the range of 2–4 h, depending on moisture content. As a rule of thumb, the drying air should have a dew point of −34 °C (−30 °F) and the capability of being heated up to 121 °C (250 °F). It takes about 1 ft^3 mid^{-1} of plastic processed when using a desiccant dryer.

The non-hygroscopic plastics collect moisture only on the surface. Drying this surface moisture can be accomplished by simply passing warm air over the material. Moisture leaves the plastic in favor of the warm air resulting in dry air. The amount of water is limited or processing can be destructive.

Determine from the material supplier and/or experience the plastic's moisture content limit. Also important is to determine which procedure will be used in determining water content. They include equipment such

as weighing, drying, and/or reweighing. These procedures have definite limitations based on the plastic to be dried. Fast automatic analyzers, suitable for use with a wide variety of plastic systems, are available that provide quick and accurate data for obtaining the in-plant moisture control of plastics.

11.1.5 Fabricating processes

Designing good products requires some familiarity with processing methods. Until the designer becomes familiar with processing, a qualified fabricator must be taken into the designer's confidence early in development. The fabricator and mold or die designer should advise the product designer on materials behavior and how to simplify the design in order to simplify processing and reducing cost. Understanding only one process and in particular just a certain narrow aspect of it should not restrict the designer.

There are dozens of popular different basic processes with each having many modifications so that there are literally hundreds of processes used. The ways in which plastics can be processed into useful end products tend to be as varied as the plastics themselves. However only a few basic processes are used worldwide for most of the products produced. Extrusion consumes approximately 36 wt% of all plastics. IM follows by consuming 32 wt%. Consumption by other processes is estimated to be 10 wt% blow molding, 8% calendering, 5% coating, 3% compression molding, and others 3%. Thermoforming, which is the fourth major process used (considered a secondary process, since it begins with extruded sheets and films where extrusion is the primary process), consumes principally about 30% of the extruded sheet and film that principally goes into packaging.

It is estimated that there are in the USA about 80,000 injection molding machines (IMMs) and about 18,000 extruders operating. This difference in the amount of machines is due the fact that there is more activity (product design, R&D, fabrication, etc.) required with IM.

If an extruder can be used to produce products it has definite operating and economical advantages compared to IM. It requires detailed process control. IM requires more sophisticated process control to fabricate many thousands of different complex and intricate products.

While the processes differ, there are elements common to many of them. In the majority of cases, TP compounds in the form of pellets, granules, flake, and powder, are melted by heat so they can flow. Pressure is often involved in forcing the molten plastic into a mold cavity or through a die and cooling must be provided to allow the molten plastic to harden. With TSs, heat and pressure also are most often used, only in this case, higher heat (rather than cooling) serves to cure or harden the TS plastic, under pressure, in the mold. When liquid TP or TS plastics incorporate certain additives, heat and/or pressure need not necessarily be used.

Understanding, controlling, and measuring the plastic melt flow behavior of plastics during processing is important. It relates to a plastic that can be fabricated into a useful product. The target is to provide the necessary homogeneous uniformly heated melt during processing to have the melt operate completely stable and working in equilibrium. Unfortunately the perfect melt does not exist. Fortunately with the passing of time where improvements in the plastics and equipment uniformity continues to occur, melt consistency and melt flow behavior continues to improve, simplifying the art of processing.

An important factor for the processor is obtaining the best processing temperature for the plastics used. A guide is obtained from past experience and/or the material producer. The set-up person determines the best process control conditions (usually requires certain temperature, pressure, and time profiles) for the plastic being processed. Recognize that if the same plastic is used with a different machine (with identical operating specifications) the probability is that new control settings will be required for each machine. The reason is that, like the material, machines have variables that are controllable within certain limits that permit meeting the designed product requirements including costs.

The secondary operations fabricating methods can be divided into three broad categories: the machining of solid shapes; the cutting, sewing, and sealing of film and sheeting; and the forming of film and sheet. The machining techniques used are quite common to metal, wood, and other industries. Plastic shapes can be turned into end products by such methods as grinding, turning on a lathe, sawing, reaming, milling, routing, drilling, and tapping.

The cutting, sewing, and sealing of film and sheet involve turning plastic film and sheeting into finished articles like inflatable toys, garment bags, shower curtains, aprons, raincoat, luggage, and literally thousands of products. In making these products, the film or sheet is first cut to the desired pattern by hand, in die-cutting presses, or by other automatic methods. The pieces are then put together using such assembly techniques such as sewing, heat bonding, welding, high-frequency vibration, or ultrasonic sealing.

There are post-finished forming methods. Film and sheet can be post-embossed with textures and letterpress, gravure, or silk screening can print them. Rigid plastic parts can be painted or they can be given a metallic surface by such techniques as metallizing, barrel plating, or electroplating. Another popular method is hot-stamping, in which heat, pressure, and dwell time are

used to transfer color or design from a carrier film to the plastic part. Popular is the in-mold decorating that involves the incorporation of a printed foil into a plastic part during molding so that it becomes an integral part of the piece and is actually inside the part under the surface. There are applications, such as with blow molded products, where the foil provides structural integrity reducing the more costly amount of plastic to be used in the products.

11.1.5.1 Extrusions

Extrusion is the method employed to form TPs into continuous films, sheeting, tubes, rods, profile shapes, filaments, coatings (wire, cable, cord, etc.), etc. In extrusion, plastic material is first loaded into a hopper using upstream equipment, then fed into a long heating chamber through which it is moved by the action of a continuously revolving screw. At the end of this plasticator the molten plastic is forced through an orifice (opening) in a die with the relative shape desired in the finished product. As the extrudate (plastic melt) exits the die, it is fed downstream onto a pulling and cooling device such as multiple rotating rolls, conveyor belt with air blower, or water tank with puller.

The multi-screw extruders are used as well as the more popular single-screw extruders. Multi-screw extruders are primarily used for compounding plastic materials. Each has benefits primarily based on the plastic being processed and the products to be fabricated. At times their benefits can overlap, so the type to be used would depend on cost factors, such as cost to produce a quality product, cost of equipment, cost of maintenance, etc.

Size of the die orifice initially controls the thickness, width, and shape of any extruded product dimension. It is usually oversized to allow for the drawing and shrinkage that occur during conveyor pulling and cooling operations. The rate of takeoff also has significant influences on dimensions and shapes. This action, called drawdown, can also influence keeping the melt extrudate straight and properly shaped, as well as permitting size adjustments. Drawdown ratio is the ratio of orifice die size at the exit to the final product size.

Each of the processes (blown film, sheet, tube, etc.) contains secondary equipment applicable to their specific product lines such as computerized fluid chillers and temperature control systems. Equipment has become more energy-efficient, reliable, and cost-effective. The application of microprocessor and computer compatible controls that can communicate within the extruder line results in the more accurate control of the line.

A major part of film, sheet, coating, pipe, profile, etc. lines involve windup rolls. They include winders, dancer rolls, lip rolls, spreader rolls, textured rolls, engraved rolls, and cooling rolls. All have the common feature that they are required to be extremely precise in all their operations and measurements. Their surface conditions include commercial grade mirror finishes, precision bearings and journals are used, and, most important, controlled variable rotating speed controls to ensure uniform product tension control.

11.1.5.1.1 Orientations

Systems have been designed to increase the degree of orientation (stretching) in order to obtain films of improved clarity, strength, heat resistance, etc. Except for special applications, where greater strength in one direction may be needed, films are normally made with balanced properties.

11.1.5.1.2 Postformings

Various methods can be used for postforming products after the hot plastic melt leaves the extruder die. Examples are netting products that are flat to round shapes, rotated mandrel die makes perforated tubing, spiral spacer web around a coated wire or tube, varying tube or pipe wall thickness, and different perforated tubing or pipe pattern.

11.1.5.1.3 Coextrusions

There is the important variation on extrusion that involves the simultaneous or coextrusion of multiple molten layers of plastic from a single extrusion system. Two or more extruders are basically joined together by a common manifold through which melts flow before entering the die face. The plastics can include the same material but with different colors. There are also systems sometimes used where one material with two melts is made from one plasticator whereby certain advantages develop vs. the usual single melt such as reducing pin holes, and/or strengthening the product.

Many advantages exist in coextrusion. The different materials used in the coextruded structure meets different performance requirements based on their combinations. A single expensive plastic could be used to meet performance requirements such as permeability resistance, however with the proper combinations of plastics cost reductions will occur.

11.1.5.2 Injection moldings

The process of IM is used principally for processing unreinforced or glass fiber RTPs; however, it also processes TSs. Examples of the importance of using different mold design approaches are reviewed in Figure 11.1.10 concerning product openings and Figure 11.1.11 highlighting different ways with or

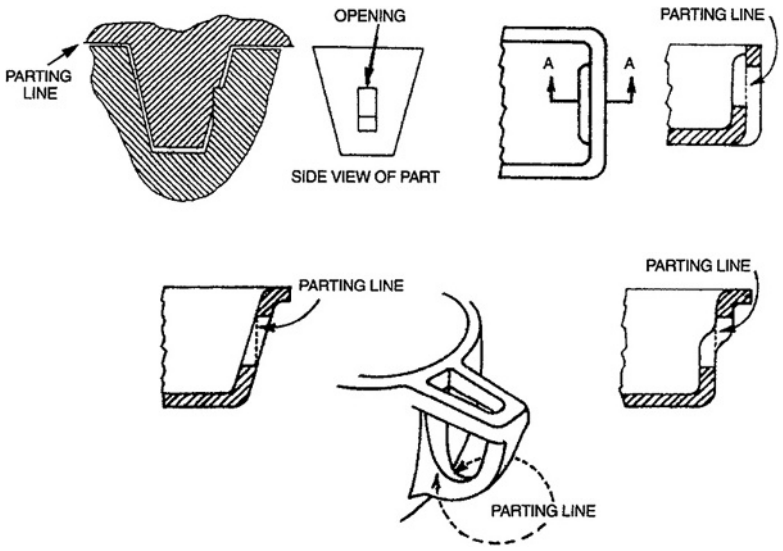


Fig. 11.1.10 Examples of simplifying mold construction to produce openings without side action movements.

without a parting line on the threads. These are examples that molds have to be properly designed to meet proper operations of product requirements. Where possible design of product shapes should make use of simplifying the design of molds.

The machines used for molding TSs are basically the same system as in molding TPs. Temperatures differ, as does the design of the screw. Unlike TPs that just melt in the plasticator and solidify in the cooled mold, the TSs melt in the plasticator and cure to a harden state in the mold that operates at a higher temperature than the plasticator.

11.1.5.2.1 Coinjections

The review in coextrusion also applies with coinjection providing similar advantages. Two or more IM barrels are basically joined together by a common manifold and nozzle through which melts flow before entering the mold cavity by a controlled device such as an open-closed valve system. The plastics can include the same material but with different colors. There are also systems sometimes used where one material with two shots is made from one plasticator whereby certain advantages develop vs. the usual single shot IM such as reducing pin holes, and/or strengthening the product. The nozzle is usually designed with a shutoff feature that allows only one melt to flow through at a controlled time.

The usual coinjection with two or more different plastics is bonded/laminated together. Proper melt flow and compatibility of the plastics is required in order to provide the proper adhesion. Some of the melt flow variable factors can be compensated by the available plasticator and mold process control adjustments.

11.1.5.2.2 Gas-assist moldings

There are different gas-assist injection molding (GAIM) processes. Other names exist that include injection molding gas-assist (IMGA), gas injection molding (GIM), or injection gas pressure (IGP). Most of the gas-assisted molding systems are patented. This review concerns the use of gas; however, there are others such as water-assist IM.

The processes use a gas that is usually nitrogen with pressures up to 20–30 MPa (2900–4400 psi). Within the mold cavity the gas in the melt forms channels. Gas pressure is maintained through the cooling cycle. In effect the gas packs the plastic against the cavity wall. Gas can be injected through the center of the IMM

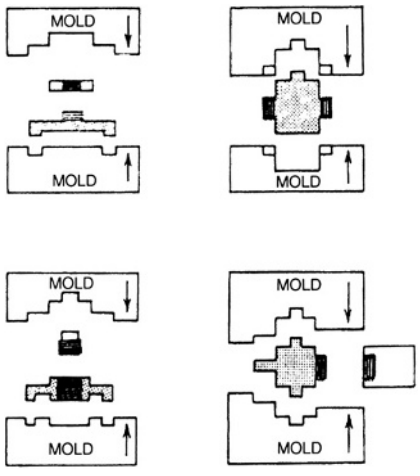


Fig. 11.1.11 Examples of molding with or without parting line on a product.

nozzle as the melt travels to the cavity or it can be injected separately into the mold cavity. In a properly designed tool run under the proper process conditions, the gas with its much lower viscosity than the melt remains isolated in the gas channels of the part without bleeding out into any thin-walled areas in the mold. The gas produces a balloon-like pressure on the melt.

The gas-assist approach is a solution to many problems associated with conventional IM and structural foam molding. It significantly reduces volume shrinkage that can cause sink marks in IM. Products are stiffer in bending and torsion than equivalent conventional IM products of the same weight. The process is very effective in different size and shape products, especially the larger molded products. It offers a way to mold products with only 10–15% of the clamp tonnage that would be necessary in conventional IM.

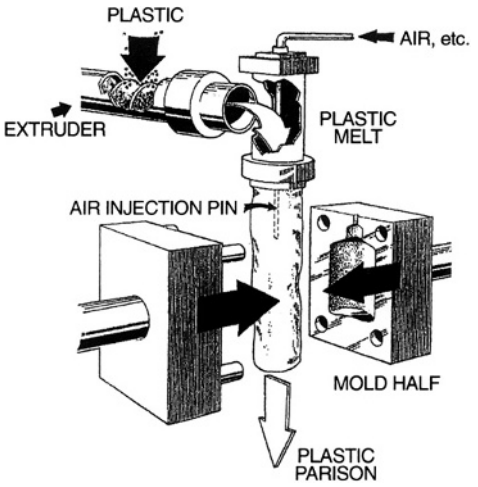


Fig. 11.1.12 Schematic of the extrusion BM process.

11.1.5.2.3 Micromoldings

As reviewed, the basic processes have many different fabricating systems. An example for IM is micromolding; precision molding of extremely small products as small as 1 mm^3. Products usually weigh less than 20 mg (0.020 g) with some even as low as 0.01 g. Products are measured in microns and have tolerances of ±10 microns or less. A micron (μm) is one-millionth of a meter; 25.4 μm make up one-thousandth of an inch. In comparison a human hair is 50–100 μm in diameter. A mil, that is about 25 times smaller than a micron, is one-thousandth of an inch.

Molding machines and tooling for small parts are not just smaller versions of their regular larger molding counterparts. Tooling is often created using electrical-discharge machining or diamond turning. It can be created with surface features below the wavelength of light by using lithographic and electrodeposition techniques. Proper venting usually has to include precision venting in the cavity as well as possibly removing air prior to entering the cavity.

11.1.5.2.4 Blow moldings

Generally used only with TPs, this process is applicable to the production of hollow plastic products such as bottles, gas tanks, and complex shaped containers/devices. The two basic systems to melt the TP are extruding (Figure 11.1.12) or IM (Figure 11.1.13). BM involves the melting of the TPs, then forming it into a tube-like or test tube shape (known as a parison when using an extruder or preform when IM), sealing the ends of the tube, and injecting air (through a tube or needle inserted in the tube or an opening in the preform core pin). The parison or preform, in a softened state, is inflated inside the mold and forced against the walls of the mold's female cavity. On cooling, the product, now conforming to the shape of the cavity, is solidified, and ejected from the mold as a finished piece. The coextrusion and coinjection already reviewed also applies to BM products.

11.1.5.2.5 Complex consolidated structural products

BM provides designers with the capability to make products ranging from the simple to rather complex 3-D shapes. Designers should become aware of the potentials BM offers since intricate and complex structural shapes can be fabricated. There are different techniques for BM these shapes (Figure 11.1.14). The techniques involve moving the preform or parison, moving the mold, or a combination of moving both the hot melt and mold.

BM permits combining in one product different parts or shapes that are to be assembled when using other processes. Result is simplifying the product design and significantly reducing cost. Some of the consolidating functions include hinges, inserts, fasteners, threads, nonplastic parts, and others somewhat similar to those used in IM. Hinges include the different mechanical types as well as integral hinges.

11.1.5.3 Thermoformings

Thermoforming consists of uniformly heating TP sheet or film to its softening temperature. Next the heated flexible plastic is forced against the contours of a mold. Force is applied by mechanical means (tools, plugs, solid molds, etc.) or by pneumatic means (differentials in air pressure created by pulling a vacuum between plastic and mold or using the pressures of compressed air to force the sheet against the mold).

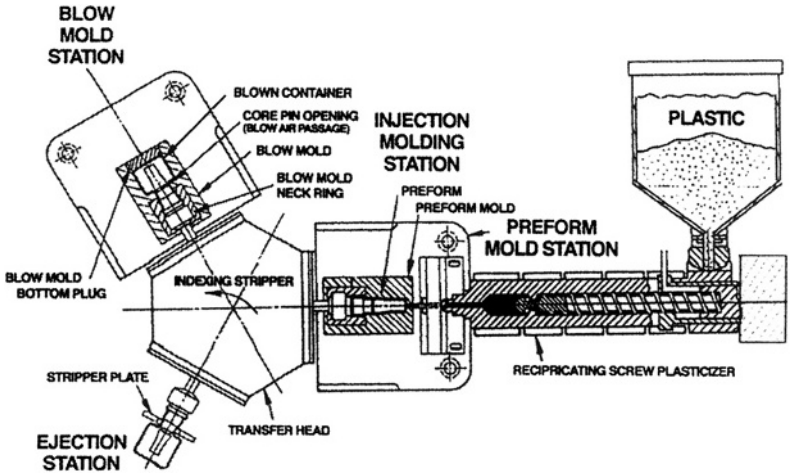


Fig. 11.1.13 Schematic of the injection BM process.

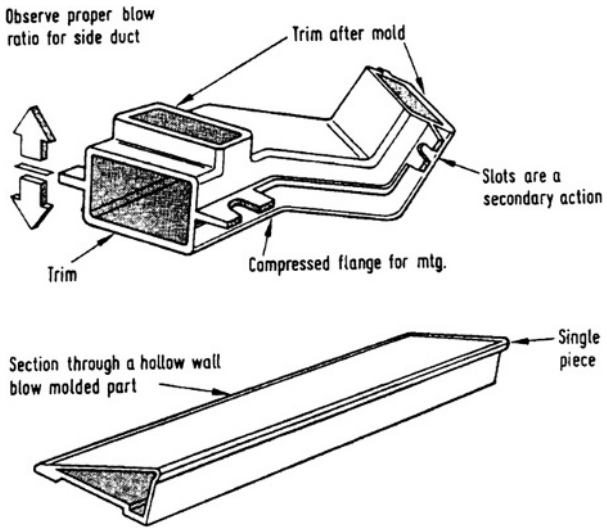


Fig. 11.1.14 Examples of complex BM products.

Almost any TP can be thermoformed. However certain types make it easier to meet certain forming requirements such as deep draws without tearing or excessive thinning in areas such as corners, and/or stabilizing of uniaxial or biaxial deformation stresses. Ease of thermoforming basically depends on stock material's thickness tolerance and forming characteristics. This ease of forming is influenced by factors such as to minimize the variation of the sheet thickness so that a uniform heat occurs in the film or sheet material thicknesswise, ability of the plastic to retain uniform and specific heat gradients across its surface and thickness, elimination or minimizing pin holes in the plastic, and stabilizing of uniaxial or biaxial deformation.

Many forming techniques are used. Each has different capabilities depending on factors such as formed product size, thickness, shape, type plastic, and/or quantity. Mold geometry with their different complex shapes vs. type of plastic material being processed will influence choice of process.

11.1.6 Foams

The manufacture of foam plastic products cuts across most of the processing techniques used. Foams can be fabricated during extrusion, IM, blow molding, casting, calendering, coating, rotational molding, etc. Typical requirements in such instances can be the incorporation of blowing agents in the plastic. They can be those that decompose under heat to generate the gasses needed to create the cellular structure. Various controls to accommodate the foaming action are used.

There are, however, some techniques unique to foamed plastics. When working with expanded polystyrene (EPS) beads, for example, to produce cups, picnic dishes, etc., various steam-chest molding methods are used. Based on the blowing agent used (pentane gas, etc.) the application of steam causes the beads to expand and fuse together in a perforated mold.

When working with polyurethane foams, it is possible to use spray guns or mixing metering machines to mix the liquid ingredients together and direct them into a product cavity, mold, etc. The mixed ingredients with their chemical reaction start to foam after leaving the dispensing equipment.

There is a unique technology of molding structural foam, foams with integral solid skins, and a cellular core resulting in a high strength-to-weight ratio. When processing structural foams, several techniques are used with most related to IM and extrusion.

11.1.7 Reinforced plastics

Different fabricating processes and materials of construction are employed to produce RP products that represent about 5 wt% of all plastic products produced worldwide. They range in fabricating pressures from zero (contact), through moderate, to relatively high pressure [14–207 MPa (2000–30,000 psi)], at temperatures based on the plastic's requirements that range from room temperature and higher. Equipment may be simple/low cost with labor costs high, to rather expensive specialized computer control sophisticated equipment with very low labor costs for the different processes. Each process provides capabilities such as meeting production quantity (small to large), performance requirements, proper ratio of reinforcement to matrix, fiber orientation, reliability/quality control, surface finish, and so forth versus cost (equipment, labor, utilities, etc.).

The plastic may be either RTSs or RTPs. The RTSs were the first major plastics to be adapted to this technology. The largest consumption of RTPs are processed by different methods such as IM (over 50 wt%), rotationally molding, or extruded on conventional equipment. There are even RTP sheets that can be "cold" stamped into shape using matching metal molds that form the products. It is called cold stamping because the molds are kept at or slightly above room temperature. The sheets, however, must be pre-heated.

11.1.7.1 Calenders

Calenders can be used to process TPs into film and sheeting, and to apply a plastic coating to textiles or other supporting/substraight materials. In calendering film and sheeting, the plastic compound is passed between a series of three or four large, heated, revolving rollers that squeeze the material between them into a sheet or film. An analogy in this case might be flattening out a pasty dough mixture with a rolling pin. The thickness of the finished material is controlled primarily by the space between the rolls. The surface of the plastic film or sheeting may be smooth or matted, depending on the surfacing on the rollers. When large quantities of particularly PVC film and sheet are to be manufactured, this process can provide lower-cost products than extrusion.

11.1.7.2 Castings

Casting may be used with TPs or TS plastics to make products, shapes, rods, tubes, film, sheet, etc. by basically pouring a liquid monomer–polymer solution into an open or closed mold where it finishes polymerizing and/or cooled into a solid. This liquid is often a monomer rather than the polymer used in most molding compounds. In turn the polymer with heat polymerizes into a solid plastic. An essential difference between casting and molding is that pressure need not be used in casting (although large-volume, complex parts can be made by low-pressure-casting methods). A variation on casting is known as liquid injection molding (LIM) and involves the proportioning, mixing, and dispensing of liquid components and directly injecting the resultant mix into a mold that is clamped under pressure.

11.1.7.3 Coatings

TPs or TS plastics may be used as a coating. The materials to be coated may be plastic, metal, wood, paper, fabric, leather, glass, concrete, ceramics, etc. Methods of coating are varied and include knife or spread coating, spraying, roller coating, dipping, brushing, and extrusion. Calendering of a film to a supporting material is also a form of coating.

Special methods can use powdered plastics for coatings. As an example the fluidized bed coating system. The object to be coated is heated and then immersed in a dense-phase fluidized bed of powdered plastic; the plastic adheres to the heated object and subsequent heating provides a smooth, pinhole-free coating. The electrostatic spray system is based on the fact that most plastic powders are insulators with relatively high-volume resistivity values. Therefore, they accept a charge (positive or negative polarity) and are attracted to a grounded or oppositely charged object (which is the one being coated).

11.1.7.4 Compression moldings

CM is the most common method of forming TS plastics. Until the advent of IM, it was the most important of plastic processes. CM is the compressing of a material into a desired shape by application of heat and pressure to the material in a mold cavity. Pressure is usually at 7–14 MPa (1000–2000 psi). Some TSs may require pressures down to 345 kPa (50 psi) or even just contact (zero pressure). The majority of TS compounds are heated to about 150–200 °C (302–392 °F) for optimum cure; but can go as high as 650 °C (1200 °F).

11.17.5 Reaction IMs

The RIM process predominantly uses TS polyurethane (PUR) plastics. Others include nylon, TS polyester, and epoxy. PUR offers a large range of product performance properties. As an example PUR has a modulus of elasticity in bending of 200–1400 MPa (29,000–203,000 psi) and heat resistance in the range of 90–200 °C (122–392 °F).

The higher values are obtained when glass fiber reinforces the PUR (also with nylon, etc.). The reinforced RIM process is called RRIM or structural RIM (SRIM). Large and very thick RIM products can be molded with or without reinforcements using fast cycles.

When compared to IM that processes a plastic compound (polymer plus additives, etc.), RIM uses two liquid PUR chemical monomer components (polyol and isocyanate) that are mixed to produce the polymer (plastic). Additives such as catalysts, surfactants, fillers, reinforcements, and/or blowing agents are also incorporated. Their purpose is to propagate the reaction and form a finished product possessing the desired properties

Mixing is by a rapid impingement in a chamber (under high pressure in a specially designed mixing head) at relatively low temperatures before being injected into a closed mold cavity at low pressure. An exothermic chemical reaction occurs during mixing and in the cavity requiring less energy than the conventional IM system. Polymerization of the monomer mixture in the mold allows for the custom formulation of material properties and kinetics to suit a particular product application.

RIM is the logical process to consider at least for molding large and/or thick products. With RIM technology, cycle times of 2 min and less have been achieved in production for molding large and thick [10 cm (3.9 in.)] products. It is less competitive for small products. Capital requirements for RIM processing equipment are rather low when compared with IM equipment (includes mold) that would be necessary to mold products of similar large size.

11.1.7.6 Rotational moldings

This method, like blow molding, is used to make hollow one-piece TP parts. RM consists of charging a measured amount of TP into a warm mold cavity that is rotated in an oven about two axes. In the oven, the heat penetrates the mold, causing the plastic, if it is in powder form, to become tacky and adhere to the mold female cavity surface, or if it is in liquid form, to start to gel on the mold cavity surface. Since the molds continue to rotate during the heating cycle, the plastic will gradually become distributed on the mold cavity walls through gravitational force. As the cycle continues, the plastic melts completely, forming a homogeneous layer of molten plastic. After cooling, the molds are opened and the parts removed.

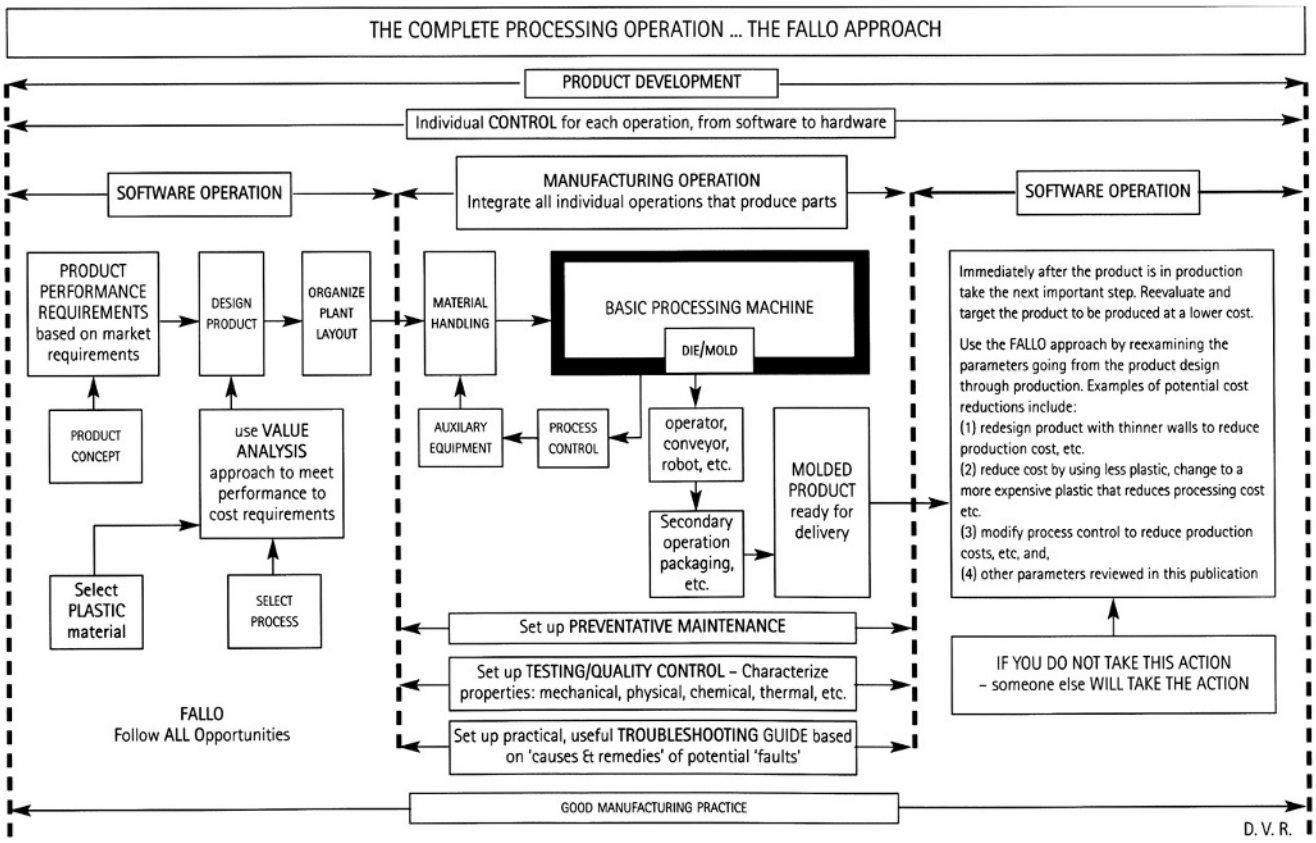


Fig. 11.1.15 The FALLO approach: *F*ollow *ALL O*pportunities.

RM can produce quite uniform wall thicknesses even when the product has a deep draw of the parting line or small radii. The liquid or powdered plastic used in this process flows freely into corners or other deep draws upon the mold being rotated and is fused/melted by heat passing through the mold's wall.

This process is particularly cost-effective for small production runs and large product sizes. The molds are not subjected to pressure during molding, so they can be made relatively inexpensively out of thin sheet metal. The molds may also be made from lightweight cast aluminum and electroformed nickel, both of them are light in weight and low in cost. Large rotational machines can be built economically because they use inexpensive gas-fired or hot air ovens with the lightweight mold-rotating equipment.

11.1.8 Variables

Even though equipment operations and plastic compounded materials have understandable and controllable variables that influence processing, the usual most uncontrollable variable in the process can be the plastic material. The degree of properly compounding or blending by the plastic manufacturer, converter, or in-house by the fabricator is important. With the passing of time and looking ahead, existing material and equipment variabilities are continually reduced due to improvement in their manufacturing and process control capabilities. However they still exist.

11.1.9 FALLO approach

Conditions that are important in making plastic products the success it has worldwide are summarized in Figure 11.1.15. All designs, processes, and materials fit into this overall FALLO (*F*ollow *ALL* Opportunities) approach flow chart that produces products meeting required performance and cost requirements.

Designers and processors, needing to produce qualified products at the lowest cost have used the basic concept of the FALLO approach. This approach makes one aware that many steps are involved to be successful, all of which must be coordinated and interrelated. It starts with the design that involves specifying the plastic and specifying the manufacturing process. The specific process (injection, extrusion, blow molding, thermoforming, and so forth) is an important part of the overall scheme.

Chapter 11.2

11.2

Product design

Dominick V. Rosato and Donald V. Rosato

11.2.1 Introduction

Plastics offer the opportunity to optimize designs by focusing on material composition as well as product structural geometry to meet different product requirements. In structural applications for plastics, which generally include those in which the product has to resist substantial static and/or dynamic loads, it may appear that one of the problem design areas for many plastics is their low modulus of elasticity. Since shape integrity under load is a major consideration for structural products, low modulus type plastic products are designed shapewise and/or thicknesswise for efficient use of the material to afford maximum stiffness and overcome their low modulus. This type of plastics and products represent most of the plastic products produced worldwide.

Throughout this chapter as the viscoelastic behavior of plastics has been described, it has been shown that deformations are dependent on such factors as the time under load and the temperature. Therefore, when structural components are to be designed using plastics it must be remembered that the extensive amount of standard equations that are available for designing springs, beams, plates, and cylinders, and so on have all been derived under certain assumptions. They are that (1) the strains are small, (2) the modulus is constant, (3) the strains are independent of the loading rate or history and are immediately reversible, (4) the material is isotropic, and (5) the material behaves in the same way in tension and compression.

Since these assumptions are not always justifiable when applied to plastics, the classic equations cannot be used indiscriminately. Each case must be considered on its merits, with account being taken of such factors as the time under load, the mode of deformation, the service temperature, the fabrication method, the environment, and others. In particular, it should be noted that the traditional equations are derived using the relationship that stress equals modulus times strain, where the modulus is a constant. The modulus of a plastic may not be a constant.

There are different design approaches to consider as reviewed in different engineering textbooks concerning specific products. They range from designing a drinking cup to the roof of a house. As an example consider a house to stand up to the forces of a catastrophic hurricane. Low pitch roofs are less vulnerable than steeper roofs because the same aerodynamic factors that make an airplane fly can lift the roof off the house. The roof also requires being properly attached to the building structure.

Example of a product design program approach follows:

1. Define the function of the product with performance requirements.
2. Identify space and load limitations of the product if they exist.
3. Define all of the environmental stresses that the product will be exposed to in its intended function.
4. Select several materials that appear to meet the required environmental requirements and strength behaviors.
5. Do several trial designs using different materials and geometries to perform the required function.
6. Evaluate the trial designs on a cost effectiveness basis. Determine several levels of performance and

Plastics Engineered Product Design; ISBN: 9781856174169
Copyright © 2003 Elsevier Ltd. All rights of reproduction, in any form, reserved.

the specific costs associated with each to the extent that it can be done with available data.
7. Determine the appropriate fabricating process for the design.
8. Based on the preliminary evaluation select the best apparent choices and do a detailed design of the product.
9. Based on the detailed design select the probable final product design, material, and process.
10. Make a model if necessary to test the effectiveness of the product.
11. Build prototype tooling.
12. Make prototype products and test products to determine if they meet the required function.
13. Redesign the product if necessary based on the prototype testing.
14. Retest.
15. Make field tests.
16. Add instructions for use.

11.2.1.1 Reinforced plastic

More extensively used are the conventional engineering plastics that are not reinforced to maximize mechanical performances of plastics. However there are reinforced plastics (RPs) that offer certain important structural and other performance requirements. These requirements provides the designer great flexibility and provides freedom practically not possible with most other materials. However, it requires a greater understanding of the interrelations to take full advantage of RPs. It is important to understand that RPs has an extremely wide range of properties, structural responses, product performance characteristics, product shapes, manufacturing processes, and influence on product performances.

The usual approach is that the designer is involved in "making the material." RP designed products have often performed better than expected, despite the use of less sophisticated fabricating tools in their design. Depending on construction and orientation of stress relative to reinforcement, it may not be necessary to provide extensive data on time-dependent stiffness properties since their effects may be small and can frequently be considered by rule of thumb using established practical design approaches. When time-dependent strength properties are required, creep, fatigue, and other data are used most effectively. These type data are available.

The arrangement and the interaction of the usual stiff, strong fibers dominate the behavior of RPs with the less stiff, weaker plastic matrix (thermoset (TS) or thermoplastic (TP)). A major advantage is that directional properties can be maximized in products by locating fibers that maximize mechanical performances in different directions.

When compared to unreinforced plastics, the analysis and design of RPs is simpler in some respects and more complicated in others. Simplifications are possible since the stress–strain behavior of RPs is frequently fairly linear to failure and they are less time-dependent. For high performance applications, they have their first damage occurring at stresses just below their high ultimate strength properties. They are also much less temperature-dependent, particularly RTSs (reinforced TSs).

When constructed from any number and arrangement of RP plies, the stiffness and strength property variations may become much more complex for the novice. Like other materials, there are similarities in that the first damage that occurs at stresses just below ultimate strength. Any review that these types of complications cause unsolvable problems is incorrect. Reason being that an RP can be properly designed, fabricated and evaluated to take into account any possible variations; just as with other materials. The variations may be insignificant or significant. In either case, the designer will use the required values and apply them to an appropriate safety factor; similar approach is used with other materials. The designer has a variety of alternatives to choose from regarding the kind, form, amount of reinforcement to use, and the process versus requirements.

With the many different fiber types and forms available, practically any performance requirement can be met and molded into any shape. However, they have to be understood regarding their advantages and limitations. As an example there are fiber bundles in lower cost woven rovings that are convoluted or kinked as the bulky rovings conform to the square weave pattern. Kinks produce repetitive variations in the direction of reinforcement with some sacrifice in properties. Kinks can also induce high local stresses and early failure as the fibers try to straighten within the matrix under a tensile load. Kinks also encourage local buckling of fiber bundles in compression and reduce compressive strength. These effects are particularly noticeable in tests with woven roving in which the weave results in large-scale reinforcement.

Fiber content can be measured in percent by weight of the fiber portion (wt%). However, it is also reported in percent by volume (vol%) to reflect better the structural role of the fiber that is related to volume (or area) rather than to weight. When content is only in percent, it usually refers to wt%.

Basic behaviors of combining actions of plastics and reinforcements have been developed and used successfully. As an example, conventional plain woven fabrics that are generally directional in the 0° and 90° angles contribute to the highest mechanical strength at those angles. The rotation of alternate layers of fabric to a layup

of 0°, +45°, 90°, and −45° alignment reduces maximum properties in the primary directions, but increases in the +45° and −45° directions. Different fabric and/or individual fiber patterns are used to develop different property performances in the plain of the molded RPs. These woven fabric patterns CAN include basket, bias, cowoven, crowfoot, knitted, leno, satin (four-harness satin, eight-harness satin, etc.), and twill.

For almost a century many different RP products have been designed, fabricated, and successfully operated in service worldwide. They range from small to large products such as small insulators for high-voltage cable lines to large 250 ft diameter deep antenna parabolic reflectors. RPs have been used in all types of transportation vehicles, different designed bridges, road surfacing such as aircraft landing strips and roads, mining equipment, water purification and other very corrosive environmental equipment, all types of electrical/electronic devices, etc.

11.2.1.2 Monocoque structure

With the flexibility in shaping and fabricating plastics provides an easy approach to designing monocoque structures. This is the type of construction in which the outer skin carries all or a major part of the stresses. Different applications take advantage of this design approach such as automotive body, motor truck, railroad car, aircraft fuselage and wings, toys, houses, rockets, and so on. This load bearing construction can integrate the product body and support/chassis into a single structure.

11.2.2 Geometric shape

Design analysis is required to convert applied loads and other external constraints into stress and strain distributions within the product, and to calculate associated deformations (Chapter 11.1). The nature and complexity of these calculations is influenced strongly by the shape of the component. It is most convenient if the component approximates to some simple idealized form, such as a plate or shell (for example a body panel), a beam or tube (for example a chassis member or bumper), or a combination of idealized forms (for example a box structure). In these cases, standard design formulas can be provided into which appropriate parameters can be substituted for a particular application.

However, there are many more cases where the component shape does not approximate to a simple standard form (for example a wheel, pump housing, or manifold) or where a more detailed analysis is required for part of a product (for example the area of a hole, boss, or attachment point). In these cases, the geometry complicates the design analysis and it may be necessary to carry out a direct analysis, possibly using finite element analysis (FEA).

One of plastics' design advantages is its formability into almost any conceivable shape. It is important for designers to appreciate this important characteristic. Both the plastic materials and different ways to fabricate products provide this rather endless capability. Shape, which can be almost infinitely varied in the early design stage, is capable for a given volume of materials to provide a whole spectrum of strength properties, especially in the most desirable areas of stiffness and bending resistance. With shell structures, plastics can be either singly or doubly curved via the different fabricating processes.

11.2.2.1 Modulus of elasticity

There are different techniques that have been used for over a century to increase the modulus of elasticity of plastics. Orientation or the use of fillers and/or reinforcements such as RPs can modify the plastic. There is also the popular and extensively used approach of using geometrical design shapes that makes the best use of materials to improve stiffness even for those that have a low modulus. Structural shapes that are applicable to all materials include shells, sandwich structures, dimple sheet surfaces, and folded plate structures (Figure 11.2.1).

11.2.2.1.1 EI theory

In all materials (plastics, metals, wood, etc.) elementary mechanical theory demonstrates that some shapes resist deformation from external loads. This phenomenon stems from the basic physical fact that deformation in beam or sheet sections depends upon the mathematical product of the modulus of elasticity (E) and the moment of inertia (I), commonly expressed as EI. It is applied to all types of constructions such as solids, foams, and sandwich structures. In many applications plastics can lend themselves in the form of a sophisticated lightweight stiff structure and the requirements are such that the structure must be of plastics. In other instances, the economics of fabrication and erection of a plastics lightweight structure and the intrinsic appearance and other desirable properties make it preferable to other materials.

This theory has been applied to many different constructions including many plastic products. In each case displacing material from the neutral plane makes the improvement in flexural stiffness. Use of this engineering principle that has been used for many centuries relates to the basic physical fact that deformation in beams or

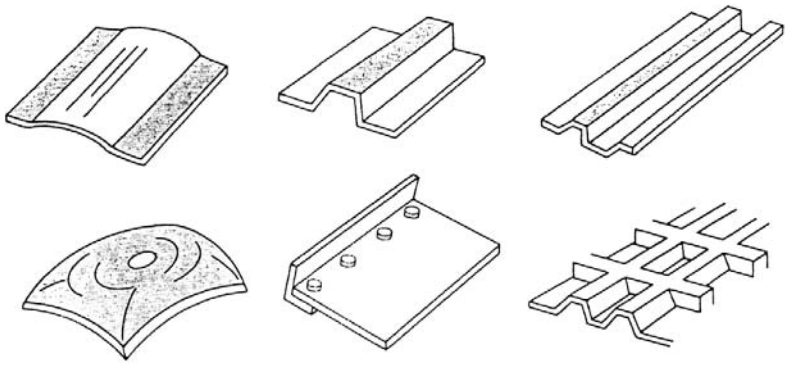


Fig. 11.2.1 Examples of shapes to increase stiffness.

sheets depends upon the mathematical product of E and I, more commonly expressed as EI.

The EI principle applies to the basic beam structures as well as hollow channel, I-shape, T-shape, etc. where it imparts increased stiffness in one direction much more than in the other. Result is more efficient strength-to-weight products and so forth. While this construction may not be as efficient as the sandwich panel, it does have the advantage that it can be easily fabricated (molded, extruded, etc.) directly in the required configuration at a low cost and the relative proportions be designed to meet the load requirements.

11.2.3 Plate

Methods for the design analysis in the past for plastics were based on models of material behavior relevant to traditional metals, as for example elasticity and plastic yield. These principles were embodied in design formulas, design sheets and charts, and in the modern techniques such as those of CAD using FEA. Design analyst was required only to supply appropriate elastic or plastic constants for the material, and not question the validity of the design methods. Traditional design analysis is thus based on accepted methods and familiar materials, and as a result many designers have little, if any, experience with such other materials as plastics, wood, and glass.

Using this approach it is both tempting and common practice for certain designers to treat plastics as though they were traditional materials such as steel and to apply familiar design methods with what seem appropriate materials constants. It must be admitted that this pragmatic approach does often yield acceptable results. However, it should also be recognized that the mechanical characteristics of plastics are different from those of metals, and the validity of this pragmatic approach is often fortuitous and usually uncertain.

It would be more acceptable for the design analysis to be based on methods developed specifically for the materials, but this action will require the designer of metals to accept new ideas. Obviously, this acceptance becomes easier to the degree that the newer methods are presented as far as possible in the form of limitations or modifications to the existing methods.

Table 11.2.1 provides examples of mechanical property data of different materials ($GN/m^2 = \text{kPa}$). A review is presented concerning the four materials in Table 11.2.1, where the exact values used are unimportant.

Higher performance types could be used for the metals and plastics but those in this table offer a fair comparison for the explanation being presented. This review identifies the need for using design analysis methods appropriate for plastics. It also indicates the uncertainty of using with plastics methods derived from metals, and demonstrates the dangers of making

Table 11.2.1 Mechanical properties of materials

Property	Aluminum	Mild steel	Polypropylene (PP)	Glass-fiber reinforced plastics (GRP)
Tensile modulus (E) 10^6 GN/m^2 (psi)	70 (10)	210 (30)	1.5 (0.21)	15 (2.2)
Tensile strength (σ) 10^3 MN/m^2 (psi)	400 (58)	450 (65)	40 (5.8)	280 (40.5)
Specific gravity (S)	2.7	7.8	0.9	1.6

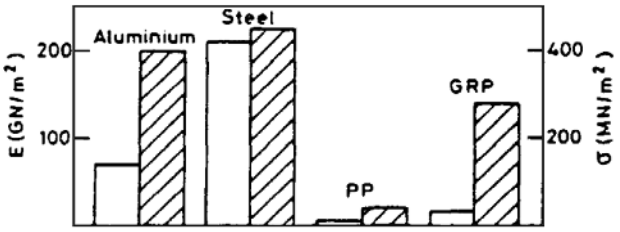


Fig. 11.2.2 Open bar illustrations represent stiffness and shaded illustrations represent strength with panels having the same dimensions.

generalized statements about the relative merits of different classes of materials.

Based on the usual data on metals, they are much stiffer and significantly stronger than plastics. This initial evaluation could eliminate the use of plastics in many potential applications, but in practice it is recognized by those familiar with the behavior of plastics that it is the stiffness and strength of the product that is important, not its material properties.

The proper approach is to consider the application in which a material is used such as in panels with identical dimensions with the service requirements of stiffness and strength in flexure. Their flexural stiffnesses and strengths depend directly on the respective material's modulus and strength. Other factors are similar such as no significantly different Poisson ratios. The different panel properties relative to stiffness and strength are shown in Figure 11.2.2. The metal panels are stiffer and stronger than the plastic ones because the panels with equal dimensions use equal volumes of materials.

By using the lower densities of plastics it allows them to be used in thicker sections than metals. This approach significantly influences the panel's stiffnesses and strengths. With equal weights and therefore different thicknesses (t) the panels are loaded in flexure. Their stiffnesses depend on (Et^3) and their strength on (σt^2) where E and σ are the material's modulus and its strength. For panels of equal weight their relative stiffnesses are governed by (E/s^3) and their relative strengths

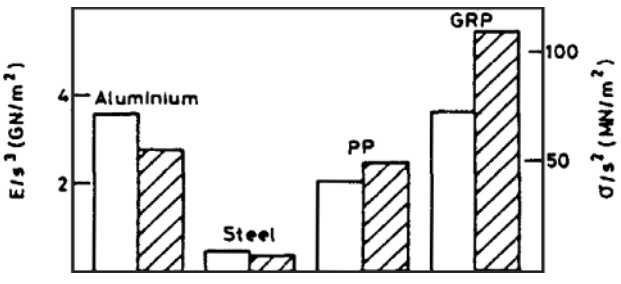


Fig. 11.2.3 Open bar illustrations represent stiffness and shaded illustrations represent strength with panels having the equal weights.

by (σ/s^2) where s denotes specific gravity. As shown in Figure 11.2.3, the plastics now are much more favorable. So depending on how one wants to present data or more important apply data either Figures 11.2.2 or 11.2.3 or Table 11.2.1 is used. Thus, the designer has the opportunity to balance out the requirements for stiffness, strength, and weight saving.

Recognize that it is easy to misinterpret property data and not properly analyze the merits of plastics. No general conclusions should be drawn on the relative merits of various materials based on this description alone. In comparing materials in Table 11.2.1 a designer can easily obtain different useful data. As an example the GRP panel has 2.4 times the thickness of a steel panel for the same flexural stiffness. It has 3.6 times its flexural strength and only half its weight. The tensile strength of the GRP panel would be 50% greater than that of the steel panel, but its tensile stiffness is only 17% that of the steel panel.

Similar remarks could be made with respect to various materials' costs and energy contents, which can also be specified per unit of volume or weight. General statements about energy content or cost per unit of stiffness or strength, as well as other factors, should be treated with caution and applied only where relevant. If these factors are to be treated properly, they too must relate to final product values that include the method of fabrication, expected lifetime, repair record, and in-service use.

This review shows what the veteran plastic designer knows; that plastic products are often stiffness critical, whereas metal products are usually strength critical. Consequently, metal products are often made stiffer than required by their service conditions, to avoid failure, whereas plastic products are often made stronger than necessary, for adequate stiffness. In replacing a component in one material with a similar product in another material is not usually necessary to have the same product stiffness and strength.

11.2.3.1 Folded plate

Capability in fabricating simple to complex folded products makes designing them easy based on analysis such as beams with rectangular, triangular, spherical, or other shapes. Products vary from those with spring actions to movable contoured walls and ceilings to bottles and outer-space structures. Elementary beam equations are used. When assemblies are plates whose lengths are large relative to their cross-sectional dimensions (thin-wall beam sections, ribbed panels, and so on) and are in large plates whose fold lines deflect identically, such as the interior bays of roofs or bridge structures, they can be analyzed as beams. More elaborate documented

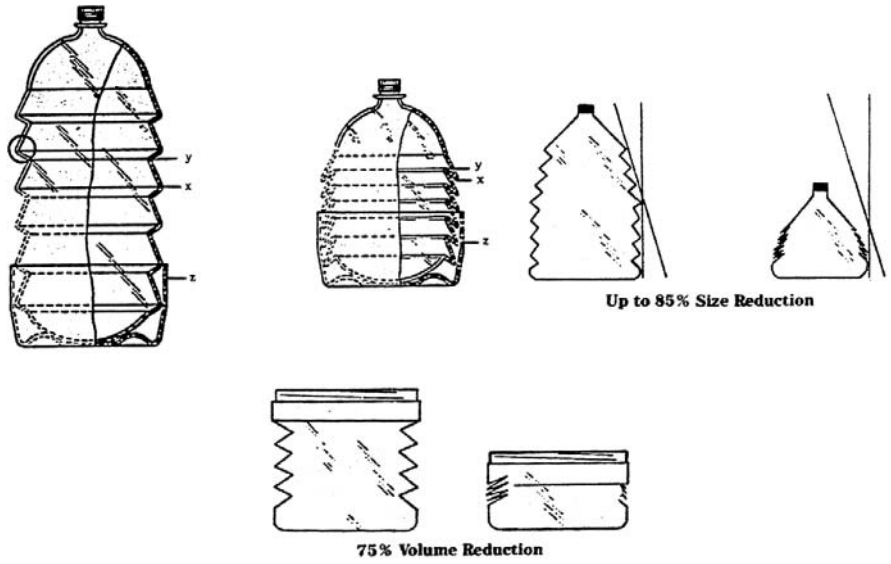


Fig. 11.2.4 Views of a collapsible bottle. (Courtesy of Collapsible Bottle of America Co.)

procedures are used to determine transverse multibending stresses in assemblies.

An example of folded plate technology are bellows-style collapsible plastic containers such as blow molded bottles that are foldable. The foldable shaped containers in contrast to that of the usual shaped bottles provides advantages and conveniences. Examples include reduced storage as the container's content is reduced, transportation volume and weight costs relate to the collapsed size, and disposal space; prolonged product freshness by reducing oxidation and loss of carbon dioxide; and provides extended life via continuous collapsing surface access to foods like mayonnaise and jams.

Figures 11.2.4 and 11.2.5 show the bellows of collapsible containers that overlap and fold to retain their folded condition without external assistance, thus providing a self-latching feature. The views show uncollapsed bottle, collapsed bottle, and top view of the bottle. This latching is the result of bringing together under pressure two adjacent conical sections of unequal proportions and different angulations to the bottle axis. On a more technical analytical level the latching is created basically by the swing action of one conical section around a fixed pivot point, from an outer to an inner, resting position. The two symmetrically opposed pivot points and rotating segments keep a near-constant diameter as they travel along the bottle axis. This action explains the bowing action of the smaller, conical section as it approaches the overcentering point.

A characteristic in molding some products that are designed to include collapsing, living hinge, etc. action requires them to be subjected to the action as soon as possible after processing. This initial action is used to create permanent fold rings and completely orient the plastic molecules. In most disposable applications these bottles would undergo three changes of volume: an initial collapsing of the container before shipment and storage; expansion of the container at its destination, before or during filling; and finally gradually collapsing the bottle for reuse or disposal.

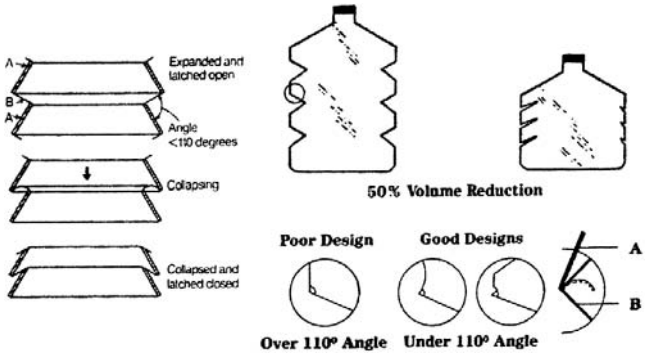


Fig. 11.2.5 Geometric concepts to the collapsible bottle. (Courtesy of Collapsible Bottle of America Co.)

11.2.4 Beam

As the engineering books explain, a beam is a bar or structural member subjected to transverse loads that tend to bend it. Any structural members act as a beam if external transverse forces induce bending. A simple beam is a horizontal member that rests on two supports at the ends of the beam. All parts between the supports have free movement in a vertical plane under the influence of vertical loads.

There are fixed beams, constrained beams, or restrained beams rigidly fixed at both ends or rigidly fixed at one end and simply supported at the other. A continuous beam is a member resting on more than two supports. A cantilever beam is a member with one end projecting beyond the point of support, free to move in a vertical plane under the influence of vertical loads placed between the free end and the support.

When a simple beam bends under its own weight, the plastic or fibers in a plastic on the upper or concave side is shortened, with the stress acting on them is compression. The fibers on the under or convex side are lengthened, and the stress acting on them is tension. In addition, shear exists along each cross section, the intensity of which is greatest along the sections at the two supports and zero at the middle section. When a cantilever beam bends under its own weight the fibers on the upper or convex side are lengthened under tensile stresses. The fibers on the under or concave side are shortened under compressive stresses, the shear is greatest along the section at the support, and zero at the free end.

The neutral surface is that horizontal section between the concave and convex surfaces of a loaded beam, where there is no change in the length of the fibers and no tensile or compressive stresses acting upon them. The neutral axis is the trace of the neutral surface on any cross section of a beam. The elastic curve of a beam is the curve formed by the intersection of the neutral surface with the side of the beam, it being assumed that the longitudinal stresses on the fibers are within the elastic limit.

The reactions, or upward pressures at the points of support, are computed by applying certain conditions necessary for equilibrium of a system of vertical forces in the same plane. They are the algebraic sum of all vertical forces that must equal zero; that is, the sum of the reactions equals the sum of the downward loads. There is also the algebraic sum of the moments of all the vertical forces that equals zero.

The first condition applies to cantilever beams and to simple beams uniformly loaded, or with equal concentrated loads placed at equal distances from the center of the beam. In the cantilever beam, the reaction is the sum of all the vertical forces acting downward, comprising the weight of the beam and the superposed loads. In the simple beam each reaction is equal to one-half the total load, consisting of the weight of the beam and the superposed loads. The second condition applies to a simple beam not uniformly loaded. The reactions are computed separately, by determining the moment of the several loads about each support. The sum of the moments of the load around one support is equal to the moment of the reaction of the other support around the first support.

The fundamental laws for the stresses at any cross section of a beam in equilibrium are: (1) sum of the horizontal tensile stresses equal sum of horizontal compressive stresses, (2) resisting shear equal vertical shear, and (3) resisting moment equal bending moment. Bending moment at any cross section of a beam is the algebraic sum of the moments of the external forces acting on either side of the section. It is positive when it causes the beam to bend convex downward, thus causing compression in upper fibers and tension in lower fibers of the beam. When the bending moment is determined from the forces that lie to the left of the section, it is positive if they act in a clockwise direction; if determined from forces on the right side, it is positive if they act in a counterclockwise direction. If the moments of upward forces are given positive signs, and the moments of downward forces are given negative signs, the bending moment will always have the correct sign, whether determined from the right or left side. The bending moment should be determined for the side for which the calculation will be the simplest.

The deflection of a beam as computed by the ordinary formulas is that due to flexural stresses only. The deflection in honeycomb and short beams due to vertical shear can be high, and should always be checked. Because of the nonuniform distribution of the shear over the cross section of the beam, computing the deflection due to shear by exact methods is difficult. It may be approximated by

$$y_s = M/AE_s \tag{11.2.1}$$

where y_s is the deflection, inch, due to shear; M the bending moment, lb-in, at the section where the deflection is calculated; A the area of cross section of beam, square inches; and E_s the modulus of elasticity in shear, psi. For a rectangular section, the ratio of deflection due to shear to the deflection due to bending, will be less than 5% if the depth of the beam is less than one-eighth of the length.

In designing a beam an approach is: (1) compute reactions; (2) determine position of the dangerous section and the bending moment at that section; (3) divide the maximum bending moment (lb-in) by the allowable unit stress (psi) to obtain the minimum value of the section modulus; and (4) select a beam section with a section modulus equal to or slightly greater than the section modulus required.

Assumptions are made in simple beam-bending theory that involve (1) all deflections are small, so that planar cross sections remain planar before and after bending; (2) the beam is initially straight, unstressed, and symmetrical; (3) its proportional limit is not exceeded; and (4) Young's modulus for the material is the same in both tension and compression. In the analysis maximum stress occurs at the surface of the beam farthest from the neutral surface, as given by the following equation:

$$\sigma = Mc/I = M/Z \tag{11.2.2}$$

where M is the bending moment in in./lbs., c the distance from the neutral axis to the outer surface where the maximum stress occurs in in., I the moment of inertia in in.^{4}, and $Z=I/c$ the section modulus in in.^{3}. Observe that this is a geometric property, not to be confused with the modulus of the material, which is a material property.

11.2.5 Rib

11.2.5.1 Plastic

When discussion problems in minimizing or increasing load-bearing requirements in wall thickness of plastics, ribbing is recommended if it is determined that space exist for adding ribs (Figures 11.2.6 and 11.2.7). If there is sufficient space, the use of ribs is a practical, economic means of increasing the structural integrity of plastic products without creating thick walls. Table 11.2.2 shows how a rib structure rigidity (2 in. rib spacings) compares with non-ribbed panels where each are 1 × 2 ft. The rib height is 0.270 in. with a thickness of 0.065 in. If it is determined after a product is produced that ribbing is required adding ribs after the tool is built is usually simple and relatively inexpensive since it involves removing steel in the fabricating tool. With thinner walls a major cost saving could develop since processing them reduces processing time and provides more heat uniformity during processing.

Handbooks reviewing stress and deflections in beams and moments of inertia provide information such as the moment of inertia and resistance to deflection that expresses the resistance to stress by the section modulus.

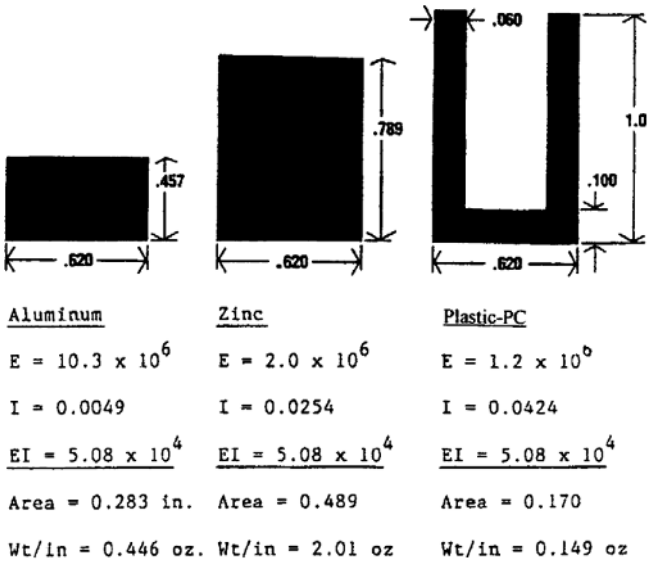


Fig. 11.2.7 Comparing a rib design strengthwise and weightwise with other materials.

Table 11.2.2 Example where ribbing is beneficial

Property	Steel	Solid plastic	Structural foam
Thickness (in.)	0.040	0.182	0.196
E (psi)	3×10^7	3.2×10^5	2.56×10^5
I (in.^{4})	0.000064	0.006	0.0075
$E \times I$ (rigidity)	1920	1920	1920
Weight (lbs.)	3.24	1.98	1.78

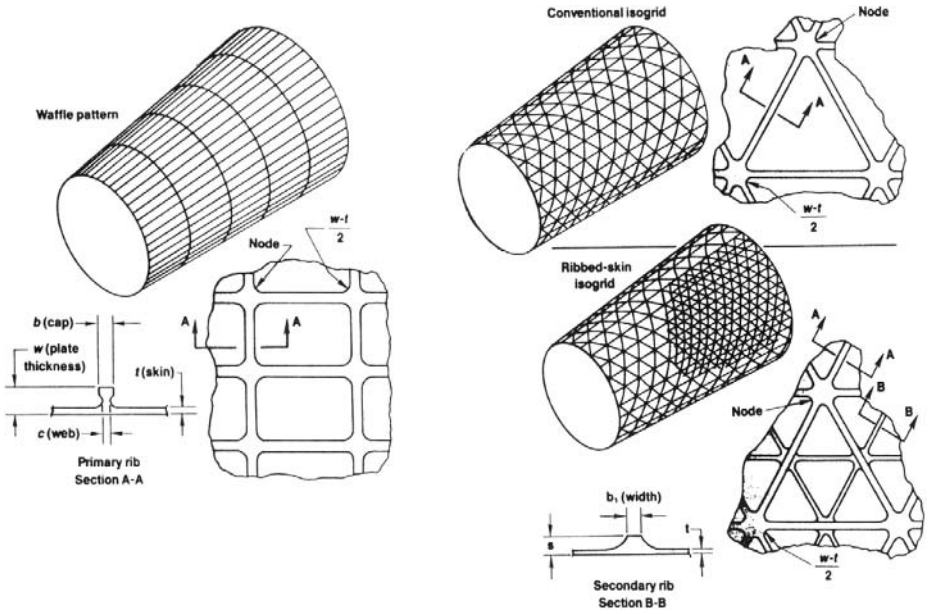


Fig. 11.2.6 Example of a design where thin wall with ribbing supports high edge loading.

By finding a cross section with the two equivalent factors, one ensures equal or better performance of the fabricated product. The moment of inertia can be changed substantially by adding ribs and other shapes such as gussets as well as their combinations.

There are available basic engineering rib-design guidelines. The most general approach is to make the rib thickness at its base a minimum of one-half the adjacent wall's thickness. With ribs opposite appearance areas, the width should be kept as thin as possible. In areas where structure is more important than appearance, or with very low shrinkage materials, ribs are often 75% or even 100% of the outside wall's thickness. A goal in rib design is to prevent the formation of a heavy mass of material that can result in a sink, void, distortion, long cycle time, or any combination of these problems.

11.2.5.2 Reinforced/foamed plastic

Frequently cases arise in which ribs are used to reinforce plastic plates such as in tanks, boat hulls, bridges, floors, towers, buildings, and so on. The design of ribbed plates such as these is somewhat analogous to reinforced concrete and T-beam design. In view (a) of Figure 11.2.8, for example, a construction is shown consisting of a plate composed of balanced fabric RP 0.15 in. thick and mat RP 0.05 in. thick, combined with a rib, making a structure whose overall depth is 1.500 in. The rib is formed of a cellular material such as foamed plastic, plus a cluster of resin-bonded parallel fibers such as roving, at the bottom. The mat is carried around the rib and serves to tie the rib and plate together.

The construction of a plate and rib form a T-beam. The principal design problem is to determine how much of the plate can be considered to be acting as a flange with the rib, that is, the magnitude of b in Figure 11.2.8(a). For purposes of illustration, b is taken as 5 in. If the T-beam is loaded in bending so as to induce compression at the top and tension at the bottom, the balanced fabric and the mat will be in compression at the top, and the roving and mat will be in tension at the bottom. Because the roving is much stronger than the mat, it is evident that the mat adjacent to it will break before the roving reaches its maximum stress. That is if the roving were stressed 50,000 psi the adjacent mat would be stressed 10,000 psi, which is double its strength. Consequently, in finding the neutral axis and computing the strength of the cross section, the mat is neglected on the tension side. Above the neutral axis the mat is in compression, but in order to simplify the computations, only the mat in the flange of the T is considered.

The foamed plastic has such a low modulus of elasticity and such low strength that it contributes little to either the stiffness or the bending strength of the T-beam. It must, however, be stiff enough to prevent buckling or wrinkling of the mat or the roving.

The active elements of the T-beam are therefore as shown in Figure 11.2.8(b). The flange consists of balanced fabric 5.000 in. wide and 0.150 in. thick plus mat 4.200 in. wide and 0.050 in. thick. The web consists of

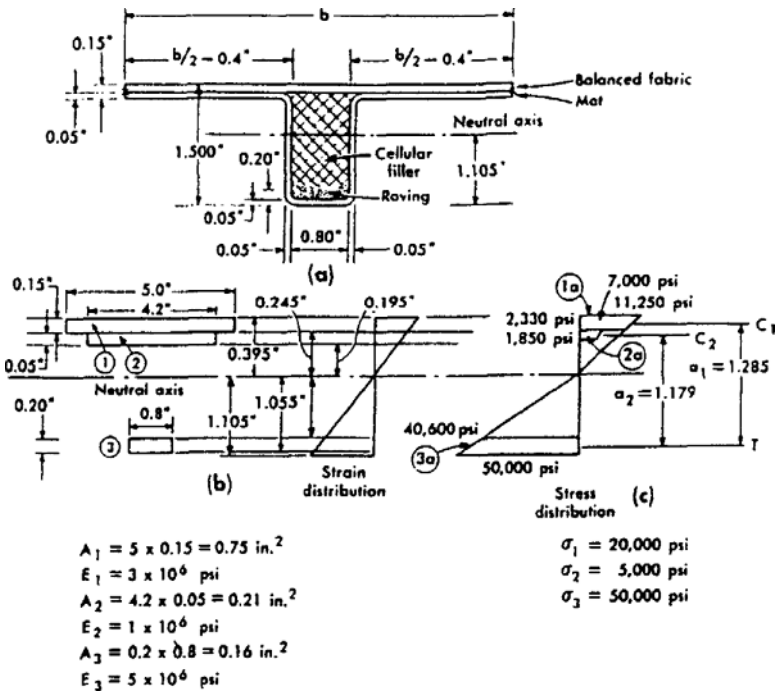


Fig. 11.2.8 Cross section of a rib applied to a plate.

the bundle of roving 0.800 in. wide and 0.200 in. thick. By the application of Equation (11.2.32), the neutral axis is found to be 1.105 in. from the bottom, or 1.055 in. above the lower edge of the roving.

The basic assumptions discussed in the introduction imply that when this beam is bent, strains at any point in both tension and compression are proportional to the distance from the neutral axis, and that stress is equal to strain multiplied by modulus of elasticity. As an example, the stress in the lower-most roving fiber is 50,000 psi, the stress in the topmost fiber of the flange is 50,000 $(3 \times 10^6/5)$ (0.395/1.055) or 11,250 psi. Similarly, the stresses in the upper edge of the bundle of rovings, at the lower edge of the balanced fabric, and at the upper and lower edges of the balanced fabric, and at the upper and lower edges of the mat in the flange are as shown in Figure 11.2.8(c). These are all less than the corresponding values of σ_1, σ_2, and σ_3 listed.

The internal resisting moment, or resistance to outside bending forces, can be found by computing the total resultant compression C_1 in the balanced fabric, total resultant compression C_2 in the mat, finding the distances α_1 and α_2 between the lines of action of these two resultants and the line of action of the total resultant tension T in the roving, computing the values $C_1\alpha_1$ and $C_2\alpha_2$, and adding.

Resultant C_1 acts at the centroid of the trapezoidal stress area 1*a*, resultant C_2 at the centroid of area *2a*, and resultant T at the centroid of area *3a*. Solving for these centroids, the distance α_1 is found to be 1.285 in. and distance α_2 is 1.179 in. These are the internal moment arms of the two resultant compressive forces C_1 and C_2:

$$\begin{aligned} C_1 &= [(11,250+7000)/2][5.000\times 0.150] = 6840 \text{ lb} \\ C_2 &= [(2330+1850)/2][4.200\times 0.05] = \underline{\ 440} \text{ lb} \\ \text{Total } C &= \overline{7280} \end{aligned} \tag{11.2.3}$$

A check on the accuracy of the computations is afforded by the fact that T must equal C:

$$\begin{aligned} T &= [(40,600+50,000)/2][0.800\times 0.200] \\ &= 7250 \text{ lb} \end{aligned} \tag{11.2.4}$$

The initial resisting moment is $C_1\alpha_1 + C_2\alpha_2$

$$\begin{aligned} C_1\alpha_1 &= 6840\times 1.285 = 8790 \text{ in-lb} \\ C_2\alpha_2 &= 440\times 1.179 = \ \ 520 \\ M_{res} &= 9310 \text{ in-lb} \end{aligned} \tag{11.2.5}$$

Evidently, if the mat in the flange had been left out of the computation ($C_2\alpha_2$), the error in the calculated result would have been approximately 5%. Omitting the mat in the rib between the flange and the neutral axis affected the result much less.

If a shear force is imposed on the rib, two critical planes of internal shear stress occur, one at the neutral axis and one at the plane between the mat and the fabric in the flange (Figure 11.2.8(a)). Shear stresses are computed by equation,

$$\tau = \frac{VQ'}{bEI} \tag{11.2.6}$$

For example, suppose the T-beam is 30 in. long (L) and carries a uniformly distributed load W. Then $M = wL/8 = 9310$ in-lb and $W = 2500$ lb. The maximum shear V is half the total load or 1250 lb.

At the neutral axis the statistical moment Q′ is the weighted moment of the flange or of the roving about the neutral axis; EI is the stiffness factor of flange plus roving; and b is the total thickness of the mat at the neutral axis plus the effective thickness of the cellular filler. This effective thickness may be computed in accordance with the principles set forth on combined action. If, for example, the shear modulus of the cellular core is 1/15 the shear modulus of the mat-reinforced material, the effective width of the core is 0.8/15 or 0.053 in. The total effective width of the mat and the core at the neutral axis is therefore $0.05 + 0.05 + 0.053 = 0.153$ in.

The computed value of Q′ is 0.725×10^6 lb-in, and the value of EI is 0.968×10^6 psi. The shear stress intensity in the mat at the neutral axis is:

$$\tau_m = \frac{1250\times 0.725}{0.153\times 0.968} = 6100 \text{ psi} \tag{11.2.7}$$

The shear stress intensity in the cellular core is

$$\tau_c = 6100/15 = 410 \text{ psi} \tag{11.2.8}$$

If either τ_m, or τ_r, is excessive it is necessary to increase the rib thickness at the neutral axis, probably by increasing the thickness of the mat.

Properties of the cellular core may not be known well enough or may be too low to warrant inclusion with the mat in calculating shear. If the core is not included, the thickness at the neutral axis is that of the mat alone, or 0.10 in. The shear stress τ_m, then becomes 9400 psi instead of 6100 psi.

At the interface between mat and fabric in the flange of the T-beam, the value of Q′ is that of the fabric alone. This is found to be 0.720×10^6 psi. The width b is 4.2 in. (neglecting the width of the cellular core). Therefore:

$$\tau = \frac{1250\times 0.720}{4.2\times 0.968} = 220 \text{ psi} \tag{11.2.9}$$

In all probability the shear stress intensity is actually higher adjacent to the rib, and lower near the outer ends of the flange, but in any event it is not likely to be excessive.

11.2.6 Column

A column or strut is a bar or structural member under axial compression, which has an unbraced length greater than about eight or ten times the least dimension of its cross section. Because of its length, it is impossible to hold a column in a straight line under a load. A slight sidewise bending always occurs, causing flexural stresses in addition to the compressive stresses induced directly by the load. The lateral deflection will be in a direction perpendicular to that axis of the cross section about which the moment of inertia is the least. Thus, in a complex shape such an H-column will bend in a direction perpendicular to its major axis. In a square shape it will bend perpendicular to its two major axes. With a tubular shape it is likely to bend in any direction.

The radius of gyration of a column section with respect to a given axis is equal to the square root of the quotient of the moment of inertia with respect to that axis, divided by the area of the section, that is,

$$k = \sqrt{\frac{I}{A}}; \quad \frac{I}{A} = k^2 \tag{11.2.10}$$

where I is the moment of inertia and A is the sectional area. Unless otherwise mentioned, an axis through the center of gravity of the section is the axis considered. As in beams, the moment of inertia is an important factor in the ability of the column to resist bending, but for purposes of computation it is more convenient to use the radius of gyration. The length of a column is the distance between points unsupported against lateral deflection. The slenderness ratio is the length l divided by the least radius of gyration k.

Various conditions may exist at the ends of columns that usually are divided into four classes. (1) Columns with round ends; the bearing at either end has perfect freedom of motion, as there would be with a ball-and-socket joint at each end. (2) Columns with hinged ends; they have perfect freedom of motion at the ends in one plane, as in compression members in bridge trusses where loads are transmitted through endpins. (3) Columns with flat ends; the bearing surface is normal to the axis of the column and of sufficient area to give at least partial fix to the ends of the columns against lateral deflection. (4) Columns with fixed ends; the ends are rigidly secured, so that under any load the tangent to the elastic curve at the ends will be parallel to the axis in its original position.

Tests prove that columns with fixed ends are stronger than columns with flat, hinged, or round ends, and that columns with round ends are weaker than any of the other types. Columns with hinged ends are equivalent to those with round ends in the plane in which they have movement; columns with flat ends have a value intermediate between those with fixed ends and those with round ends. Usually columns have one end fixed and one end hinged, or some other combination. Their relative values may be taken as intermediate between those represented by the condition at either end. The extent to which strength is increased by fixing the ends depends on the length of column; fixed ends have a greater effect on long columns than on short ones.

There is no exact theoretical formula that gives the strength of a column of any length under an axial load. Different formulas involving the use of empirical coefficients have been deduced, however, and they give results that are consistent with the results of tests. These formulas include the popular Euler's formula, different eccentric formulas, cross-bend formulas, wood and timber column formulas, and general principle formulas.

11.2.6.1 Euler's formula

The Euler's formula developed during 1759 by Leonard Euler (Swiss Mathematician, 1707–1783) is used in product designs and also in designs using columns in molds and dies that processes plastic. Euler's formula assumes that the failure of a column is due solely to the stresses induced by sidewise bending. This assumption is not true for short columns that fail mainly by direct compression, nor is it true for columns of medium length. The failure in such cases is by a combination of direct compression and bending.

Column formulas are found in most machine and tooling handbooks as well as strength of materials textbooks. Euler first published this critical-load formula for columns in 1759. For slender columns it is usually expressed in the following form:

$$F = \frac{m\pi^2 EI}{I^2} = \frac{m\pi^2 EA}{(I/K)^2} \tag{11.2.11}$$

where F is the collapsing load on the column in pounds, I the length of the column in inches, A the area of the section in square inches, k the least radius of gyration, which is I/A, E the modulus of elasticity, I the least moment of inertia of the section, m is a constant depending on the end conditions of the column.

Euler's formula is strictly applicable to long and slender columns, for which the buckling action predominates over the direct compression action and thus makes no allowance for compressive stress. The slenderness ratio is defined as the ratio of length I to the radius of gyration k, represented as l/k.

When the slenderness ratio exceeds a value of 100 for a strong slim column, failure by buckling can be expected. Columns of stiffer and more brittle materials

Table 11.2.3 Moments of inertia and radii of gyration

	Moment of inertia I	Radius of gyration $\sqrt{\frac{I}{A}} = k$
	$\frac{\pi D^4}{64} = \frac{\pi r^4}{4}$	$\frac{D}{4} = \frac{r}{2}$
	$\frac{bh^3}{12}$	$\frac{h}{\sqrt{12}} = 0.289h$
	$\frac{\pi D^4}{128} = \frac{\pi r^4}{8}$	$\frac{D}{5.66}$

will buckle at lower slenderness ratios. The constant factor m in Euler's critical-load formula clearly shows that the failure of a column depends on the configuration of the column ends. The basic four types with their respective m are:

1. Both ends pivoted or hinged ($m = 1$)
2. One end fixed and the other free ($m = 1/4$)
3. One end fixed and the other pivoted ($m = 2$)
4. Both ends fixed ($m = 4$).

Table 11.2.3 shows cross sections of the three common slender column configurations. Formulas for each respective moment of inertia I and radius of gyration k are given. With the above formulas buckling force F can be calculated for a column configuration. Table 11.2.4 lists values of slim ratios (I/k) for small-nominal-diameter column lengths.

Most failures with the slender columns occur because the slenderness ratio exceeds 100. The prudent designer devises ways to reduce or limit the slenderness ratio.

In the following formula P is the axial load; l the length of column; I the least moment of inertia; k the least radius of gyration; E the modulus of elasticity; y the lateral deflection, at any point along a larger column, that is caused by load P. If a column has round ends, so that the bending is not restrained, the equation of its elastic curve is

$$EI\frac{d^2y}{dx^2} = -Py \tag{11.2.12}$$

When the origin of the coordinate axes is at the top of the column, the positive direction of x being taken downward and the positive direction of y in the direction of the deflection. Integrating the above expression twice and determining the constants of integration give:

$$P = \Omega\pi^2\frac{EI}{l^2} \tag{11.2.13}$$

which is Euler's formula for long columns. The factor Ω is a constant depending on the condition of the ends. For round ends $\Omega = 1$; for fixed ends $\Omega = 4$; for one end round and the other fixed $\Omega = 2.05$. P is the load at which, if a slight deflection is produced, the column will not return to its original position. If P is decreased, the column will approach its original position, but if P is increased, the deflection will increase until the column fails by bending.

For columns with value of l/k less than about 150, Euler's formula gives results distinctly higher than those observed in tests. Euler's formula is used for long members and as a basis for the analysis of the stresses in some types of structural parts. It always gives an ultimate and never an allowable load.

11.2.7 Torsion

A bar is under torsional stress when it is held fast at one end, and a force acts at the other end to twist the bar. In a round bar (Figure 11.2.9) with a constant force acting,

Table 11.2.4 Slenderness ratio l/k of round columns

	Diameter (in.)						
Column length (in.)	**0.031**	**0.047**	**0.0625**	**0.078**	**0.083**	**0.125**	**0.1875**
1.0	128	85	64	51	43	32	21
1.5	192	128	96	77	64	48	32
1.75	224	149	112	90	75	56	37
2.0	256	171	128	102	85	64	43
2.25	288	192	144	115	96	72	48
2.5	320	213	160	128	107	80	53
3.0	384	256	192	154	128	96	64
3.25	416	277	206	166	139	104	69

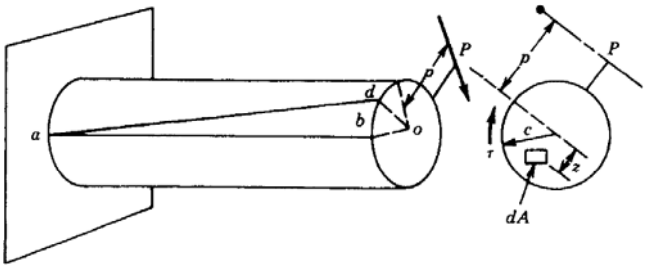


Fig. 11.2.9 Round bar subject to torsion stress,

the straight-line *ab* becomes the helix *ad*, and a radial line in the cross section, *ob*, moves to the position *ad*. The angle *bad* remains constant while the angle *bod* increases with the length of the bar. Each cross section of the bar tends to shear off the one adjacent to it, and in any cross section the shearing stress at any point is normal to a radial line drawn through the point. Within the shearing proportional limit, a radial line of the cross section remains straight after the twisting force has been applied, and the unit shearing stress at any point is proportional to its distance from the axis.

The twisting moment, *T*, is equal to the product of the resultant, *P*, of the twisting forces, multiplied by its distance from the axis, *p*. Resisting moment, T_r, in torsion, is equal to the sum of the moments of the unit shearing stresses acting along a cross section with respect to the axis of the bar. If d*A* is an elementary area of the section at a distance of *z* units from the axis of a circular shaft (Figure 11.2.9 (b)), and *c* is the distance from the axis to the outside of the cross section where the unit shearing stress is τ, then the unit shearing stress acting on d*A* is $(\tau z/c)\mathrm{d}A$, its moment with respect to the axis is $(\tau z^2/c)\mathrm{d}A$, and the sum of all the moments of the unit shearing stresses on the cross section is $f(\tau z^2/c)\mathrm{d}A$. In this expression the factor fz^2 d*A* is the polar moment of inertia of the section with respect to the axis. Denoting this by *J* the resisting moment may be written $\tau J/c$.

The polar moment of inertia of a surface about an axis through its center of gravity and perpendicular to the surface is the sum of the products obtained by multiplying each elementary area by the square of its distance from the center of gravity of its surface; it is equal to the sum of the moments of inertia taken with respect to two axes in the plane of the surface at right angles to each other passing through the center of gravity section of a round shaft.

The analysis of torsional shearing stress distribution along noncircular cross sections of bars under torsion is complex. By drawing two lines at right angles through the center of gravity of a section before twisting, and observing the angular distortion after twisting, it has been found from many experiments that in noncircular sections the shearing unit stresses are not proportional to their distances from the axis. Thus in a rectangular bar there is no shearing stress at the comers of the sections, and the stress at the middle of the wide side is greater than at the middle of the narrow side. In an elliptical bar the shearing stress is greater along the flat side than at the round side.

It has been found by tests as well as by mathematical analysis that the torsional resistance of a section, made up of a number of rectangular parts, is approximately equal to the sum of the resistances of the separate parts. It is on this basis that nearly all the formulas for noncircular sections have been developed. For example, the torsional resistance of an I-beam is approximately equal to the sum of the torsional resistances of the web and the outstanding flanges. In an I-beam in torsion the maximum shearing stress will occur at the middle of the side of the web, except where the flanges are thicker than the web, and then the maximum stress will be at the midpoint of the width of the flange. Reentrant angles, as those in l-beams and channels, are always a source of weakness in members subjected to torsion.

The ultimate/failure strength in torsion, the outer fibers of a section are the first to shear, and the rupture extends towards the axis as the twisting is continued. The torsion ula for round shafts has no theoretical basis after the shearing stresses on the outer fibers exceed the proportional limit, as the stresses along the section then are no longer proportional to their distances from the axis. It is convenient, however, to compare the torsional strength of various materials by using the formula to compute values of τ at which rupture takes place.

11.2.8 Sandwich

The same or different materials are combined in the form of sandwich structures (Figure 11.2.10). They can be used in products with an irregular distribution of the different materials, and in the form of large structures or sub-structures. A sandwich material composed of two skins and a different core material is similar to RP laminates. Overall load-carrying capabilities depend on average local sandwich properties, but materials failure criteria depend on local detailed stress and strain distributions. Design analysis procedures for sandwich materials composed of linear elastic constituents are well developed. In principle, sandwich materials can be analyzed as composite structures.

Structures and sub-structures composed of a number of different components and/or materials, including traditional materials, obey the same principles of design analysis. Stresses, strains, and displacements within individual components must be related through the characteristics (anisotropy, viscoelasticity, and so on) relevant to the particular material, and loads and displacements

must be compatible at component interfaces. Thus, each individual component or sub-component must be treated.

Load and support conditions for individual components depend on the complete structure (or system) analysis, and are unknown to be determined in that analysis. As an example, if a plastic panel is mounted into a much more rigid structure, then its support conditions can be specified with acceptable accuracy. However, if the surrounding structure has comparable flexibility to the panel, then the interface conditions will depend on the flexural analysis of the complete structure. In a more localized context, structural stiffness may be achieved by ribbing and relevant analyses may be carried out using available design formulas (usually for elastic behavior) or FEA but necessary anisotropy or viscoelasticity complicate the analysis, often beyond the ability of the design analyst.

11.2.8.1 Design

A structural sandwich is a specially shaped product in which two thin facings of relatively stiff, hard, dense, strong material is bonded to a relatively thick core material. With this geometry and relationship of mechanical properties, facings are subjected to almost all the stresses in transverse bending or axial loading. The geometry of the arrangement provides for high stiffness combined with lightness, because the stiff facings are at a maximum distance from the neutral axis, similar to the flanges of an I-beam. Overall load-carrying capabilities depend on average local sandwich properties, but material failure criteria depend on local detailed stress and strain distributions. Design analysis procedures and fabricating procedures for sandwich materials composed of linear elastic constituents are well developed and reported in the literature. In principle, sandwich materials can be analyzed as RP composites.

The usual objective of a sandwich design is to save weight, increase stiffness, use less expensive materials, or a combination of these factors, in a product. Sometimes, other objectives are also involved such as reducing tooling and other costs, achieving smooth or aerodynamic smoothness, reducing reflected noise, or increasing durability under exposure to acoustic energy. The designers consider factors such as getting the loads in, getting the loads out, and attaching small or large load-carrying members under constraints of deflection, contour, weight, and cost. To design properly, it is important to understand the fabrication sequence and methods, use of the correct materials of construction, the important influence of bond between facing materials and core, and to allow a safety factor that will be required on original, new developments. Use of sandwich panels are extensively used in building and, construction, aircraft, containers, etc.

The primary function of the face sheets is to provide the required bending and in-plane shear stiffness, and to carry the axial, bending, and in-plane shear loading. In high-performance structures, facings most commonly chosen are RPs (usually prepreg), solid plastic, aluminum, titanium, or stainless steel.

The primary function of a core in structural sandwich parts is that of stabilizing the facings and carrying most of the shear loads through the thickness. In order to perform this task efficiently, the core must be as rigid and as light as possible, and must deliver uniformly predictable properties in the environment and meet performance requirements. Several different materials are used such as plastic foam, honeycomb (using RP, film (plastic, steel, aluminum, paper, etc.), balsa wood, etc.).

Different fabricating processes are used. These include bag molding, compression molding, reinforced reaction rejection molding (RRIM), filament winding,

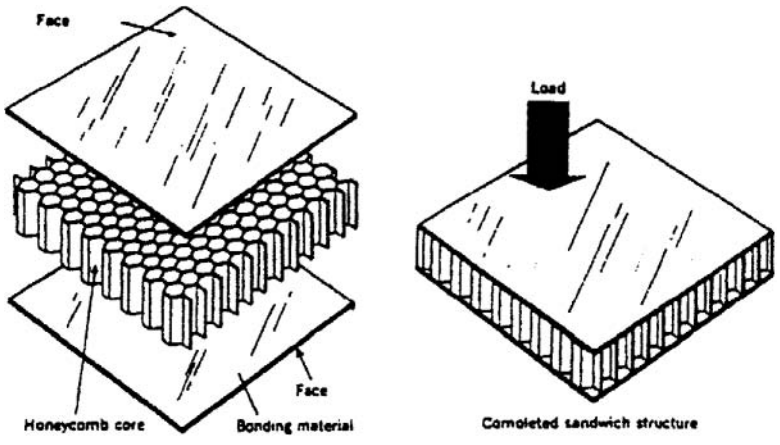


Fig. 11.2.10 Honeycomb core sandwich structure. (Courtesy of Plastics FALLO.)

corotational molding, etc. There is also the so-called structural foam (SF) that is also called integral skin foaming or reaction injection molding. It can overlap in lower performance use with the significantly larger market of the more conventional sandwich. Up until the 1980s in the U.S., the RIM and SF processes were kept separate. Combining them in the marketplace was to aid in market penetration. During the 1930s to 1960s, liquid injection molding (LIM) was the popular name for what later became RIM and SF (Chapter 11.1).

These structures are characterized as a plastic structure with nearly uniform density foam core and integral near-solid skins (facings). When these structures are used in load-bearing applications, the foam bulk density is typically 50–90% of the plastic's unfoamed bulk density. Most SF products (90 wt%) are made from different TPs, principally PS, PE, PVC, and ABS. Polyurethane is the primary TS plastic. Unfilled and reinforced SFs represents about 70% of the products. The principal method of processing (75%) is modified low-pressure injection molding. Extrusion and RIM account for about 10% each.

In a sandwich design, overall proportions of structures can be established to produce an optimization of face thickness and core depth which provides the necessary overall strength and stiffness requirements for minimum cost of materials, weight of components, or other desired objectives. Competing materials should be evaluated on the basis of optimized sandwich section properties that take into account both the structural properties and the relative costs of the core and facing materials in each combination under consideration. For each combination of materials being investigated, thickness of both facings and core should be determined to result in a minimum cost of a sandwich design that provides structural and other functional requirements.

Sandwich configurations are used in small to large shapes. They generally are more efficient for large components that require significant bending strength and/or stiffness. Examples of these include roofs, wall and floor panels, large shell components that are subject to compressive buckling, boat hulls, truck and car bodies, and cargo containers. They also provide an efficient solution for multiple functional requirements such as structural strength and stiffness combined with good thermal insulation, or good buoyancy for flotation.

Sandwich materials can be analyzed as composite structures. Structures and sub-structures composed of a number of different components and/or materials, including traditional materials, obey the same principles of design analysis. Stresses, strains, and displacements within individual components must be related through the characteristics (anisotropy, viscoelasticity, etc.) relevant to the particular material; also loads and displacements must be compatible at component interfaces. Thus, each individual component or sub-component must be treated using the relevant methods.

Load and support conditions for individual components depend on the complete structure (or system) analysis. For example, if a panel is mounted into a much more rigid structure, then its support conditions can be specified with acceptable accuracy. However, if the surrounding structure has comparable flexibility to the panel, then the interface conditions will depend on the flexural analysis of the complete structure. In a more localized context, structural stiffness may be achieved by ribbing, and relevant analyses may be carried out using available design formulas (usually for elastic behavior) or FEA. But necessary anisotropy or viscoelasticity complicate the analysis, often beyond the ability of the design analyst.

Primary structural role of the face/core interface in sandwich construction is to transfer transverse shear stresses between faces and core. This condition stabilizes the faces against rupture or buckling away from the core. It also carries loads normally applied to the panel surface. They resist transverse shear and normal compressive and tensile stress resultants. For the most part, the faces and core that contain all plastics can be connected during a wet lay-up molding or, thereafter, by adhesive bonding. In some special cases, such as in a truss-core pipe, faces and core are formed together during the extrusion process, resulting in an integral homogeneous bond/connection between the components. Fasteners are seldom used to connect faces and core because they may allow erratic shear slippage between faces and core or buckling of the faces between fasteners. Also, they may compromise other advantages such as waterproofing integrity and appearance.

For RP-faced sandwich structures the design approaches includes both the unique characteristics introduced by sandwich construction and the special behavior introduced by RP materials. The overall stiffness provided by the interaction of the faces, the core, and their interfaces must be sufficient to meet deflection and deformation limits set for the structures. Overall stiffness of the sandwich component is also a key consideration in design for general instability of elements in compression.

In a typical sandwich constructions, the faces provide primary stiffness under in-plane shear stress resultants (N_{xy}), direct stress resultants (N_x, N_y), and bending stress resultants (M_x, M_y). Also as important, the adhesive and the core provide primary stiffness under normal direct stress resultants (N_z), and transverse shear stress resultants (Q_x, Q_y). Resistance to twisting moments (T_{xz}, T_{yz}), which is important in certain plate configurations, is improved by the faces. Capacity of faces is designed not to be limited by either material strength or resistance to local buckling.

The stiffness of the face and core elements of a sandwich composite must be sufficient to preclude local buckling of the faces. Local crippling occurs when the two faces buckle in the same mode (antisymmetric). Local wrinkling occurs when either or both faces buckle locally and independently of each other. Local buckling can occur under either axial compression or bending compression. Resistance to local buckling is developed by an interaction between face and core that depends upon the stiffness of each.

With the structural foam (SF) construction, large and complicated parts usually require more critical structural evaluation to allow better prediction of their load-bearing capabilities under both static and dynamic conditions. Thus, predictions require careful analysis of the structural foam's cross section.

The composite cross section of an SF part contains an ideal distribution of material, with a solid skin and a foamed core. The manufacturing process distributes a thick, almost impervious solid skin that is in the range of 25% of overall wall thickness at the extreme locations from the neutral axis where maximum compressive and tensile stresses occur during bending.

When load is applied flatwise the upper skin is in compression while the lower one is in tension, and a uniform bending curve will develop. However, this happens only if the shear rigidity or shear modulus of the cellular core is sufficiently high. If this is not the case, both skins will deflect as independent members, thus eliminating the load-bearing capability of the composite structure. In this manner of applying a load the core provides resistance against shear and buckling stresses as well as impact (Figure 11.2.11). There is an optimum thickness that is critical in designing this structure.

When the SF cross section is analyzed, its composite nature still results in a twofold increase in rigidity, compared to an equivalent amount of solid plastic, since rigidity is a cubic function of wall thickness. This increased rigidity allows large structural parts to be designed with only minimal distortion and deflection when stressed within the recommended values for a particular core material. When analyzing rigidity and the moment of inertia (I) can be evaluated three ways. In the first approach, the cross section is considered to be solid material (Figure 11.2.12).

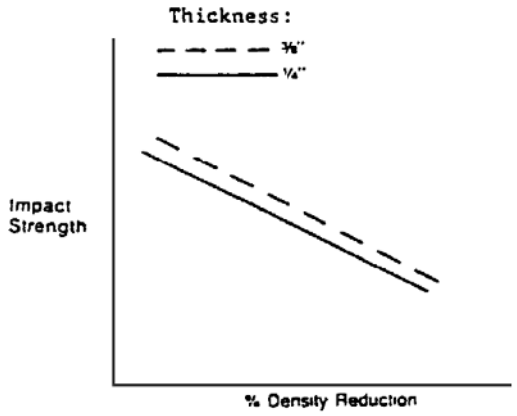


Fig. 11.2.11 Core thickness vs. density impact strength.

The moment of inertia (I_x) is then equal to:

$$I_x = bh^3/12 \tag{11.2.14}$$

where b is the width and h the height.

This commonly used approach provides acceptable accuracy when the load-bearing requirements are minimal. An example is the case of simple stresses or when time and cost constraints prevent more exact analysis.

The second approach ignores the strength contribution of the core and assumes that the two outer skins provide all the rigidity (Figure 11.2.13).

The equivalent moment of inertia is then equal to:

$$I_x = b(h^3 - h^3/12) \tag{11.2.15}$$

This formula results in conservative accuracy, since the core does not contribute to the stress-absorbing function. It also adds a built-in safety factor to a loaded beam or plate element when safety is a concern.

A third method is to convert the structural foam cross section to an equivalent I-beam section of solid resin material (Figure 11.2.14).

The moment of inertia is then formulated as:

$$I_x = [bh^3 - (b - b_1)(h - 2t_x)^3]/12 \tag{11.2.16}$$

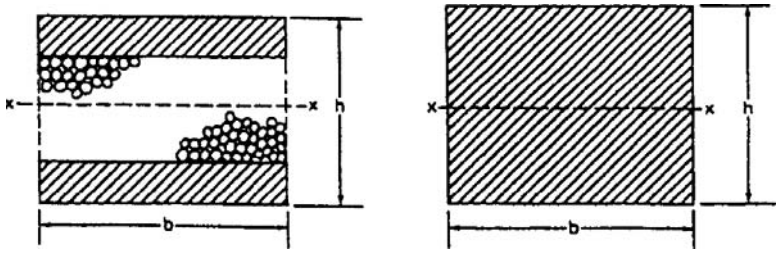


Fig. 11.2.12 Sandwich and solid material construction.

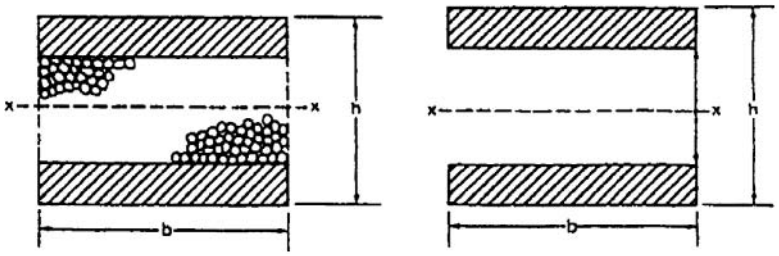


Fig. 11.2.13 Sandwich cross section with and without a core.

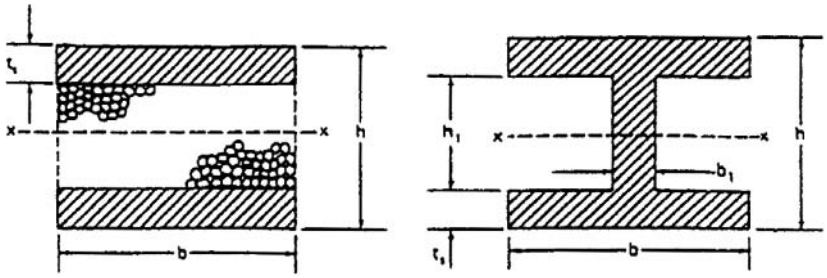


Fig. 11.2.14 Sandwich and I-beam cross section.

where $b_1 = b(E_c)/(E_s)$, E_c is the modulus of core, E_s the modulus of skin, t_s the skin thickness, and h_1 the core height.

This approach may be necessary where operating conditions require stringent load bearing capabilities without resorting to overdesign and thus unnecessary costs. Such an analysis produces maximum accuracy and would, therefore, be suitable for FEA on complex parts. However, the one difficulty with this method is that the core modulus and the as-molded variations in skin thicknesses cannot be accurately measured.

The following review relates to the performance of sandwich constructions such as those with RP skins and honeycomb core. For an isotropic material with a modulus of elasticity (E), the bending stiffness factor (EI) of a rectangular beam b wide and h deep is

$$EI = E(bh^3/12) \tag{11.2.17}$$

A rectangular structural sandwich with the same dimensions whose facings and core have moduli of elasticity E_f and E_c, respectively, and a core thickness c, the bending stiffness factor EI becomes:

$$EI = (E_f b/12)(h^3 - c^3) + (E_c b/12)c^3 \tag{11.2.18}$$

This equation is OK if the facings are of equal thickness, and approximate or approximately equal, but the approximation is close if the facings are thin relative to the core. If, as is usually the case, E_c is much smaller than E_f, the last term in the equation is deleted.

Asymmetrical sandwich structures with different materials or different thicknesses in their facings, or both, the more general equation for EI may be used. With isotropic materials, the shear modulus G is high compared to the elastic modulus E, and the shear distortion of a transversely loaded beam is so small that it can be neglected in calculating deflection. Sandwich core shear modulus G is usually so much smaller than E_f of the facings that the shear distortion of the core may be large and therefore contribute significantly to the deflection of a transversely loaded beam. The total deflection of a sandwich beam involves the two factors of the deflection caused by the bending moment alone and the deflection caused by shear, that is,

$$\delta = \delta_m + \delta_s \tag{11.2.19}$$

where δ is the total deflection, δ_m the moment deflection, and δ_s the shear deflection.

Under transverse loading, bending moment deflection is proportional to the load and the cube of the span and inversely proportional to the stiffness factor, EI.

11.2.9 Gear

Designing gears can be very complex since many interfacing load factors are involved. There are bending, shear, rolling, tension, and sliding stresses all acting upon a mechanism whose purpose is to transmit uniform motion and power. This situation is well understood by those designing gears. For over a century plastic gears have been extensively used in all industries worldwide and with time passing the plastic industry has provided lightweight and quieter operating gears. They provide

a means of cutting cost, weight, and noise without reducing performance.

Information on designing gears is extensive. Knowledge of gear fundamentals is a prerequisite for the understanding of applying appropriate plastic information into the gear formulas so that the application results in favorable operation. Textbooks, technical handbooks, and industrial literature of gear suppliers provide information such as teeth load requirements, transmitting motion and power by means of gears, their construction, and detail performance requirements. Reviews on teeth of heavily loaded gears require tip relief to reduce effects of deflection, and have full fillet radii to reduce stress concentrations. If the pinion in a pair of gears has a small number of teeth, undercutting may result. Undercutting weakens teeth, causes undue wear, and may affect continuity of action.

Designing gears involves ribbing, coring, and shaping material. When going from a metal to a plastic gear it is redesigned to meet the behaviors of plastics. The popular injection molded plastic gear saves material, eliminates stresses from having thick and thin sections, provides uniform shrinkage in teeth and the remainder of the gear, and provides a full load-carrying capacity for the teeth. Plastic gears are dimensioned so that they will provide sufficient backlash at the highest temperatures likely to be encountered in service.

Wear, scoring, material flow, pitting, fracture, creep, and fatigue cause plastic and metal gears to fail. Continuous lubrication can increase the allowable bending stress by a factor of at least 1.5. However there are plastics (acetals, nylons, fluoropolymers, and others) that operate efficiently with no lubrication. There are plastics with wear resistance and durability of plastic gears makes them exceptionally useful.

The bending stress in engineering TPs is based on fatigue tests run at specific pitch-line velocities. A velocity factor should be used if the operating pitch-line velocity exceeds the test speed. Plastic gears should have a full fillet radius at the tooth root, so they are not subjected to stress concentration as are metal gears. If a gear is lubricated, bending stress will be important to evaluate. As with bending stresses, calculating surface-contact stress requires using a number of correction factors. As an example, a velocity factor is used when the pitch-line velocity exceeds the test velocity. A correction factor is also used to account for changes in operating temperature, gear materials, and the pressure angle. Stalled torque, another important factor, could be considerably more than the normal loading torque.

A damaging situation for gears is to operate over a specified temperature for the plastic used. Reducing the rate of heat generation or increasing the rate of heat transfer will stabilize the gear's temperature so that they will run indefinitely until stopped by fatigue failure. Using unfilled engineering plastics usually gives them a fatigue life on an order of magnitude higher than metal gears.

Plastic gears are subject to hysteresis heating, particularly at high speeds. If the proper plastic is not used to meet the gears service requirements the hysteresis heat may be severe enough that the plastic melts. Avoid this failure by designing the gear drive so that there is favorable thermal balance between the heat that is generated and that which is removed by cooling processes. Hysteresis heating in plastics can be reduced by several methods, the usual one being to reduce the peak stress by increasing the tooth root area available for torque transmission. Another way to reduce stress on the teeth is by increasing the gear's diameter.

Materials used such as stiffer plastics can reduce hysteresis heating. Crystalline TPs for example (the popularly used acetal and nylon) can be stiffened by 25–50% with the addition of fillers and reinforcements. Others used include ABS, polycarbonates, polysulfones, phenylene oxides, polyurethanes, and TS polyesters. Additives, fillers, and reinforcements are used in plastics gears to meet different performance requirements (Chapter 11.1), Examples include glass fiber for added strength, and fibers, beads, and powders for reduced thermal expansion and improved dimensional stability. Other materials, such as molybdenum disulfide, polytetrafluoroethylene (PTFE), and silicones, may be added as lubricants to improve wear resistance.

Choice of plastics gear material depends on requirements for size and nature of loads to be transmitted, speeds, required life, working environment, type of cooling, lubrication, and operating precision. The strength of these TPs varies with temperature. If the incorrect plastic is used, the higher temperatures can reduce root stress and permit tooth deformation. In calculating power to be transmitted by spur, helical, and straight bevel gearing, the following formulas should be used with the factors given in Table 11.2.5.

For internal and external spur gears:

$$\mathrm{HP} = \frac{S_S FYV}{55(600+V)PC_S} \tag{11.2.20}$$

For internal and external helical gears,

$$\mathrm{HP} = \frac{S_S FYV}{423(78+\sqrt{V})P_n C_S} \tag{11.2.21}$$

For straight bevel gears,

$$\mathrm{HP} = \frac{S_S FYV(C-F)}{55(600+V)PCC_S} \tag{11.2.22}$$

where S is the safe stress in bending (Table 11.2.5a); F the width in inches; Y the tooth form factor (Table 11.2.5b); C the pitch cone distance in inches;

Table 11.2.5 Plastic gear (a) safe bending stress (psi), (b) tooth form examples of Y factors, and (c) service factors

(a)

	Safe stress	
Plastics type	**Unfilled**	**Glass-filled**
ABS	3000	6000
Acetal	5000	7000
Nylon	6000	12,000
Polycarbonate	6000	9000
Polyester	3500	8000
Polyurethane	2500	–

(b)

				20-deg internal full depth	
Number of teeth	**14½-deg involute or cycloidal**	**20-deg full depth involute**	**20-deg stub tooth involute**	**Pinion**	**Gear**
12	0.210	0.245	0.311	0.327	–
13	0.220	0.261	0.324	0.327	–
14	0.226	0.276	0.339	0.330	–
15	0.236	0.289	0.348	0.330	–
16	0.242	0.259	0.361	0.333	–
17	0.251	0.302	0.367	0.342	–
18	0.261	0.308	0.377	0.349	–
19	0.273	0.314	0.386	0.358	–
20	0.283	0.320	0.393	0.364	–
21	0.289	0.327	0.399	0.371	–
22	0.292	0.330	0.405	0.374	–
24	0.298	0.336	0.415	0.383	–
26	0.307	0.346	0.424	0.393	–
28	0.314	0.352	0.430	0.399	0.691
30	0.320	0.358	0.437	0.405	0.679
50	0.352	0.480	0.474	0.437	0.613
100	0.371	0.446	0.506	0.462	0.565
150	0.377	0.459	0.518	0.468	0.550
300	0.383	0.471	0.534	0.478	0.534
Rack	0.390	0.484	0.550	–	–

(*Continued*)

Table 11.2.5 Plastic gear (a) safe bending stress (psi), (b) tooth form examples of Y factors, and (c) service factors—cont'd

(c)

Type of load	8–10 h/day	24 h/day	Intermittent 3 h/day	Occasional ½ h/day
Steady	1.00	1.25	0.80	0.50
Light shock	1.25	1.5	1.00	0.80
Medium shock	1.5	1.75	1.25	1.00
Heavy shock	1.75	2.00	1.5	1.25

C_S the service factor (Table 11.2.5c); P the diarnetral pitch; P_n the normal diametral pitch; and V the velocity at pitch circle diameter in ft/min.

The surrounding condition, whether liquid, air, or oil (most efficient) will have substantial cooling effects. A fluid like oil is at least ten times better at cooling than air. Agitating these mediums increases their cooling rates, particularly when employing a cooling heat exchanger.

Methods of fabricating gears involve cutting/hobbing from processed blocks or sheet plastics, compression molding laminated (W) material, or the most popular injection molding. Use is made of unfilled and filled or reinforced laminated TPs or TSs. Phenolic laminated gears are in a class of their own. One can make all the perfect calculations and insert the necessary values for plastic gears, but if molding conditions and molding materials are not processed properly one may end up with mediocre or even unsatisfactory results.

Being not as strong as steel, plastics perform far closer to their design limits than do metal gears. Metal and plastic in gear design differ. Designs for metal are based on the strength of a single tooth, but plastic shares the load among the various gear teeth to spread it out. In plastics the allowable stress for a specific number of cycles to failure increases as the tooth size decreases to a pitch of about 48. Very little increase is seen above a 48 pitch, because of the effects of size and other considerations.

11.2.9.1 Contact stress

Stresses caused by the pressure between two elastic contacting parts are of importance in design such as gears and bearings. Centuries ago Hertz developed the mathematical theory for the surface stresses and the deformation produced by pressure between curved parts, and the results of this analysis are supported by research. Formulas based on this theory give the maximum compressive stresses that occur at the center of the surfaces of contact, but do not consider the maximum subsurface shear stresses or the maximum tensile stresses that occur at the boundary of the contact area.

11.2.10 Bearing

Plastic used in bearings, as in gears, have many success stories because certain plastics have the required properties to meet different product application performances. Similarly to plastic gears, plastic bearings have a long history of successful performance. Small to large bearings are used in applications requiring light to heavy duty. Some TPs have inherent lubricating characteristics and additives such as molybdenuni disulfide, polytetrafluoroethylene (PTFE), and others enhance other TPs as well as TSs. Other TPs, as well as TSs, by the addition of PTFE and/or molybdenum disulfide, become excellent candidates for bearing materials.

There are high-performance laminated (RP) fabric, bonded with phenolic plastic incorporating antifiction ingredients. They have given excellent service when properly applied in various applications particularly in the past. This group of bearings has a low coefficient of friction, antiscoring properties, and adequate strength for use in steel mills and other heavy-duty applications.

11.2.10.1 PV factor

Bearings are designed to keep their frictional heat at a low value and have conditions that lead to dissipation of such generated heat. Major heat contributors are the magnitude pressures P exerted on the projected area of the bearing and the velocity V or the speed of the rotating bearing. Experience has set limits on this PV value. Limits within PV factors have been developed for specific plastics that provide the industries with successful bearings. Other heat contributors are coefficient of friction of mating materials, lubrication, clearances between bearing and shaft, rusted shaft, surrounding temperature, surface finish, hardness of the mating materials, contaminants, and bearing wall thickness that relates to heat dissipation.

The basic rule is that neither the pressure nor the velocity should exceed a value of 1000 (psi or ft/min). As an example with a PV limit for acetal of 3000, the PV factor could be 1000 ft/min × 3 pounds or 1000 pounds × 3 ft/min at the extreme, provided heat conditions resulted in uniform rate of wear. The coefficient of friction data, available from suppliers, can also provide guidelines to the efficiency in comparing the different materials.

The limit of the PV factor for each material or the internally lubricated materials for the constant wear of bearing is usually available from the supplier of the plastic. Lubrication whether incorporated in a plastic or provided by feeding the lubricant to the bearing will raise the PV limit 2.5 or more times over the dry system.

11.2.11 Grommet

Damping designed products may be required. As an example, large flat areas may require damping so that they do not act like loudspeakers. The damping action can quiet a cabinet that resonates. Various plastics have helped alleviate problems in all types of noisemakers. Different damping approaches can be used, such as applying plastic foam sound insulator or plastic panels that have low damping characteristics.

A popular approach is to use plastic grommets where applicable. Sound-absorbing grommets are used on equipment such as motors' bolt attachments and trash compactors. Testing and all other types of equipment can take advantage of grommets or be redesigned to use plastic. Grommets provide their greatest noise reduction through damping in the octave frequency bands above 500 Hz where the ear is most sensitive and sound most annoying.

Grommets have replaced assembled linkages (Figure 11.2.15). In addition to reducing noise, the usual injection molded polyurethane (PUR) grommet eliminated the time consuming/costly metal assembly. During assembly it is snapped into a hole in the steel lever, then a grooved rod is inserted into the grommet. The grommet isolates vibration from the metal parts and eliminates the hardening and cracking that used to shorten the life of the old assembly. The plastic design mechanically is at least as strong as the metal assembly that includes withstanding high-load pull-outs and can withstand high cyclic loads applied at different degrees off the rod axis at temperatures up to 300 °F (149 °C).

To quiet a noise-generating mechanism, the first impulse is often to enclose it. Enclosure can be the best solution, but not always. By determining what is causing the noise, appropriate action can be taken to be more specific and provide a cost-effective fix. A plastic enclosure can be used to suppress noise. Recognize that with a metal enclosure a small noise is transmitted to the metal structure that serves to amplify the sound.

11.2.12 Gasket

Different plastics are used to meet different gasket and seal requirements. An example is PTFE (polytetrafluoroethylene) that provides a virtually inert exposure to all kinds of elements and has outstanding high-temperature performance. Unreinforced PTFE can be vulnerable to creep and stress relaxation. However, reinforced PTFE modifies these limitations.

Gaskets are designed to meet different requirements such as retaining loads or meeting stress relaxation requirements, chemical or heat resistance, severe environments, and containment of liquids, greases, and so on. There are different industry tests and standards to meet many different service requirements. There are tests for applying compressive stress simulating the way

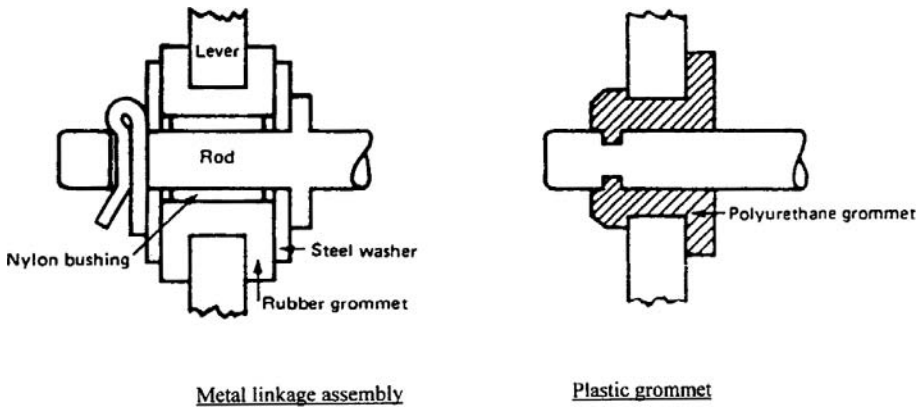


Fig. 11.2.15 Grommet replaces a five individual metal assembly. (Courtesy of Mobay/Bayer.)

many gaskets and seals are stressed in service (ASTM F 38). They measure the effects of such pertinent variables as stress relaxation in regard to time and environment.

Stress analysis is used to determine their capability to seal against leakage resulting from the pressure of a confined fluid. Generally high pressures or stress relates to increasing the tendency of a gasket to creep. Stress relaxation data to the designer provides a guide to developing a suitable design compromise, without overdesigning. These data show that the thinner the gasket, the less stress relaxation occurs. A gasket is redesign to be stronger and more expensive in construction. In material evaluations, stress relaxation can be related with geometric variables by means of a shape factor such as

$$\text{Shape factor} = \text{Annular area/total lateral area} = (OD - ID)/4t \qquad (11.2.23)$$

where OD is the outside diameter; ID the inside diameter; and t the thickness.

The trend of this factor is generally consistent with plastics' behaviors as reviewed in Chapter 11.1. The relaxation-test data has to be determined for the plastic to be used. Individual behavior of one plastic is usually different when compared to another plastic.

11.2.13 Shape

11.2.13.1 Overview

In addition to what has been reviewed in meeting structural requirements, analyses of product shapes also includes factors such as the size available equipment can handle including thicknesses and product complexity and capability to package and ship to the customer (Chapter 11.1). The ability to achieve specific shapes and design details is dependent on the way the process operates and plastics to be processed. Generally the lower the process pressure, the larger the product that can be produced. With most labor-intensive fabricating methods, such as RP hand lay-up with TS plastic there is virtually no limit on size.

An important requirement concerns meeting dimensional tolerances of shaped products. Reported are different shrinkages for different plastics per standard tests that may have a relation to the designed product. The probability is that experience with prototyping will only provide the true shrinkage conditions of the shaped products. Minimum shrink values are included in the design of mold cavities and die openings so that if the processed plastic does not meet required dimensions all that is required is to cut the metal in the tools.

If the reverse occurs, expensive tool modifications may be required, if not replacing the complete tool. It is vital to set up a complete checklist of product requirements, to preclude the possibility that a critical requirement may be overlooked initially. Fortunately there are occasions where changes in process control during fabrication can be used to produce the required dimensional product.

11.2.13.1.1 Filament wound shape

Filament winding (FW) shapes are principally circular (cylinders, pipes, tubing, etc.) or enclosed vessel (storage tanks, oxygen tanks, etc.). They produce spherical, conical, and geodesic shapes. The fabricating process permits tightly controlled fiber netting orientation and exceptional quality control in different fiber–resin matrix ratios required by design. Structures can be fabricated into shapes such as rectangular or square beams or boxes, longitudinal leaf or coil springs, etc. Filaments can be set up in a part to meet different design stresses.

There are two basic patterns used by industry to produce FW structures, namely, circumferential winding and helical winding. Each winding pattern can be used alone or in various combinations in order to meet different structural requirements. The circumferential winding pattern involves the circumferential winding at about a 90° angle with the axis of rotation interspersed with longitudinal reinforcements. Maximum strength is obtainable in the hoop direction. This type of pattern generally does not permit winding of slopes over 20° when using a wet winding reinforcement or 30° when using a dry winding process. It also does not result in the most efficient structure when end closures are required. With end closures and/or steep slopes, a combination of helical and circumferential winding is used.

With helical winding, the reinforcements are applied at any angle from 25° to 85° to the axis of rotation. No longitudinal filament need be applied because low-winding angles provide the desired longitudinal strength as well as the hoop strength. By varying the angle of winding, many different ratios of hoop to longitudinal strengths can be obtained.

Two different techniques of applying the reinforcements in helical windings are used by industry. One technique is the application of only one complete revolution around the mandrel from end to end. The other technique involves a multi-circuit winding procedure that permits a greater degree of flexibility of wrapping and length of cylinder.

11.2.13.1.2 Netting analysis

Continuous reinforced filaments should be used to develop an efficient high-strength to low-weight FW structure. Structural properties are derived primarily

from the arrangement of continuous reinforcements in a netting analysis system in which the forces, owing to internal pressure, are resisted only by pure tension in the filaments (applicable to internal-pressure systems).

There is the closed-end cylinder structure that provides for balanced netting of reinforcements. Although the cylinder and the ends require two distinctly different netting systems, they may be integrally fabricated. The structure consists of a system of low helix angle windings carrying the longitudinal forces in the cylinder shell and forming integral end closures which retain their own polar fittings. Circular windings are also applied to this cylindrical portion of the vessel, yielding a balanced netting system. Such a netting arrangement is said to be balanced when the membrane generated contains the appropriate combination of filament orientations to balance exactly the combination of loadings imposed.

The girth load of the cylindrical shell is generally two times the axial load. The helical system is so designed that its longitudinal strength is exactly equal to the pressure requirement. Such a low-angle helical system has a limited girth strength. The circular windings are required in order to carry the balance of the girth load.

The end dome design contains no circular windings since the profile is designed to accommodate the netting system generated by the terminal windings of the helical pattern. It is termed an ovaloid; that is, it is the surface of revolution whose geometry is such that fiber stress is uniform throughout and there is no secondary bending when the entire internal pressure is resisted by the netting system.

There is the ovaloid netting system that is the natural result of the reversal of helical windings over the end of the vessel. The windings become thicker as they converge near the polar fittings. In order to resist internal pressure by constant filament tension only, the radius of curvature must increase in this region. It can also be equal to one half the cylinder radius when the helix angle $\alpha = 0°$, and equal to the cylinder radius when $\alpha = 45°$. The profile will also be affected by the presence of an external axial force.

In the application of bidirectional patterns, the end domes can be formed by fibers that are laid down in polar winding patterns. The best geometrical shape of the dome is an oblated hemispheroid. Theoretically, the allowable stress level in the two perpendicular directions should be identical. However, the efficiency of the longitudinal fibers is less than that of the circumferential fibers. It is possible to estimate an optimum or length-to-diameter ratio of a cylindrical case for a given volume.

The filament-wound (FW) sphere design structure provides another example of a balanced netting analysis system. It is simpler in some respects than the closed-end cylinder. The sphere must be constructed by winding large circles omni-directionally and by uniform distribution over the surface of the sphere. In practice, distribution is limited so that a small polar zone is left open to accommodate a connecting fitting.

The netting pattern required generates a membrane in which the strength is uniform in all directions. The simplest form of such a membrane would have its structural fibers running in one direction and the other half at right angles to this pattern. This layup results in the strength of the spherical membrane being one-half of the strength of a consolidated parallel fiber system.

The oblated spheroid design structure relates to special spherical shapes. Practical design parameters have shown that the sphere is the best geometric shape when compared to a cylinder for obtaining the most efficient strength-to-weight pressure vessel. The fiber RPs is the best basic constituents. Certain modifications of the spherical shape can improve the efficiency of the vessel. One modification involves designing the winding pattern of the fibers so that unidirectional loading can be maintained. In this type of structure, it is generally assumed that the fibers are under equal tension. This type of structure is identified as an isotensoid. The geometry of this modified sphere is called oblated spheroid, ovaloid, or ellipsoid.

The term isotensoid identifies a pressure vessel consisting entirely of filaments that are loaded to identical stress levels. The head shape of an isotensoid is given by an elliptical integral, which can most readily be solved by a computer. Its only parameter is the ratio of central opening to vessel diameter. This ratio determines the variation of the angle of winding for the pressure vessel. During pressurization the vessel is under uniform strain; consequently, no bending stresses or discontinuity stresses are induced.

A short polar axis and a larger perpendicular equatorial diameter characterize the vessel. The fibers are oriented in the general direction of a polar axis. Their angle with this axis depends on the size of the pole openings (end closures). For glass fiber-TS polyester RP vessels, levels of 200,000 psi (1.4 GPa) can be obtained.

The toroidal design structure is a pressure vessel made with two sets of filaments symmetrically arranged with respect to the meridians. They meet two basic requirements: (1) static equilibrium at each point, which determines the angle between the two filaments, and (2) stability of the filaments on the surface, which requires the filaments to follow geodesic paths on the surface. When the equation of the surface is given, these two requirements are generally incompatible. One way to reconcile the correct angularity of the filaments (equilibrium) with the correct paths of the filaments (stability) is to take some freedom in determining the geometry of the surface.

11.2.13.2 Cylinder

Standard engineering analysis can be used. Consider a cylinder of inside radius r, outside radius R, and length L containing a fluid under pressure p. The circumferential or hoopwise load in the wall (t, thickness) is proportional to the pressure times radius $= pr$, and the hoop stress:

$$f_h = \text{hoopwise load/cross-sectional area} = pr/t \text{ or } pd/2t \quad (11.2.24)$$

similarly, the longitudinal stress:

$$f_l = pd/4t \quad (11.2.25)$$

assuming $\pi(R^2 - r^2) = 27\pi rt$ for a thin-walled tube.

Thus, this condition of the hoop stress being twice the longitudinal stress is normal for a cylinder under internal pressure forces only. The load in pounds acts on the tube at a distance from one end and a bending moment M is introduced. This produces a bending stress in the wall of the cylinder of

$$f_b = My/I \quad (11.2.26)$$

where $y = R$ and $I =$ moment of inertia. For a cylinder:

$$I = \pi(D^4 - d^4)/64 \text{ in.}^4 \quad (11.2.27)$$

This stress must then be considered in addition to the longitudinal stress already presented because of internal pressure. If the end closures are in the form of flat plates, bending stresses due to the internal pressure are introduced as

$$F_b = 1.25\ (pr^2/t_I) \quad (11.2.28)$$

where t_I is the thickness of end. This necessitates the wall of a flat disk-type end being extremely thick compared with a hemispherical end which is found to be the most efficient shape where the stress in the wall is:

$$pd/4t \quad (11.2.29)$$

Figure 11.2.16 compares the thicknesses and corresponding volumes of the two types of ends for varying values of r (assuming $p = 2000$ psi and ultimate stress in the wall material of 100,000 psi).

The volume of the flat end is found to be approximately four times the volume of the hemispherical end for any given radius of tube, resulting in increased weight and material cost. Other end shapes such as ellipse will have a volume of weight somewhere between the two, depending on the actual shape chosen.

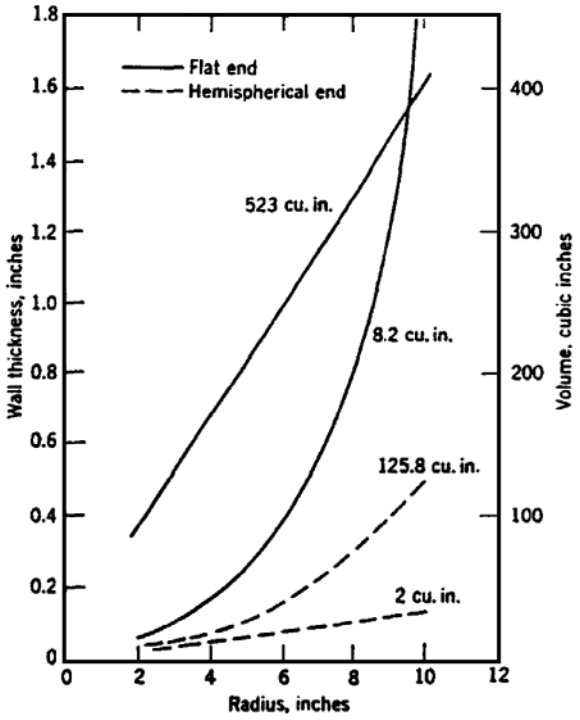


Fig. 11.2.16 Cylinder comparison of thickness for a flat end and a hemispherical end.

11.2.13.3 Sphere

Circumferential load in the wall of all the spheres under internal pressure is equal to the pressure times the internal cross-sectional area, and the hoop stress, using the previous engineering assumption, is found to be

$$f_h = pd/4t \quad (11.2.30)$$

It will readily be seen that no matter which section is chosen, provided the plane of the section passes through the center of the sphere, the condition will be the same, and it can be said that the hoop stress will be the same in all directions.

When it is assumed that

$$\pi(R^2 - r^2) = 27\pi rt \quad (11.2.31)$$

it is determined that for wall thicknesses up to approximately 3 in., the error is negligible. It can also be determined that the percent of error decreases as the ratio r/t increases. Table 11.2.6 provides size vs. weight of RP spheres.

For a sphere with the stresses uniform in all directions, it follows that the fibers require equal orientation in all directions. The problem of orientation resolves itself purely into one of practical application of the fibers. In the cylinder, the fibers are specifically oriented to meet

Table 11.2.6 Data for 3000 psi glass fiber-epoxy spheres

Capacity, cubic inches	Outside diameter, inches	Nominal weight, pounds	Maximum weight, pounds	Overall length, inches
50	5¼	1.50	1.60	5⅝
100	6½	2.50	2.63	6 15/16
200	8⅛	4.44	4.62	8½
300	9¼	6.25	6.56	9⅝
400	10⅛	8.06	8.48	10 9/16
500	10 15/16	9.87	10.35	11 5/16
650	11 15/16	12.56	13.18	12 5/16
880	13 1/16	16.00	16.80	13⅜
1070	14	20.06	21.07	14⅜
1325	15⅛	24.50	25.73	15 7/16
1575	15⅞	28.75	30.15	16¼
1800	16⅝	32.75	34.42	16 15/16
2500	18½	44.81	47.06	18⅞
3200	20	56.50	59.32	20 7/16

any condition of stressing. The simplest method of doing this is to employ a single helical pattern.

Theory shows that this is highly sensitive to variations in the longitudinal hoop-stress ratio and also the accuracy of the angle wound. The addition of pure hoop windings to the helix gives a theoretical gain in stability with no loss of strength or efficiency. In order to develop the most satisfactory orientation, the winding is performed so those two different helix angles are used.

11.2.13.4 Tank

11.2.13.4.1 Fabricating RP tank

Classical engineering stress analysis shows that hoop stress (stress trying to push out the ends of the tank) is twice that of longitudinal stress. To build a tank of conventional materials (steel, aluminum, etc.) requires the designer to use sufficient materials to resist the hoop stresses that results in unused strength in the longitudinal direction. In RP, however, the designer specifies a laminate that has twice as many fibers in the hoop direction as in the longitudinal direction.

An example is a tank 0.9 m (3 ft) in diameter and 1.8 m (6 ft) long with semi-spherical ends. Such a tank's stress calculations (excluding the weight of both the product contained in it, and the support for the tank) are represented by the formulas

$$s = pd/2t \text{ for the hoop stress} \tag{11.2.32}$$

and

$$s = pd/4t \text{ for the end and longitudinal stresses} \tag{11.2.33}$$

where s is the stress, p the pressure, d the diameter, and t the thickness.

Tensile stresses are critical in tank design. The designer can assume the pressure in this application will not exceed 100 psi (700 Pa) and selects a safety factor of 5. The stress must be known so that the thickness can be determined. The stress or the strength of the final laminate is derived from the makeup and proportions of the resin, mat, and continuous fibers in the RP composite material.

Representative panels must be made and tested, with the developed tensile stress values then used in the formula. Thus, the calculated tank thickness and method of lay-up or construction can be determined based on

$$t_h = pd/2s_h \quad \text{or} \quad t_h = pd/4s_h \tag{11.2.34}$$

where

$$t_h = \frac{100 \times 3 \times 12}{2 \times (20 \times 10^3)/5} = 0.450 \text{ in.} \qquad (11.2.35)$$

$t_l = 1/2\ t_h = 0.225$ in. (or the same thickness with half the load or stress); t_h is the hoop thickness; t_l the longitudinal thickness; s_h the hoop stress; s_l the longitudinal stress; $s_h = 20 \times 10^3$ psi (140 MPa); safety factor = 5; $p = 100$ psi (700 Pa); $d = 3$ ft (0.9 m).

If the stress values had been developed from a laminate of alternating plies of woven roving and mat, the lay-up plan would include sufficient plies to make 1 cm (0.40 in.) or about four plies of woven roving and three plies of 460 g/m^2 (1½ oz) mat. However, the laminate would be too strong axially. To achieve a laminate with 2 to 1 hoop to axial strength, one would have to carefully specify the fibers in those two right angle directions, or filament wind the tank so that the vector sum of the helical wraps would give a value of 2 (hoop) and 1 (axial), or wrap of approximately 54° from the axial.

Another alternative would be to select a special fabric whose weave is 2 to 1, wrap to fill, and circumferentially wrap the cylindrical sections to the proper thickness, thus getting the required hoop and axial strengths with no extra, unnecessary strength in the axial (longitudinal) direction, as would inevitably be the case with a homogeneous metal tank.

As can be seen from the above, the design of RP products, while essentially similar to conventional design, does differ in that the materials are combined when the product is manufactured. The RP designer must consider how the load-bearing fibers are placed and ensure that they stay in the proper position during fabrication.

11.2.13.4.2 Underground storage tank

GFRP (glass fiber-TS polyester RP) underground tanks for storing gasoline and other materials have been in use worldwide since at least the 1950s. Experience with them initiated many tank standards for different materials. RPs provided much longer life than their steel counterparts. In fact, steel tanks previously had no "real life" or no requirement standards until the RPs entered the market. It has been estimated that more than 200,000 GFRP tanks were installed in the USA from 1960–1990. A previous study by the Steel Tank Institute (Lake Zurich, IL) reported 61% were of steel and 39% of GFRP. At present, at least 50% of all tanks are GFRP, This RP vs. steel debate escalated when the EPA gave service stations and fleet refueling areas 10 years to remove steel tanks that leaked.

Historically a Chicago service station documented the long life of RPs. A May 1963 installation remained leak tight and structurally sound when unearthed in May 1988. After testing the vessel, engineers buried it at another gas station. This tank was one of 60 developed by Amoco Chemical Co. It was fabricated in two semi-cylinder sections of glass fiber woven roving and chopped strand mat impregnated by an unsaturated isophthalic TS-polyester resin selected for its superior resistance to acids, alkalis, aromatics, solvents, and hydrocarbons. Two sections were bonded to each other and to end caps with RP lap joints. Today, the tanks are fabricated by using chopped glass fiber mixed with the isophthalic resin. This mixture is dropped from above onto a rotating steel mandrel. The glass–resin mix is sprayed to make the end caps.

Demand for this type of petroleum storage tank has grown rapidly as environmental regulations have become more stringent. Marina installation have taken advantage of these RP tanks. They permit for boat owners to purchase gasoline at the pier. Before they were installed, gasoline either had to be carried to the marina or purchased elsewhere, because of corrosive conditions underground for metal or other tanks, particularly ones next to salt water.

Standards require that today's underground tanks must last thirty or more years without undue maintenance. To meet these criteria, they must be able to maintain structural integrity and resist the corrosive effects of soil and gasoline, including gasoline that has been contaminated by moisture and soil. The tank just mentioned that was removed in 1991 met these requirements, but two steel tanks unearthed from the same site at that time failed to meet them. One was dusted with white metal oxide and the other showed signs of corrosion at the weld line. Rust had weakened this joint so much that it could be scraped away with a pocketknife. Tests and evaluations were conducted on the RP tank that had been in the ground for 25 years; tests were also conducted on similarly constructed tanks unearthed at 51 and 71 years that showed the RP tanks could more than meet the service requirements.

Prior to the development of the GFRP tanks, no standards were required for buried tanks such as loads or loading conditions, minimum depths of earth cover, or structural safety factors were available. At that time, sizes were 22,704–45,408 l (6000–12,000 gal), with a nominal width and height of 2.44 m (8 ft) for truck shipments to local gasoline stations. Standards have developed listing requirements for stored fluid type, environmental resistance, minimum earth cover, ground water submerged limits, and surface wheel load over tank. Increasing acceptance of buried GFRP tanks has widened the size range from 2081 to 181,632 l (550 to at least 48,000 gal) and the range in typical diameters from about 1.22 to 3.35 m (4 ft to at least 11 ft).

The tank configuration is basically cylindrical, in order to provide the required design volumes within the established envelope of heights and widths. Length

ranges are from 5.5 to 11 m (18–36 ft); they are well within practical truck shipment limits. A circular shape is required to support the substantial internal and external fluid and earth pressures with good structural efficiency. Other considerations in selection of an efficient configuration are used.

With a vertical axis, tank underlay requires that much less land area than with axis horizontal, but very deep excavation is required where expensive ledge or ground water conditions will frequently be encountered. Both internal and external pressures are large, requiring a substantial increase in wall thickness and rib stiffness, compared with a horizontally placed tank. With axis horizontal, maximum external and internal pressures do not vary with size (length), and can be resisted with economically feasible wall thickness and rib proportions. Tanks underlay a larger ground area. Uniform bedding is more difficult to attain.

Hemispherical shells, low rise dished-shaped heads and flat plate closures were all considered. Hemispherical shells were found structurally very efficient because of good buckling resistances under external fluid and earth pressure, good strength under internal pressure, no requirement for edge ring, and low discontinuity stresses at the junction with the cylinder. Flat end closures result in excessive deflections and large edge bending moments on the cylinder. Sandwich construction could be used to improve structural efficiency of flat ends. Sandwich wall construction was also investigated for attaining necessary buckling resistance of spherical shell end closures and found to be feasible but less cost effective.

Use of rib stiffening was required. It was found necessary to stiffen the cylindrical shell against buckling under external pressure from ground water and dimpling from local earth pressure due to surface wheel loads. Sandwich wall construction was investigated as an alternative to use of stiffening ribs and found to be feasible but less cost effective.

Shape-wise, a hollow trapezoid provides efficient bending strength and stiffness, a wide base for proper spacing of cylinder shell support against local buckling, and a narrower top to resist local buckling with high circumferential flange forces. Figure 11.2.17 provides a design concept for a tank with hollow trapezoidal ribbing.

Base width and clear spacing between ribs are established to minimize the number of ribs while providing adequate local buckling resistance for a cylinder wall of approximately the minimum practical shell thickness. Clear spacing must also be sufficient to permit installation of sleeves or nozzles for fill pipes and vents between ribs.

The RPs selected for detailed consideration are designed to provide both structural and liquid-sealing qualities. For example, a smooth liquid tight inner surface is obtained with a resin rich surface layer reinforced with glass filament surfacing veil. It is backed up by a 3.2 mm (1/8 in.) thick liquid seal layer of chopped glass/

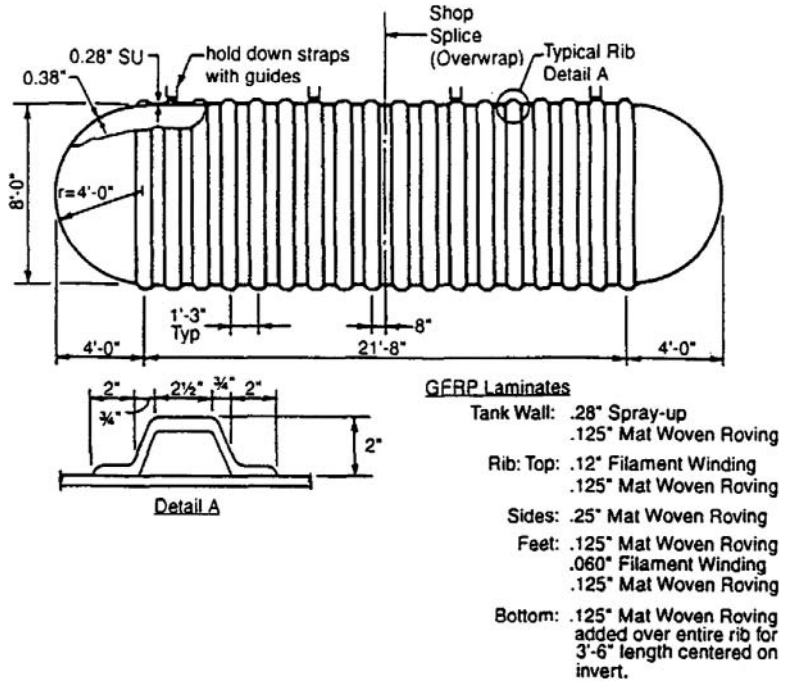


Fig. 11.2.17 Example of a design for 10,000 gallon gasoline RP storage tank.

polyester spray-up. Discontinuous fibers are provided to avoid a continuous path for liquid migration into and along the reinforcement. Additional thickness for structural purposes is provided, either by adding more chopped fiber reinforced spray-up, or by filament winding. An outer resin rich surface layer is reinforced with a surfacing veil. A silica sand filler may be used of bulk, improved compressive stiffness and economy.

Minimum practical total RP thickness is established as 4.8 mm (3/16 in.) for the combined spray-up liquid seal and FW structural layers and 6.4 mm (l/4 in.) for an all-chopped fiber spray-up laminate with sand filler. The choice for any construction is made on the basis of comparative design thickness, weight, and fabrication costs. The all-chopped fiber reinforced construction using somewhat greater wall thickness than the composite FW chopped fiber wall is determined to provide the lowest tank cost; filament winding provides lower weight.

11.2.13.4.3 Hopper rail car tank

In the past (1973) a severe shortage of railroad covered hopper cars for the transportation of grain developed. Cargill, Inc. provided a contract to Structural Composites, Inc. for determining feasibility studies on the potential of using RP in the design and fabrication of these cars. Test results showed structural deficient existed. By 1978 an acceptable design resulted fabricating the Glasshopper (registered name). It was used in rail service March 1981. Cargill Inc., Southern Pacific Transportation Co., and ACF Industries, Inc. (Figure 11.2.18). It was larger and lighter in weight than the conventional steel covered hopper car resulting in being able to deliver more commodity per fuel dollar. Other advantages included corrosion resistance, and lower maintenance costs.

The first to be built was Glasshopper I. It successfully passed all of the required American Association of Railroads (AAR) tests including the 454,000 kg (1,000,000 lb) static end compression test and the 568,000 kg (1,250,000 lb) coupler force impact test in the laboratory, and then successfully completed a round trip between St. Louis, MO and Oakland, CA (9700 km (6000 mile)).

From outward appearance, the RP designs were very similar to the standard ACF steel-covered hopper car. The first RP prototype, Glasshopper I that was in grain service, had four compartments. The car had a total capacity of 142 m^3 (5000 ft^3) and an overall length of about 16 m (53 ft). Its basic specifications are shown in Table 11.2.7.

The second prototype car Glasshopper II that was later put into service had three compartments. The tare weight of the second car was 24,600 kg (54,200 lb), which was 4000 kg (8800 lb) lighter than a standard steel car weight of 28,600 kg (63,000 lb).

Construction details for Glasshopper I consist of a filament wound FW RP car body, RP/balsawood core sandwich panel bulkheads and slope sheets, steel side sills and shear plates, steel bolster webs, and RP hatch covers. Standard running gear and safety appliances were utilized, as were standard gravity outlets. Several changes in construction details such as the use of single laminate slope sheets were made in the design of Glasshopper II to reduce weight and manufacturing costs.

Fig. 11.2.18 RP railroad covered hopper car.

Table 11.2.7 Glasshopper / basic specifications	
Length inside	50 ft 3½ in
Length over end sills	51 ft $5\frac{5}{8}$ in
Length over strikers	52 ft 11 in
Length over coupler pulling face	55 ft 6½ in
Length over running boards	53 ft $\frac{7}{8}$ in
Length between truck centres	42 ft 3 in
Extreme width	10 ft 8 in
Height, rail to top of running boards	15 ft $1\frac{27}{32}$ in
Height, rail to bottom of outlet	12 in
Extreme height, rail to top of hatch bumper	15 ft 6 in
Number of discharge outlets	4
Roof hatch opening, continuous	20 in x 44 ft $7\frac{3}{4}$ in
Curve negotiability, uncoupled	150 ft
Cubic capacity	500 ft^3
Tare weight	59,000 lb
Gross rail load	263,000 lb
AAR clearance diagram	Plate 'C'

Table 11.2.8, shows the weight percentages of steel or RP materials. A significant amount (30 wt%) of the RP car structure is fabricated using RP materials. By subtracting the trucks steel weight, the remaining structure is RP. This construction allows the significant weight reduction to be possible. FEA modeling was used throughout the design stages of the program to aid the structural analysis effort. The structural response in both static and dynamic loading conditions was characterized prior to initiation of the car construction.

FW process was used to fabricate both Glasshopper car bodies. It was determined that this process afforded the best mechanical properties for the lowest cost. Fabricating processes exist that can be highly automated which would help toward having RP-covered hopper cars compete economically with conventional steel-covered hopper cars in the marketplace.

Resin matrix material system chosen for fabrication of the car bodies was a proprietary isophthalic polyester resin system developed by Cargill specifically for the Glasshopper project. PPG, Certain Teed, and OCF supplied the reinforcement of E-glass rovings. Both OCF 450 and 675 yield glass were used successfully in conjunction with the Cargill resin during FW operation. To provide adequate mechanical properties in the directions required to withstand the externally applied service loads, the FW apparatus was programmed to provide the multi-axial filament directional orientation capability.

Secondary bonding operations involving the attachment of stiffeners, etc., for the first RP car, used Hysol's epoxy adhesive (EA 919). This same adhesive was used in joints where both bonding and bolting with mechanical fasteners were employed. Lord Corp.'s acrylic adhesive system (TS 3929-70) was used successfully for Glasshopper II. Hat section stiffeners and wide flange beams were fabricated using the hand lay-up and pultrusion processes, respectively. The material used in the construction of the hat stiffeners included 1½ mat, 24 oz woven rovings, 22½ oz unidirectional fabric, and the isophthalic polyester resin. Pultrusions were purchased, finished, and were fabricated using standard pultrusion processes.

In order to demonstrate structural adequacy, Glasshopper I was tested in the ACF test laboratory located in St. Charles, MO. The test program was designed to show that the car meets and exceeds all requirements as specified by the AAR. Both static and dynamic tests were included in the testing. To determine the car's structural response under various applied loading conditions, Glasshopper I was instrumented with a total of 224 strain gauges, located at various areas determined through structural analysis to be of greatest importance and to provide maximum information. Glasshopper II, instrumented with 310 strain gauges, successfully completed the test program in 1983.

A series of six different static tests were successfully passed by Glasshopper I, including end compression, draft, vertical coupler-up, vertical coupler-down, coupler shank, and torsional jacking. The end compression test consisted of "squeezing" the car, while empty, with a hydraulic ram until a coupler force of 1,000,000 lb was measured. The draft test was conducted on the loaded car (105.9 tons) and consisted of pulling on the coupler until a force of 630,000 lb was experimentally observed. The remaining static tests were all conducted on the loaded car and involved using calibrated hydraulic rams to:

1. Jack the car upward with a vertical force of 22,700 kg (50,000 lb) applied at the coupler pulling face.
2. Jack the car downward with a vertical force of 22,700 kg (50,000 lb) applied at the coupler pulling face.
3. Lift the car free of the truck bolster by jacking at the coupler shank, a vertical force of 50,400 kg (111,000 lb) was required.
4. Lift the car free of the truck bolster by jacking at the lifting lug/jacking pad assembly to verify torsional rigidity and stability, a vertical force of 31,780 kg (70,000 lb) was required.

Table 11.2.8 Glasshopper / component weight summary

Component	Material	Weight, lbs	$\frac{\text{Component weight}}{\text{Total weight}} \times 100$
Car body	RP	6800	11.5
Sandwich panels	RP	4410	7.5
Wide flange beams	RP	640	1.1
Stiffeners	RP	2910	4.9
Top sill	RP	460	0.8
Roof/side angles	RP	510	0.9
Adhesive/bonding strip	RP	1070	1.8
Hatches	RP	860	1.5
Outlets	Steel	1800	3.0
End arrangement	Steel	9640	16.3
Side sills	Steel	4420	7.5
Running boards/safety appln.	Steel	1650	2.8
Brake system	Steel	1570	2.7
Misc. hardware	Steel	1060	1.8
Trucks	Steel	21,200	35.9
	Total weight	59,000	100.0

ΣRP component percentages = 30.
ΣSteel component percentages = 70.

An analysis of test results show the experimentally observed strains to be very close to those predicted using FEA techniques and "hand" calculations. This fact made it possible to use these techniques to further optimize the Glasshopper II design.

After successfully passing all required static tests, Glasshopper I was subjected to a series of impact tests. For these tests, the car that was fully loaded was pulled by cable up an inclined ramp and released to impact another fully loaded standing car that had its brakes released. Velocity of the car at impact was controlled by its height on the ramp at the time of release. Car's velocity was incrementally increased until an experimentally measured coupler force of 113,500 kg (250,000 lb) was developed during the impact.

Velocity of 14.9 km/h (9.24 m/h) was required to obtain the AAR specified load. It is noted that this velocity is significantly higher than the velocity required to reach the specified force with conventional steel covered hopper cars, which is about 12.1 km/h (7.5 m/h). Glasshopper I, with modified bulkhead joints, successfully passed the AAR impact test required and was subsequently prepared for the extended road test.

Following completion of laboratory testing, Glasshopper I was tested over a 9700 km (6000 m) route on the Southern Pacific system. Fully loaded car with 9600 kg (211,000 lb) made the trip from St. Charles, MO to Oakland, CA and back to Houston, TX. The car was unloaded at the Cargill export grain terminal in the Houston area and then returned empty to the facility in St. Charles, MO. The car was accompanied on the trip by the fully instrumented ACF test car used for data acquisition that monitored key strain gauges and load cells throughout the trip. All test results and visual observations showed the car performed well, and as predicted. During certain segments of the testing, speeds of 113 km/h (70 m/h) were reached with no dynamic problems (flutter, hunting, etc.) being observed.

It was determined that two major advantages of the RP covered hopper rail car are its tare weight and its corrosion resistance. As a result of its significantly lower weight and large size, the car is capable of carrying more payload per fuel dollar. This fact is extremely important

in today's conditions of escalating and high fuelled prices. Glasshopper is able to carry many highly corrosive commodities (salts, potash, fertilizer, ore, etc.) without the need for expensive linings and with significantly reduced car maintenance costs. Also, the car's service lifetime would be greatly extended in these severe service environments. Other advantages include the potential to eliminate painting requirements, reduced labor costs in manufacturing, lower center of gravity in the unloaded condition, ability to easily adapt to internal pressure designs, and rapid production changeover to alternate capacity cars.

11.2.13.4.4 Highway tank

RP tanks on firm ground have been holding corrosive materials safely since the 1940s. Later the same technology, with some enhancements material-wise and design-wise, was applied to over-the-road tankers. Some tankers fabricated by Comptank Corp., Bothwell, Ontario, Canada have been on the road in the USA and Canada carrying a wide range of corrosive and hazardous liquids. These RP tank trailers are coded 312 for hauling corrosive and hazardous materials; special designed models haul acids or other corrosive chemicals; they unload by pressure, vacuum, or gravity.

The tankers are FW using E-glass rovings with polyester resin (Reichhold Atlac 4010 AC) and surfacing veil. RP moldings are integrated parts of the shell that is usually 15.88 mm (0.625 in.) thick. These parts include external rings/ribs, covers for steel rollover guards, spill dam, catwalk, hose trays, etc.

11.2.13.4.5 Very large tank

Large filament wound 150,000 gal (568 m^3) tank has been fabricated by the Rucker Co. for Aerojet-General (1966). An example is shown schematically in Figure 11.2.19 of the so-called racetrack-fabricating machine. A fabricated 56 m (22 ft) high, 152 m (60 ft) wide, 318 m (125 ft) long that weighed 32 tons, of all RP, large tank is shown in Figure 11.2.20.

Just the mandrel for this FW machine weighed 100 tons all of metal. Total weight of the steel-constructed machine was 200 tons. The tank contained about 251 million km (158 million miles) of glass fiber, used 8 tons of textile creel containing 60 spools of glass fiber moving up to 7.24 km/h (4½ mph), and took three weeks to manufacture the epoxy-glass fiber RP tank in the Todd shipyard in Los Angeles, CA.

11.2.13.4.6 Corrosive resistant tank

Chemical and corrosive resistant property of many plastics make them useful to contain different liquids ranging from water to acids. They are used extensively in water treatment plants and piping to handle drainage, sewage, and water supply. Glass fiber TS polyester RP water filtration tank is shown in Figure 11.2.21. It is a 20 ft diameter, 32 ft high structure made in sections by a low pressure RP fabricating method. This bonded, assembled tank was shipped on a water barge to its destination. Structural shapes such as this tank for use under corrosive conditions often takes advantage of the properties of RPs and other plastics.

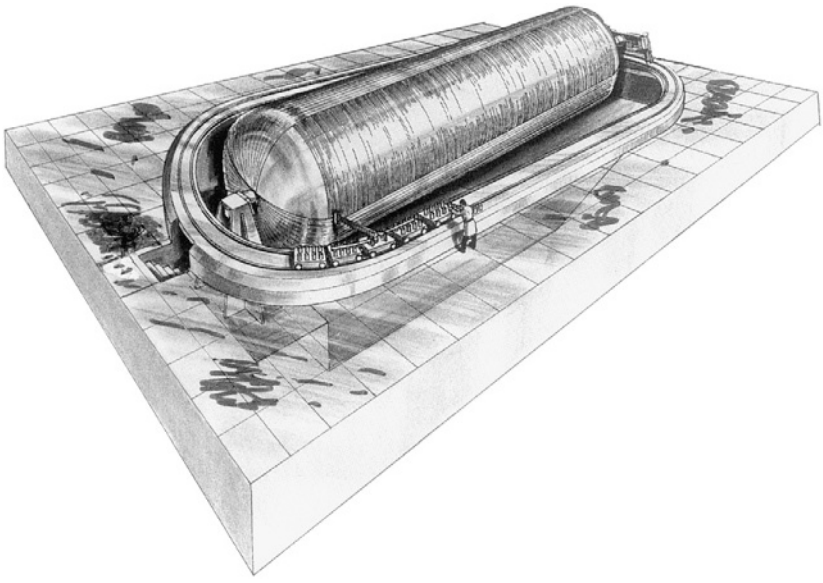


Fig. 11.2.19 Tank's fabricating process.

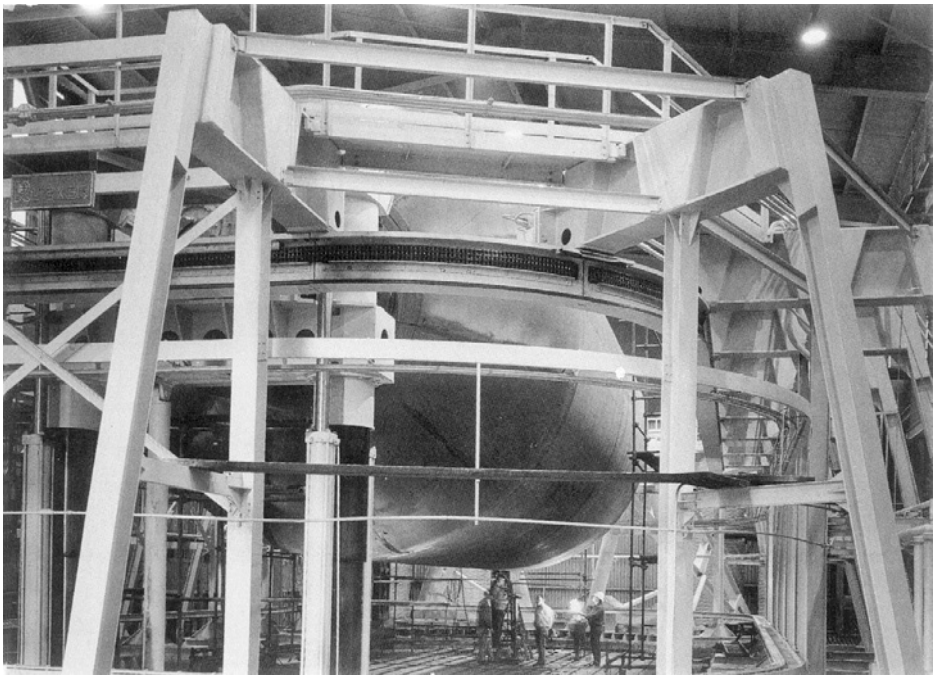


Fig. 11.2.20 RP 150,000 gal tank.

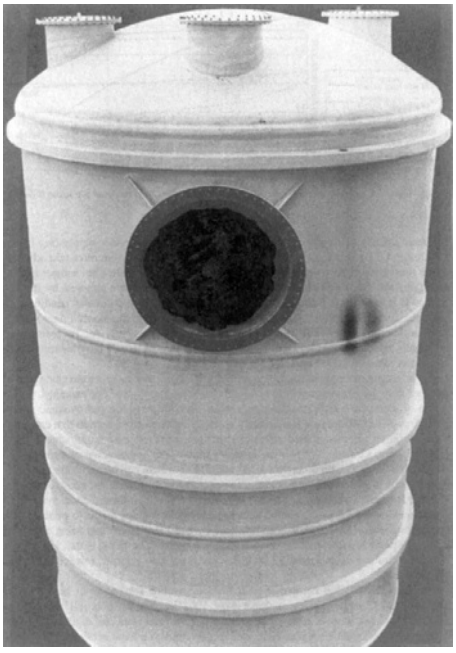


Fig. 11.2.21 Large water filtration tank with 6 ft opening.

11.2.14 Pipe

11.2.14.1 Thermoplastic pipe

Extensive amount of plastic pipe is used worldwide to move different types of liquids, gases, and solids. With the different properties of plastics (such as corrosion resistance, toughness, and strength), pipes can be fabricated to handle practically any type of material. A major and important market for plastics is in producing pipe (tube) for use such as on the ground, underground, in water, and electrical conduits. The largest use is in transporting water, gas, waste matter, industrial mining, etc. Use of TP, such as HDPE, PVC, and PP, provide most of the world using extruders. Plastic pipe represents about 30% of the dollar share compared to other materials (iron/steel at 45%, copper at 12%, concrete at 8%, aluminum at 4%, etc.).

Figure 11.2.22 provides a method to determine pipe thickness subjected to uniform internal pressure P using the standard engineered thin-wall-tube hoop-stress equation. Top view provides an equation that is approximately accurate for $t < d/10$. However, when the wall thickness increases the error becomes large. It is useful in

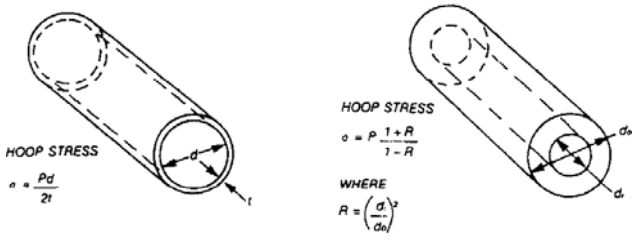


Fig. 11.2.22 Pipe wall thickness determination based on internal pressure.

determining an approximate wall thickness, even when condition ($t < d/10$) is not met. Bottom view provides an equation for the maximum hoop stress that occurs on the surface of the inside wall of the pipe. After the thin-wall stress equation is applied, the thick-wall stress equation can be used to verify the design.

11.2.14.2 RP pipe

An important product even though it represents a small portion of the market is reinforced TS (RTS) plastic; also called reinforced thermoset resin (RTR) according to ASTM standards. Its major material construction is glass fiber with TS polyester plastic that uses fabricating methods ranging from bag molding to filament winding (Chapter 11.1). These RTR pipes provide high load performance both internally and externally.

There are large diameter filament-wound pipes (RTRs) used and accepted in underground burial because they provide conditions such as corrosion resistance and installation-cost savings. Pipe design equations have been used that specifically provide useful information to meet internal and/or external pressure loads. More recently FEA has been used to design RP pipes and other structural products. These design approaches utilize performance standards based upon internal pressures and pipes' stiffness. Other requirements must be met such as longitudinal effects of internal pressure, temperature gradients, and pipe bridging.

When compared to steel pipes there are similarities and dissimilarities. They both differ from concrete (includes asbestos filled type) pipe which is a rigid pipe that cannot tolerate bending or deflection to the same extent as RTR and steel pipe. The following review provides information on the design approach and results of tests conducted on these type pipes (rigid RTR, rigid steel, and flexible concrete). They were buried in trenches under 25 ft of the same dirt and subjected to actual load testing. Specific pressures varied from installation to installation, but the relationship in the way these pipes react to the same burial condition generally remains constant.

As shown in Figure 11.2.23 the load on the surface of the (a) rigid pipe (concrete) is higher at the crown and is transmitted through the pipe directly to the bed of the trench in which the pipe rests and using some side support. The RTR or steel flexible conduit (b) deflects under covering load of earth, this deflection transfers portions of the load to the surrounding envelope of soil that increases the strength of the flexible conduit. Analyzing the type and consolidation of backfill materials is to be considered an integral part of the design process. Because less of a load on the trench bed occurs, the trench requires less bedding bearing strength reducing the installed cost.

Steel pipe is considered a homogeneous isotropic material (equally strong in both hoop and longitudinal directions) where RTR is an anisotropic material (different strength in both the hoop and longitudinal directions). These directional behavior results in the modulus of elasticity (E) to be equal in all directions for steel and not equal for RTR.

The RTR pipe structure is shown in Figure 11.2.24 where the glass fibers are filament wounded at a helix angle that is at 55–65° to the horizontal to maximize hoop and longitudinal stress efficiency. Glass fiber content is at a minimum of 45 wt%. An internal corrosion resistant leak proof barrier liner is usually included that is not included in the stress analysis.

Required is to design a pipe wall structure of sufficient stiffness and strength to meet the combined loads that the pipe will experience in long time service. One design is a straight wall pipe in which the wall thickness controls the stiffness of the pipe. Another way is to

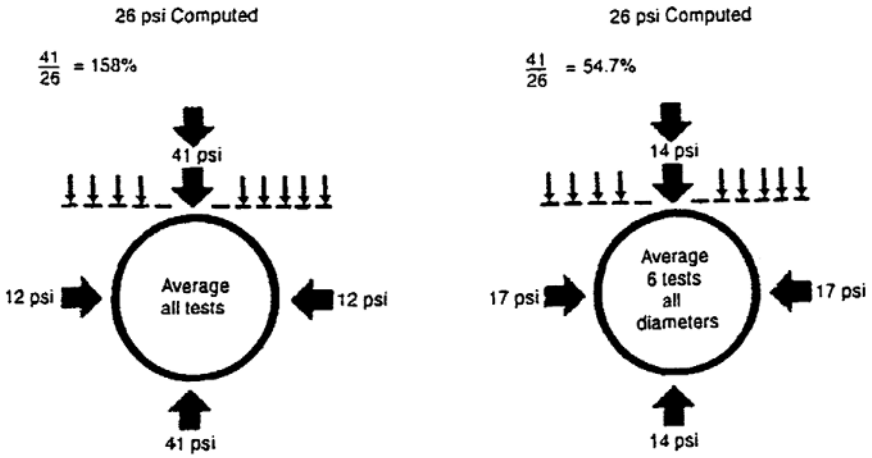


Fig. 11.2.23 Loading (26 psi) rigid concrete pipe and flexible plastic pipe.

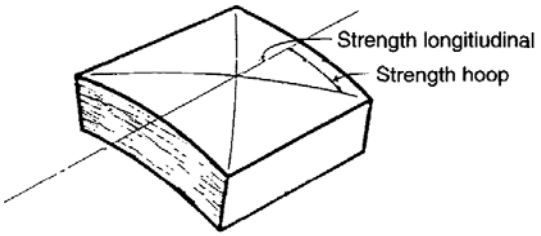


Fig. 11.2.24 Cross section of filament wound layup of the RTR pipe.

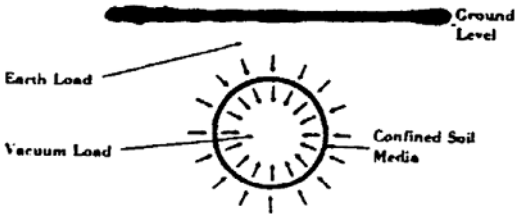


Fig. 11.2.26 Buckling analysis based on conditions such as dead loads, effects of possible flooding, and the vacuum load it is expected to carry.

design a rib-wall pipe on which reinforcement ribs of a specific shape and dimension are wound around the circumference of the pipe at precisely designed intervals. As previously reviewed (RIB section) the advantage of a rib wall pipe is that the wall thickness of the pipe can be reduced (also reducing costs) while maintaining or even increasing its overall strength-to-weight ratio. Where burial conditions are extreme or difficult underwater installations exist a rib-wall pipe design should be considered.

Maximum allowable pipe deflection should be no more than 5% using Figure 11.2.25 equation where ΔX, deflection. This value is the standard of the pipe industry for steel conduit and pipe (AWWA M-11, ASTM, and ASME). Deflection relates to pipe stiffness (EI), pipe radius, external loads that will be imposed on the pipe, both the dead load of the dirt overburden as well as the live loads such as wheel and rail traffic, modulus of soil reaction, differential soil stress, bedding shape, and type of backfill.

To meet the designed deflection of no more than 5% the pipe wall structure could be either a straight wall pipe with a thickness of about 1.3 cm (0.50 in.) or a rib wall pipe that provides the same stiffness. It has to be determined if the wall structure selected is of sufficient stiffness to resist the buckling pressures of burial or superimposed longitudinal loads. The ASME Standard of a four-to-one safety factor on critical buckling is used based on many years of field experience. To calculate the stiffness or wall thickness capable of meeting that design criterion one must know what anticipated external loads will occur (Figure 11.2.26).

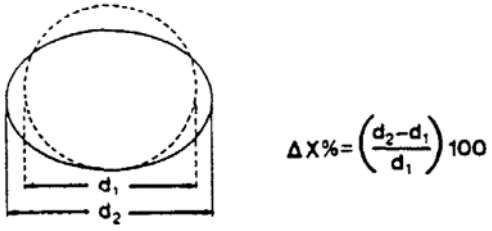


Fig. 11.2.25 Pipe deflection equation.

As reviewed the strength of KTR pipe in its longitudinal and hoop directions are not equal. Before a final wall structure can be selected, it is necessary to conduct a combined strain analysis in both the longitudinal and hoop directions of the RTR pipe. This analysis will consider longitudinal direction and the hoop direction, material's allowable strain, thermal contraction strains, internal pressure, and pipe's ability to bridge soft spots in the trench's bedding. These values are determinable through standard ASTM tests such as hydrostatic testing, parallel plate loading, coupon test, and accelerated aging tests.

Stress–strain (S–S) analysis of the materials provides important information. The tensile S–S curve for steel-pipe material identifies its yield point that is used as the basis in their design. Beyond this static loaded yield point the steel will enter into the range of plastic deformation that would lead to a total collapse of the pipe. The allowable design strain used is about two-thirds of the yield point.

RTR pipe designers also use an S–S curve but instead of a yield point, they use the point of first crack (empirical weep point). Either the ASTM hydrostatic or coupon test determines it. The weep point is the point at which the RTR matrix (plastic) becomes excessively strained so that minute fractures begin to appear in the structural wall. At this point it is probable that in time even a more elastic liner on the inner wall will be damaged and allow water or other liquid to weep through the wall. Even with this situation, as is the case with the yield point of steel pipe, reaching the weep point is not catastrophic. It will continue to withstand additional load before it reaches the point of ultimate strain and failure. By using a more substantial, stronger liner the weep point will be extended on the S–S curve.

The filament-wound pipe weep point is less than 0.009 in./in. The design is at a strain of 0.0018 in./in. providing a 5 to 1 safety factor. For transient design conditions a strain of 0.0030 in./in. is used providing a 3 to 1 safety factor.

Stress or strain analysis in the longitudinal and hoop directions is conducted with strain usually used, since it is easily and accurately measured using strain gauges,

whereas stresses have to be calculated. From a practical standpoint both the longitudinal and the hoop analysis determine the minimum structural wall thickness of the pipe. However, since the longitudinal strength of RTR pipe is less than it is in the hoop direction, the longitudinal analysis is first conducted that considers the effects of internal pressure, expected temperature gradients, and ability of the pipe to bridge voids in the bedding. Analyzing these factors requires that several equations be superimposed, one on another. All these longitudinal design conditions can be solved simultaneously, the usual approach is to examine each individually.

Poisson's ratio can have an influence since a longitudinal load could exist. The Poisson's effect must be considered when designing long or short length of pipe. This effect occurs when an open-ended cylinder is subjected to internal pressure. As the diameter of the cylinder expands, it also shortens longitudinally. Since in a buried pipe movement is resisted by the surrounding soil, a tensile load is produced within the pipe. The internal longitudinal pressure load in the pipe is independent of the length of the pipe.

Several equations can be used to calculate the result of Poisson's effect on the pipe in the longitudinal direction in terms of stress or strain. Equation provides a solution for a straight run of pipe in terms of strain. However, where there is a change in direction by pipe bends and thrust blocks are eliminated through the use of harness welded joints, a different analysis is necessary. Longitudinal load imposed on either side of an elbow is high. This increased load is the result of internal pressure, temperature gradient, and/or change in momentum of the fluid. Because of this increased load, the pipe joint and elbow thickness may have to be increased to avoid overstraining.

The extent of the tensile forces imposed on the pipe because of cooling is to be determined. Temperature gradient produces the longitudinal tensile load. With an open-ended cylinder cooling, it attempts to shorten longitudinally. The resistance of the surrounding soil then imposes a tensile load. Any temperature change in the surrounding soil or medium that the pipe may be carrying can also produce a tensile load. Engineering-wise the effects of temperature gradient on a pipe can be determined in terms of strain.

Longitudinal analysis includes examining bridging if it occurs where the bedding grade's elevation or the trench bed's bearing strength varies, when a pipe projects from a headwall, or in all subaqueous installations. Design of the pipe includes making it strong enough to support the weight of its contents, itself, and its overburden while spanning a void of two pipe diameters.

When a pipe provides a support the normal practice is to solve all equations simultaneously, then determine the minimum wall thickness that has strains equal to or less than the allowable design strain. The result is obtaining the minimum structural wall thickness. This approach provides the designer with a minimum wall thickness on which to base the ultimate choice of pipe configuration. As an example, there is the situation of the combined longitudinal analysis requiring a minimum of $\frac{5}{8}$ in. (1.59 cm) wall thickness when the deflection analysis requires a ½ in. (1.27 cm) wall, and the buckling analysis requires a $\frac{3}{4}$ in. (1.9 cm) wall. As reviewed the thickness was the $\frac{3}{4}$ in. wall. However with the longitudinal analysis a $\frac{5}{8}$ in. wall is enough to handle the longitudinal strains likely to be encountered.

In deciding which wall thickness, or what pipe configuration (straight wall or ribbed wall) is to be used, economic considerations are involved. The designer would most likely choose the $\frac{3}{4}$ in. straight wall pipe if the design analysis was complete, but it is not since there still remains strain analysis in the hoop direction. Required is to determine if the combined loads of internal pressure and diametrical bending deflection will exceed the allowable design strain.

There was a tendency in the past to overlook designing of joints. The performance of the whole piping system is directly related to the performance of the joints rather than just as an internal pressure-seal pipe. Examples of joints are bell-and-spigot joints with an elastomeric seal or weld overlay joints designed with the required stiffness and longitudinal strength. The bell type permits rapid assembly of a piping system offering an installation cost advantage. It should be able to rotate at least two degrees without a loss of flexibility. The weld type is used to eliminate the need for costly thrust blocks.

11.2.15 Spring

There is a difference when comparing the plastic to metal spring shape designs. With metals shape options are the usual torsion bar, helical coil, and flat-shaped leaf spring. The TPs and TSs can be fabricated into a variety of shapes to meet different product requirements. An example is TP spring actions with a dual action shape (Figure 11.2.27) that is injection molded. This stapler illustrates a spring design with the body and curved spring section molded in a single part. When the stapler is depressed, the outer curved shape is in tension and the

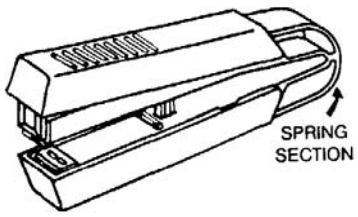


Fig. 11.2.27 TP Delrin acetal plastic molded stapler. (Courtesy of Du Pont.)

ribbed center section is put into compression. When the pressure is released, the tension and compression forces are in turn released and the stapler returns to its original position.

Other TPs are used to fabricate springs. Acetal plastic has been used as a direct replacement for conventional metal springs as well as for providing the capability to use different spring designs such as in zigzag springs, un-coil springs, cord locks with molded-in springs, snap fits, etc. A special application is where TP replaced a metal pump in a PVC plastic bag containing blood. The plastic spring hand-operating pump (as well as other plastic spring designs) did not contaminate the blood.

RP leaf springs have the potential in the replacements for steel springs. These unidirectional fiber RP have been used in trucks and automotive suspension applications. Their use in aircraft landing systems dates back to the early 1940s taking advantage of weight savings and performances. Because of the material's high specific strain energy storage capability as compared to steel, a direct replacement of multileaf steel springs by monoleaf composite springs can be justified on a weight-saving basis.

The design advantages of these springs is to fabricate spring leaves having continuously variable widths and thicknesses along their length. These leaf springs serve multiple functions, thereby providing a consolidation of parts and simplification of suspension systems. One distinction between steel and plastic is that complete knowledge of shear stresses is not important in a steel part undergoing flexure, whereas with RP design shear stresses, rather than normal stress components, usually control the design.

Design of spring has been documented in various SAE and MTM-STP design manuals. They provide the equations for evaluating design parameters that are derived from geometric and material considerations. However, none of this currently available literature is directly relevant to the problem of design and design evaluation regarding RP structures. The design of any RP product is unique because the stress conditions within a given structure depend on its manufacturing methods, not just its shape. Programs have therefore been developed on the basis of the strain balance within the spring to enable suitable design criteria to be met. Stress levels were then calculated, after which the design and manufacture of RP springs became feasible.

11.2.15.1 Leaf spring

RP/composite leaf springs constructed of unidirectional glass fibers in a matrix, such as epoxy resin, have been recognized as a viable replacement for steel springs in truck and automotive suspension applications. Because of the material's high specific strain energy storage capability compared with steel, direct replacement of multi-leaf steel springs by mono-leaf composite springs is justifiable on a weight saving basis. Other advantages of RP springs accrue from the ability to design and fabricate a spring leaf having continuously variable width and/or thickness along its length. Such design features can lead to new suspension arrangements in which the composite leaf spring will serve multiple functions thereby providing part consolidation and simplification of the suspension system.

The spring configuration and material of construction should be selected so as to maximize the strain energy storage capacity per unit mass without exceeding stress levels consistent with reliable, long life operation. Elastic strain energy must be computed relative to a particular stress state. For simplicity, two materials are compared, steel and unidirectional glass fibers in an epoxy matrix having a volume fraction of 0.5 for the stress state of uniaxial tension. If a long bar of either material is loaded axially the strain energy stored per unit volume of material is given by

$$U = (\delta_A^2/2E)(\text{in-lb/in}^3) \tag{11.2.36}$$

where δ_A is the allowable tensile stress and E is Young's modulus for the material.

In Table 11.2.9 the appropriate E for each material has been used and a conservative value selected for δ_A. On a volume basis the RP is about twice as efficient as steel in storing energy; on a weight basis it is about eight times as efficient.

The RP has an advantage because it is an anisotropic material that is correctly designed for this application whereas steel is isotropic. Under a different loading condition (such as torsion) the results would be reversed unless the RP were redesigned for that condition. The above results are applicable to the leaf spring being reviewed because the principal stress component in the spring will be a normal stress along the length of the spring that is the natural direction for fiber orientation.

In addition to the influence of material type on elastic energy storage, it is also important to consider spring configuration. The most efficient configuration (although not very practical as a spring) is the uniform bar in uniaxial tension because the stresses are completely

Table 11.2.9 Glass fiber-epoxy RP leaf spring design

Material	σ_A (ksi)	U (lb/in^2)	U/w[a] (in)
Steel	90	135	470
Glass/epoxy	60	325	4880

[a] w, specific weight.

homogeneous. If the elastic energy storage efficiency is defined as the energy stored per unit volume, then the tensile bar has an efficiency of 100%. On that basis a helical spring made of uniform round wire would have an efficiency of 32% (the highest of any practical spring configuration) while a leaf spring of uniform rectangular cross section would be only 11% efficient.

The low efficiency of this latter configuration is due to stress gradients through the thickness (zero at the mid-surface and maximum at the upper and lower surfaces) as well as along the length (maximum at mid-span and zero at the tips). Recognition of this latter contribution to inefficiency led to development of so-called constant strength beams which for a cantilever of constant thickness dictates a geometry of triangular plan-form. Such a spring would have an energy storage efficiency of 33%. A practical embodiment of this principle is the multi-leaf spring of constant thickness, but decreasing length plates, which for a typical five leaf configuration would have an efficiency of about 22%.

More sophisticated steel springs involving variable leaf thickness bring improvements of energy storage efficiency, but are expensive since the leaves must be forged rather than cut from constant thickness plate. However, a spring leaf molded of the RP can have both thickness and width variations along its length. For instance, a practical RP spring configuration having a constant cross-sectional area and appropriately changing thickness and width will have an energy storage efficiency of 22%. This approaches the efficiency of a tapered multi-leaf configuration and is accomplished with a material whose inherent energy storage efficiency is eight times better than steel.

In this design, the dimensions of the spring are chosen in such a way that the maximum bending stresses (due to vertical loads) are uniform along the central portion of the spring. This method of selection of the spring dimensions allows the unidirectional long fiber RP material to be used most effectively. Consequently, the amount of material needed for the construction of the spring is reduced and the maximum bending stresses are evenly distributed along the length of the spring. Thus, the maximum design stress in the spring can be reduced without paying a penalty for an increase in the weight of the spring. Two design equations are given in the following using the concepts described above.

To develop design formulas for RP springs, we model a spring as a circular arc or as a parabolic arc carrying a concentrated load $2F_v$ at mid-length (Figure 11.2.28).

The governing equation for bending of the spring

$$\frac{1}{\rho} - \frac{1}{R} = \frac{M}{EI} \tag{11.2.37}$$

where R is the radius of curvature of unloaded spring; ρ the radius of curvature of deformed spring; M, the bending moment, E Young's modulus; and I the moment of inertia of spring cross section.

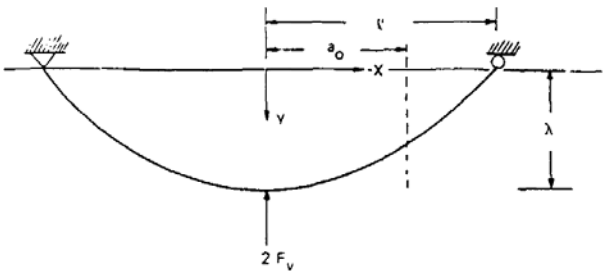


Fig. 11.2.28 RP spring model.

Using the coordinate system shown in Figure 11.2.28, Equation (11.2.37) is rewritten as

$$\frac{d^2y}{dx^2} = \frac{F_v(l-x)}{EI} - \frac{1}{R}, \quad 0 < x < l \tag{11.2.38}$$

where the coordinate y is used to denote the deformed configuration of the spring. Once the maximum allowable design stress in the spring is chosen, Equation (11.2.38) will be used to determine the load carrying capability of the spring. Due to the symmetry of the spring at $x = 0$, only half of the spring needs to be analyzed. It should be noted that Equation (11.2.38) is only an approximate representation of the deformation of the spring. However, for small values of λ/l, it is expected to give reasonably good prediction of the spring rate. Here λ is the arc height and $2i/l$ is the chord length of the spring. Although a nonlinear relation can be used in place of Equation (11.2.38), it would be difficult to derive simple equations for design purposes.

For this particular design, the thicknesses of the spring decrease front the center to the two ends of the spring. Hence, the cross-sectional area of the spring varies along its length. The maximum bending stresses at every cross section of the spring from $x = 0$ to $x = \alpha_0$ are assumed to be identical (Figure 11.2.28). The value of α_0 is a design parameter that is used to control the thickness and the load carrying capability of the spring. If α is the maximum allowable design stress, then the thickness of the spring for $0 \le x \le \alpha_0$ is determined by equating the maximum bending stress in the spring to α_0, thus:

$$h = \left[\frac{6(1-\nu^2)F_v(l-x)}{b\sigma_0}\right]^{1/2}, \quad x \le \alpha_0 \tag{11.2.39}$$

where ν is Poisson's ratio, and b and h are the width and thickness of the spring, respectively.

The factor $(1 - \nu^2)$ is introduced to account for the fact that b could be several times larger than h. If b and h are of the same order of magnitude, a zero value of ν is suggested to be used with Equation (11.2.39). This

equation shows that the thickness of the spring should be a function of F_v σ_ω, l, and b. Once F_v, σ_ω, and l are fixed, then the value of h is inversely proportional to the square root of the width of the spring.

For $x > \sigma_\omega$ the thickness of the spring is assumed to remain constant. The minimum value of h is governed by the ability of the unidirectional composite to carry shear stresses. Using Equation (11.2.38) and the appropriate boundary and continuity conditions, the following equation for the determination of the spring rate is obtained,

$$k_b = \frac{16R^2\sigma_0^3}{27(1-\nu^2)\lambda l^4 E^2}[2l^{3/2} - (l-\alpha_0)^{3/2}]^2 \tag{11.2.40}$$

where k_b is the spring rate per unit width of the spring in lb per in. of vertical deflection. In deriving Equation (11.2.40), the maximum bending stress σ_0 is assumed to develop when $y = 0$ at which the center of the spring rate has undergone a deflection equal to λ. If the actual design value of $2F_v$ is less than or greater than bk_b, the appropriate value of σ_0 to be used in Equation (11.2.74) can be determined easily from the maximum design stress by treating σ_0 as a linear function of $2F_v$.

A constraint on the current fabricating method of the RP leaf spring is that the cross-sectional area of the spring has to remain constant along the length of the spring. This imposes a restriction on the use of variable cross-sectional area design since additional work is required to trim a constant cross-sectional area spring to fit a variable cross-sectional area design. Unless the design stresses in the spring are excessively high, it is preferable to use the less labor-intensive constant cross-sectional area spring. This section describes the design formulas for this type of spring design.

Using the same coordinate system and symbols as shown in Figure 11.2.28, Equations (11.2.37) and (11.2.38) remain valid for the constant cross-sectional area spring. The mid-section thickness of the spring h_0 is related to the maximum bending stress σ_0 by

$$h_0^2 = \frac{6(1-\nu^2)F_v l}{b_0\sigma_0} \tag{11.2.41}$$

where b_0 is the corresponding mid-section width of the spring. Imposing the constant cross-sectional area constraint,

$$b_0 h_0 = bh \tag{11.2.42}$$

the thickness of the spring at any other section is given by

$$h = \frac{6(1-\nu^2)F_v(l-x)}{b_0 h_0 \sigma_0} \tag{11.2.43}$$

The corresponding width of the spring is then obtained from Equation (11.2.42). Based on Equations (11.2.42) and (11.2.43) that the width of the spring will continue to increase as it moves away from the mid-section. In general, there is a limitation on the maximum allowable spring width. Using b_a, to denote the maximum width, the value of x beyond which tapering of the spring is not allowed can be determined by imposing the constant cross-sectional area constraint. One can use α_0 to denote this value of x, then

$$\alpha_0 = l\left[1 - \frac{b_0}{b_a}\right] \tag{11.2.44}$$

Thus, Equation (11.2.43) holds only for $x \le \alpha_0$. Beyond $x = \alpha_0$, the thickness of the spring remains constant and is given by

$$h = \frac{6(1-\nu_2)F_v(l\alpha_0)}{b_0 h_0 \sigma_0}, \quad x \ge \alpha_0 \tag{11.2.45}$$

An implication of Equations (11.2.43) and (11.2.44) is that the maximum bending stresses will remain constant along the length of the spring for $[x] \ge \alpha_0$. Equation (11.2.38) with the appropriate boundary and continuity conditions, the spring rate, k, can be shown to be

$$k = \frac{3El_0}{R\lambda l(1-\nu_2)[1 + 2(1 - b_0/b_a)]} \tag{11.2.46}$$

where I_0 is the moment of inertia of the cross section of the spring at mid-section. In the design of a spring, the values of b_a, l, R, λ, and k are usually given and it is required to determine the values of h_0 for a desirable value of σ_0,. The following equation has been obtained for the determination of h_0:

$$h_0^3 - \frac{4R\sigma_0 h_0^2}{E} + \frac{8Rk\lambda l(1-\nu^2)}{b_a E} = 0 \tag{11.2.47}$$

Once h_0 is determined, the corresponding value of b_0 is then obtained from Equation (11.2.37). In Equation (11.2.47), the value of σ_0 corresponds to a center deflection equal to λ. If the actual design value of $2F_v$, is less than or greater than $k\lambda$, the appropriate value of σ_0 to be used in Equation (11.2.47) can be determined easily from the maximum design stress by treating σ_0 as a linear function of $2F_v$.

Consider, as an example, the design of a pair of longitudinal rear leaf springs for a light truck suspension. The geometry of the middle surface of the springs is given as

$l = 23$ in.
$\lambda = 6$ in.
$R = 44.08$ in.
$b_a = 4.5$ in.

The design load per spring is 2200 lb and a spring rate of 367 lb/in. is required. If σ_0 is set equal to 53 ksi in Equation (11.2.47), two possible design values of h_0 are obtained. Using Equation (11.2.41), the corresponding values of b_0 are determined. Thus, there are two possible constant cross-sectional area designs for this particular spring: (S) $h_0 = 1.074$ in., $b_0 = 2.484$ in. and (SS) $h_0 = 1.190$ in., $b_0 = 2.023$ in. A value of Young's modulus of 5.5×10^6 psi is used in the design of these springs. This corresponds to the modulus of a unidirectional RP with 50 vol% of glass fibers. If a value of σ_0 less than 53 ksi is used in the design, negative and complex values of h_0 are obtained from Equation (11.2.47).

This indicates that it is impossible to design a constant cross-sectional area spring to fit the given design parameters with a maximum bending stress of less than 53 ksi. If a constant width design is required, it can be shown from Equation (11.2.40) that a spring with a constant width of 2.484 in. and a maximum thickness of 1.074 in. will satisfy all the design specifications. The corresponding value of a_0 is 18 in. If a constant width of greater than 2.484 in. is allowed, then a maximum design stress of less than 53 ksi can be obtained.

The above example shows that two plausible constant cross-sectional area designs are obtained to satisfy all the design requirements. If the spring were subjected only to vertical loading, the second design would be selected since it involves less material. However, if the spring is expected to experience other loadings in addition to the vertical load, then it is necessary to investigate the response of the spring to these loadings before a decision can be made.

The effects of these loadings can be determined easily using Castigliano's theorem, together with numerical integration. For illustration, a comparison summation of the responses of the two constant cross-sectional area spring designs are reviewed:

1. *Rotation due to axle torque, M_T*: The rate of rotation of the center portion of the spring due to the axle torque, M_T, is: design (S) $= 1.901 \times 10^5$ in-lb/radians and design (SS) $= 1.895 \times 10^5$ in-lb/radians.

   If an axle torque of 15,000 in-lb is used for M_T, the rotation and the maximum bending stresses for these two springs are in table form:

	Rotation, degree	Maximum stress
Design (S)	4.5	15.7
Design (SS)	4.5	15.7

   The responses of these two designs to the axle torque are, for all practical purposes, identical. As in the case of transverse loading, the maximum bending stresses are uniform along the springs for $[x] \leq \alpha_0$.

2. *Effect of longitudinal force, F_L*: The longitudinal force F_L, will produce a longitudinal and vertical displacement of the spring. Using k_L and k_v, to denote the corresponding spring rate associated with F_L, results in:

	k_L	k_v
Design (S)	2663 lb/in.	1033 lb/in.
Design (SS)	2516 lb/in.	1008 lb/in.

   Assuming that a maximum value of F_L equal to the design load is expected to be carried by the spring, the deflection and the maximum stress experienced by the spring are:

	Longitudinal disp., in.	Vertical disp., in.	Maximum stress, ksi
Design (S)	0.83	2.13	13.8
Design (SS)	0.87	2.18	13.8

   The responses of the two designs to the longitudinal force are essentially identical, The maximum bending stresses are uniform along the springs for $[x] \leq \alpha_0$.

3. *Effect of twisting torque, M_L*: In the usual suspension applications, leaf springs may be subjected to twisting, for example, by an obstacle under one wheel of an axle. For the two springs studied here, the rate of twist is: design (S) $= 1.47 \times 10$ in-lb/radians and design (SS) $= 1.23 \times 10^4$ in-lb/radians.

   In addition, due to the geometry of the spring, the twisting torque M_L will cause the spring to deflect in the transverse direction. The rate of transverse deflection is: design (S) $= 3319$ in-lb/in. and design (SS) $=$ 2676 in-lb/in.

   If a maximum total angle of twist of 10 degrees is allowed, the response of the spring will be:

	Twisting torque (in-lb)	Lateral deflection (in.)	Maximum shear stress (ksi)
Design (S)	2559	0.77	2.55
Design (SS)	2150	0.80	2.58

   In calculating the effect of the twisting torque, the transverse shear modulus of the unidirectional RP has been used. For an RP with 50 vol% of glass fibers, the modulus has a value of 4.6×10^5 psi. The maximum shearing stress occurs at $[x] = \alpha_0$. For designs (S) and (SS), the values of α_0, are 10.3 in. and 12.66 in., respectively. The values of the bending stresses associated with the twisting torque are negligibly small.

4. *Effect of tranverse force*, F_T: As in the case of the twisting torque, the transverse force, F_T, will cause the spring to twist as well as to deflect transversely. The spring rates associated with the transverse force are:

	Twist (in-lb/radian)	Deflection (lb/in)
Design (S)	3319	600
Design (SS)	2676	458

Assuming that a maximum value of F_T is equal to 0.5 times the design load expected to be carried by the spring, the deflection and the maximum stress experienced by the spring will be:

	Angle of twist (degree)	Transverse deflection (in.)	Max. bending stress (ksi)	Max. shear stress (ksi)
Design (S)	19	1.83	11.5	6.05
Design (SS)	23.6	2.40	15.6	7.38

The angle of twist and the maximum shear stresses associated with this lateral force are rather high. In practice, the spring will have to be properly constrained to reduce the angle of twist and the maximum shear stress to lower values. Assuming that a maximum angle of twist of no more than 10 degrees is allowed, the deflection and the maximum stresses experienced by the spring are:

	Angle of twist (degree)	Transverse deflection (in.)	Max. bending stress (ksi)	Max. shear stress (ksi)
Design (S)	10	1.14	11.5	3.93
Design (SS)	10	1.31	15.6	4.09

The maximum bending stresses occur at the center of the spring while the maximum shear stresses occur at the ends of the spring. Based on the above numerical simulations, it appears that both designs respond approximately the same to all different types of loadings. However, design (S) will be preferred since it provides a better response to the lateral and twisting movement of the vehicle.

The maximum bending stress that will be experienced by the spring is obtained by assuming a simultaneous application of the vertical and the longitudinal forces together with the axle torque. A maximum bending stress of 82.5 ksi is obtained. This bending stress is uniform along the spring for $[x] \leq \alpha_0$. In view of the infrequent occurrence of this maximum bending stress, it is expected that the service life of the spring is guaranteed to be long in service. However, a maximum shearing stress close to 6.3 ksi can be reached when the spring is subjected to both twisting torque and transverse force at the same time. The value of this shearing stress may be too high for long-life application. However, a more complete assessment of the suitability of the design can only be obtained through interaction with the vehicle chassis designers.

11.2.15.2 Special spring

As RP leaf springs find more applications, innovations in design and fabrication will follow. As an example, certain processes are limited to producing springs having the same cross-sectional area from end to end. This leads to an efficient utilization of material in the energy storage sense. However, satisfying the requirement that the spring become increasingly thinner toward the tips can present a difficulty in that the spring width at the tip may exceed space limitations in some applications. In that case, it will be necessary to cut the spring to an allowable width after fabrication. There are special processes such as basic filament winding that can fabricate these type structures.

A similar post-molding machining operation is required to produce variable thickness/constant width springs. In both instances end to end continuity of the fibers is lost by trimming the width. This is of particular significance near the upper and lower faces of the spring that are subject to the highest levels of tensile and compressive normal stresses. A practical compromise solution is illustrated in Figure 11.2.29. Here excess material is forced out of a mid-thickness region during molding that maintains continuity of fibers in the highly stressed upper and lower face regions. A further advantage is that a natural cutoff edge is produced. The design of such a feature into the mold must be done carefully so that the molding pressure (desirable for void-free parts) can be maintained.

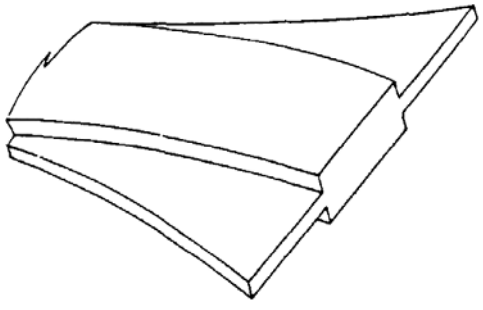


Fig. 11.2.29 Spring with a practical loading solution.

An area of importance is that of attaching the spring to the vehicle. Since the RP spring is a highly anisotropic part especially designed as a flexural element, attachments involving holes or poorly distributed clamping loads may be detrimental. For example, central clamping of the spring with U bolts to an axle saddle will produce local strains transverse to the fibers that in combination with transverse strains due to normal bending may result in local failure in the plastic matrix. The use of a hole for a locating bolt in the highly stressed central clamped region should also be avoided.

Load transfer from the tips of the RP spring to the vehicle is particularly difficult if it is via transverse bushings to a hanger bracket or shackles since the bushing axis is perpendicular to all the reinforced fibers. One favorable design is shown in Figure 11.2.30. This design utilizes a molded random fiber RP (SMC; Chapter 11.1) bracket that is bonded to the spring. Load transfer into this part from the spring occurs gradually along the bonded region and results in shear stresses that are conservative for the adhesive as well as both composite parts.

11.2.15.3 Cantilever spring

The cantilever spring (unreinforced plastics or RPs) can be employed to provide a simple format from a design standpoint. Cantilever springs, which absorb energy by bending, may be treated as a series of beams. Their deflections and stresses are calculated as short-term individual beam-bending stresses under load.

The calculations arrived at for multiple-cantilever springs (two or more beams joined in a zigzag configuration, as in Figure 11.2.31) are similar to, but may not be as accurate as those for a single-beam spring. The top beam is loaded (F) either along its entire length or at a fixed point. This load gives rise to deflection y at its free end and moment M at the fixed end. The second beam load develops a moment M (upward) and load F (the effective portion of load F, as determined by the various angles) at its free end. This moment results in deflection y_2 at the free end and moment M_2 at the fixed end (that is, the free end of the next beam). The third beam is loaded by M_2 (downward) and force F_2 (the effective portion of F_1). This type action continues.

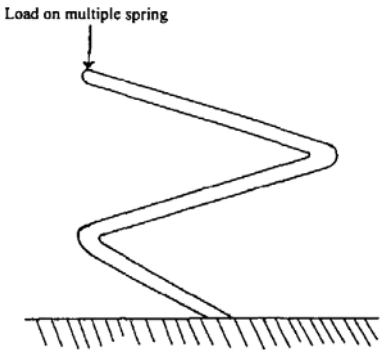


Fig. 11.2.31 Multiple-cantilever zigzag beam spring. (Courtesy of Plastics FALLO.)

Total deflection, y, becomes the sum of the deflections of the individual beams. The bending stress, deflection, and moment at each point can be calculated by using standard engineering equations. To reduce stress concentration, all corners should be fully radiused. The relative lengths, angles, and cross-sectional areas can be varied to give the desired spring rate F/y in the available space. Thus, the total energy stored in a cantilever spring is equal to

$$E_c = \frac{1}{2}F_y \tag{11.2.48}$$

where F is the total load in lb, y the deflection in in., and E the energy absorbed by the cantilever spring, in-lbs.

11.2.15.4 Torsional beam spring

Torsional beam spring design absorbs the load energy by its twisting action through an angle zero. Figure 11.2.32 is an example of its behavior is that of a shaft in torsion so that it is considered to have failed when the strength of the material in shear is exceeded.

For a torsional load the shear strength used in design should be the value obtained from the industry literature

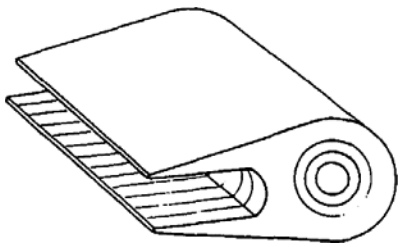


Fig. 11.2.30 Spring has a bonded bracket.

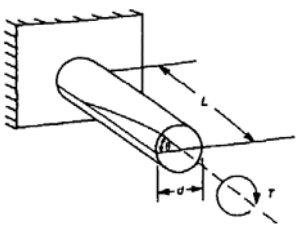


Fig. 11.2.32 Example of a shaft under torque.

(material suppliers, etc.) or one half the ultimate tensile strength, whichever is less. Maximum shear stress on a shaft in torsion is given by the following equation using the designations from Figure 11.2.32:

$$\tau = T_c/J \tag{11.2.49}$$

where T is the applied torque in in-lb, c the distance from the center of the shaft to the location on the outer surface of the shaft where the maximum shear stress occurs, in. and J the polar moment of inertia, in.^{4}.

Using mechanical engineering handbook information the angular rotation of the shaft is caused by torque that is developed by

$$\theta = TL/GJ \tag{11.2.50}$$

where L is the length of shaft, in., G the shear modulus, psi $= E/2(1 + \nu)$, E the tensile modulus of elasticity, psi, and ν Poisson's ratio.

The energy absorbed by a torsional spring deflected through angle θ equals:

$$E_t = \frac{1}{2} M_\theta \times \theta \tag{11.2.51}$$

where M_θ is the torque required for deflection θ at the free end of the spring, in-lb.

11.2.16 Hinge

Since many different plastics are flexible (Chapter 11.1) they are used to manufacture hinges. They can operate in different environments. Based on the plastic used they can meet a variety of load performance requirements. Land length to thickness ratio is usually at least 3 to 1.

Hinges can be fabricated by using different processes such as injection molding, blow molding, compression molding, and cold worked. So called "living hinges" use the TP's molecular orientation to provide the bending action in the plastic hinge. With proper mold design (proper melt flow direction, eliminate weld line, etc.) and process control fabricating procedures these integral hinge moldings operate efficiently.

Otherwise problems in service immediately or shortly after initial use delamination occurs. Immediate post-mold flexing while it is still hot is usually required to ensure its proper operation.

Hinges depend not only on processing technique but also using the proper dimensions based on the type plastic used. Dimensions can differ if the hinge is to move 45–180°. If the web land length is too short for the 180° it will self-destruct due to excessive loads on the plastic's land.

11.2.17 Press fit

Press fits that depend on having a mechanical interface provide a fast, clean, economical assembly. Common usage is to have a plastic hub or boss that accepts either a plastic or metal shaft or pin. Press-fit procedure tends to expand the hub, creating tensile or hoop stress. If the interference is too great, a high strain and stress will develop. Thus it may fail immediately, by developing a crack parallel to the axis of the hub to relieve the stress, which is a typical hoop-stress failure. It could survive the assembly process, but fail prematurely in use for a variety of reasons related to its high induced-stress levels. Or, it might undergo stress relaxation sufficient to reduce the stress to a lower level that can be maintained.

Hoop-stress equations for press-fit situations are used. Allowable design stress or strain will depend upon the particular plastic, the temperature, and other environmental considerations. Hoop stress can be obtained by multiplying the appropriate modulus. For high strains, the secant modulus will give the initial stress; the apparent or creep modulus should be used for longer-term stresses. The maximum strain or stress must be below the value that will produce creep rupture in the material. There could be a weld line in the hub that can significantly affect the creep-rupture strength of most plastics.

Complications could develop during processing with press fits in that a round hub or boss may not be the correct shape. Strict processing controls are used to eliminate these type potential problems. There is a tendency for a round hub to be slightly elliptical in cross-section, increasing the stresses on the part. For critical product performance and in view of what could occur, life-type prototyping testing should be conducted under actual service conditions in critical applications.

The consequences of stress occurring will depend upon many factors, such as temperature during and after assembly of the press fit, modulus of the mating material, type of stress, usage environment and probably the most important is the type of material being used. Some substances will creep or stress relaxes, but others will fracture or craze if the strain is too high. Except for light press fits, this type of assembly design can be risky enough for the novice, because a weld line might already weaken the boss.

Associated with press fit assembly methods are others such as molded-in inserts usually used to develop good holding power between the insert and the molded plastic.

11.2.18 Snap fit

Snap fits are used in all kinds of products ranging from toys to highly loaded mechanical tools. There are both temporary and permanent assemblies, principally in

injection and blow molded products. The following guidelines are recommended regarding the position of the snap joint to the injection molded gate and the choice of the wall thicknesses in the area of flow to the place of joining: (1) there should be no binding seams at critical points; (2) avoid binding seams created by stagnation of the melt during filling; (3) the plastic molecules and the filler should be oriented in the direction of stress; and (4) any uneven distribution of the filler should not occur at high-stress points.

Use of snap fits provides an economical approach where structural and nonstructural members can be molded simultaneously with the finished product and provide rapid assembly when compared with such other joining processes as screws. As in other product design approaches (nothing is perfect), snap fits have limitations such as those described in Table 11.2.10.

Snap fits can be rectangular or of a geometrically more complex cross section. The design approach for the snap fit beam is that either its thickness or width tapers from the root to the hook. Thus, the load-bearing cross section at any location relates more to the local load. Result is that the maximum strain on the plastic can be reduced and less material will be used. With this design approach, the vulnerable cross section is always at the root.

Geometry for snap joints should be chosen in such a manner that excessive increases in stress do not occur. The arrangement of the undercut should be chosen in such a manner that deformations of the molded product from shrinkage, distortion, unilateral heating, and loading do not disturb its functioning.

Snap fits can be applied to any combination of materials, such as plastic and plastic, metal and plastics, glass and plastics, and others. All types of plastics can be used.

Table 11.2.10 Snap fit behaviors

Advantages
Compact, space-saving form.
Takes over other functions like bearing, spring cushioning, fixing.
Higher forces can also be transmitted with proper designing.
Small number of individual parts.
Assembly of a construction system with little expenditure of production facilities and time.
Can be easily integrated into the structural member.
Disadvantages
Influence of environmental effects (such as distortion due to temperature differences) on the functioning.
Effects of processing on the properties of the snap joints (orientation of the molecules and of the filler, distribution of the filler, binding seams, shrinkage, surface, roughness and structure).
The fixing of the joined parts is weaker than in welding, bonding, and screw joining.
The conduct of force at the joining place is lesser than in areal joining (bonding, welding).

Their strength comes from its mechanical interlocking, as well as from friction. Pullout strength in a snap fit can be made hundreds of times larger than its snap in force. In the assembly process, a snap fit undergoes an energy exchange, with a clicking sound. Once assembled, the components in a snap fit are not under load, unlike the press fit, where the component is constantly under the stress resulting from the assembly process. Therefore, stress relaxation and creep over a long period may cause a press fit to fail, but the strength of a snap fit will not decrease with time. They compete with screw joints when used as demountable assemblies. The loss of friction under vibration can loosen bolts and screws whereas a snap fit is vibration proof.

The interference in a snap fit is the total deflection in the two mating members during the assembly process. Note that too much interference will create difficulty in assembly, but too little will cause low pullout strength. A snap fit can also fail from permanent deformation or the breakage of its spring action components. A drastic change in the amount of friction, created by abrasion or oil contamination, may ruin the snap. These conditions influence the successful designing of snap fits that basically depend on observing their shape, dimensions, materials, and interaction of the mating parts.

Most common snaps are the cantilever type, the hollow-cylinder type (as in the lids of pill bottles) and the distortion type. Cantilever category includes any leaf-spring components, and the cylinder type is used also to include noncircular section tubes. These snaps include those in any shape that is deformed or deflected to pass over interference. The shapes of the mating parts in a hollow cylinder snap is the same, but the shapes of the mating parts in a distortion snap are different, by definition.

Snap fits flex like a spring and quickly return, or at least nearly return to its unflexed position. Target is to provide sufficient holding power without exceeding the elastic limits of the plastics. Using the engineering beam equations one can calculate the maximum stress during assembly. If it stays below the yield point of the plastic, the flexing beam will return to its original position. However, for certain designs there will not be enough holding power, because of the low forces or small deflections. It has been found that with many plastics the calculated flexing stress can far exceed the yield point stress if the assembly occurs too rapidly. The flexing finger will just momentarily pass through its condition of maximum deflection or strain, and the material will not respond as if the yield stress had been greatly exceeded.

Another popular approach to evaluate the design of snap fits is to calculate their strain rather than their stress. Then compare this value with the allowable dynamic strain limits for the particular plastics. In designing the beams it is important to avoid having sharp comers or structural discontinuities that can cause stress risers.

Tapered finger provides a more uniform stress distribution, which makes it advisable to use where possible.

11.2.19 Tape

Plastic tapes are used to meet many different requirements that range from being flexible to strong, water to chemical and other environmental resistance, soft to wear resistance, and so on. This review is on the overall performance of tapes. Even though tape is a market in its own, there are other markets such as for belts that have some similar features that they both meet rigid and versatile requirements.

One of many different performing types is a low-profile long conveyor belts with prolonged high-speed operating life and minimal maintenance for use in plants, in underground mines, etc. The conveyors can have different belts for different applications including part accumulation, hot or cold processes, or chemical resistance. Closed-top and open-mesh versions are available. There are accumulation belts that use a blend of acetal and Teflon and meets FDA standards. It can withstand temperatures up to 180 °C. For higher temperatures, a flat or cleated line of nylon belts operates at temperatures as high as 375 °F. Chemical-resistant applications can use a flat, side walled, or cleated belt that resists bleaches and acids while functioning. For electronic applications, the flat or cleated static-conductive belt made from polypropylene meets FDA standards for Class 11 type charges.

This review on tapes highlights the historically Du Pont's research leading to Dymetrol® elastomeric tape that began in 1974. The General Motors Corp. in the USA had developed a new lightweight window regulator, to replace the heavy metal segment window regulators, but cold not make it work adequately with the metal and plastic tapes used at that time. Using its plastics processing know-how, Du Pont developed what is now known as highly engineered oriented elastomeric tape, or as the Dymetrol mechanical drive tape, and General Motors have been using it since 1979 in manual and electric window car regulators. Today this tape with its applications has evolved into a multi-tape/multi-application proposition that include safety passive restrainers, windshield wiper linkages, sliding car roofs, garage door openers, vending machines, etc.

This high modulus material composition provides tape with steep stress–strain characteristics. In other words higher dimensional stability under applied loads or higher tensile loading capability at the same elongation (vs. the standard material composition tapes). High modulus material composition tapes also have higher stiffness, resulting in a much-improved push vs. pull load transfer efficiency. In practice this means for instance that window regulator mechanisms can be constructed with tape lifiing the window as well in the compressive as in the tension mode which provides the automotive design engineer with more possibilities and flexibility to conceive car doors with optimum cost, performance and design characteristics.

Another novelty is the abrasion resistant material composition option that confers much improve abrasion resistance and somewhat lower coefficient of friction. The mechanical drive tape will also transfer tension and compressive forces when used in non-linear directions. Contributing factors are not only the tape's axial stiffness, providing the push and pull, but also its torsional and edge-bend flexibility.

Figure 11.2.33 shows the flexibility of high modulus vs. standard tape materials and also the effect of punching and of temperature on tape stiffness. Naturally tape thickness and width also affect a tape's stiffness. The obvious user benefit of this tape flexibility is greater versatility in the design of energy transfer mechanisms, since allowing for push/pull in non-linear modes, thereby advantageously replacing more complex movement transmission devices such as lever arm systems, pulleys, or gear systems. By using this feature, a cigarette vending machine offers a 50% increase in brand choice without increase in machine dimensions. Metal wire cable may rapidly fail in energy transfer mechanisms due to its frequently inadequate alternate flex cycle life, and spiral metal cable due to its spiral collapse, these mechanical drive tapes have proved to be extremely tough.

From these data, taken from various points of the tape's stress–strain curve, it can be concluded that the strength of the mechanical drive tapes. and in particular of the high modulus materials composition, is appreciable and adequate indeed for the low to medium load transfer service applications to which it addressed itself, and more than what is needed for window regulators for instance. The user benefit here naturally is long-term performance dependability or tape driven energy transfer mechanisms and proved for instance by the low GM car after-sale replacement rates of tape driven window regulators. Another example is the. use of tape to drive outdoor venetian blinds in which the previous drive device failed frequently causing expensive repair.

Contributing significantly to this tape toughness is its property to recover from strain caused by permanent or intermittent operating stress, even after exposure to temperatures as high as 80 °C.

There is virtually no creep and very little permanent deformation. Similar tests have furthermore shown that there is not much more deformation even after 8000 h exposure to 4000 psi. The added benefit of this low creep and strain recovery characteristic of tape is that it confers operating shock absorbency and smoothness to energy transfer mechanisms, not or less provided by other energy transmission devices since featuring steeper stress–strain characteristics or no stress elasticity at all.

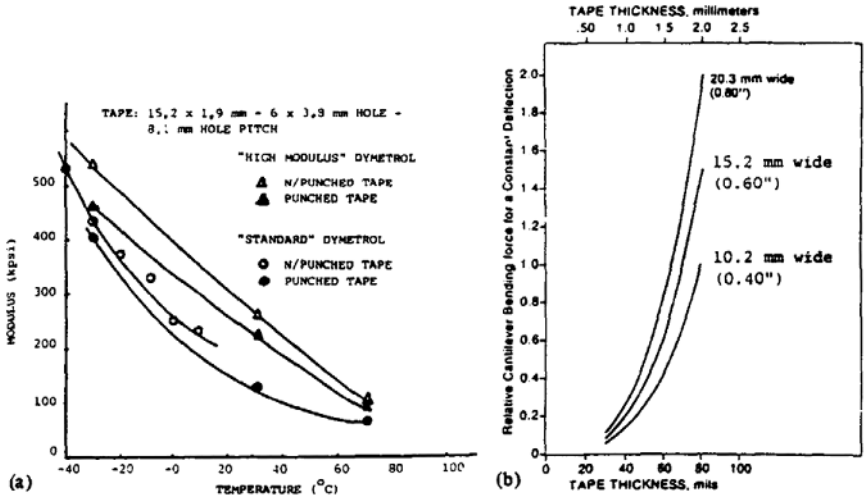


Fig. 11.2.33 Dymetrol mechanical drive tape: (a) flexural modules and (b) beam flexure versus tape thickness (unpunched).

11.2.20 Packaging

Plastics are used to package many different forms and shapes of products. Their performance requirements are very diversified ranging from relatively no strength to extremely high strength, flexible to rigid, non-permeable to permeable (in many different environments), and so on. They require design performance requirements. There is extensive literature on the subject of packaging and all their ramifications.

Packaging industry and its technology is the major outlet for plastics where it consumes about 30 wt% of all plastics (yearly sales above $40 billion) with building and construction in second place consuming about 20 wt%. If plastic packaging were not used, the amount of packaging contents (food, soaps, etc.) discarded from USA households would more than double. Plastics are the most efficient packaging materials due to their higher product-to-package ratio as compared to other materials. One ounce of plastic packaging can hold about 34 ounces of product. A comparison of product delivered per ounce of packaging material shows 34.0 plastics, 21.7 aluminum, 6.9 paper, 5.6 steel, and 1.8 glass.

When packaging problems are tough, plastics often are the answer and sometimes the only answer. They can perform tasks no other materials can and provide consumers with products and services no other materials can provide. As an example plastics have extended the life of vegetables after they are packaged.

If plastic packaging were not used, the amount of packaging contents (food, etc.) discarded from USA households would more than double. Plastics are the most efficient packaging materials due to their higher product-to-package ratio as compared to other materials. One ounce of plastic packaging can hold about 34 ounces of product. A comparison of product delivered per ounce of packaging material shows 34.0 plastics, 21.7 aluminum, 6.9 paper, 5.6 steel, and 1.8 glass.

Different designs and processing techniques are used to produce many different packaging products. These different products show how innovative designs have created different products based on plastic behaviors and they are processing capabilities. Most of these products are extruded film and sheet that are usually thermoformed. Other processes are used with injection molding and blow molding being the other principal types used.

The largest market at 35% of the total is for carded blister packs. The second major product is window packaging at 24%. The others are clamshell packaging at 20%, skin packaging at 18%, and others at 3%. The following information provides examples of packaging products that meet different performance requirements that relates to the capability of the plastic used: *aseptic* in food processing; *bag-in-box* refers to a sealed, sprouted plastic film bag inside a molded rigid container; *beverage can* with aluminum cans dominating the USA market for soft drink containers with about 70% of the market, PET plastic and glass compete for second place-note that most aluminum cans have an inside coating, usually epoxy, to protect its contents from the aluminum; *beverage container* with carbonated soft drinks being the largest market at about 50% followed by beer at about 25%, fruit juices and drinks, and milk; *beer bottle* potential in bioriented stretched plastic bottles in USA is on the horizon using coinjection or coextruded plastics such as PET plastic and/or PEN plastic using various barrier plastics or systems; *biological substance* that are classified as hazardous requiring specialty packaging where plastics play an important role to meet strict requirements; *blister* also called blister carded packaging; *bubble pack* is plastic cushioning material used in packaging; *contour packaging* is also called skin

packaging; *dual-ovenable tray* are used for frozen foods; *electronic packaging* with plastic ease of processing and low cost has given them a wide application in solving problems in electronic packaging; *film breathable* identified as controlled-atmospheric packaging (cap); *food packaging* with plastics provides the most efficient packaging materials due to their higher product-to-package ratio as compared to other materials; *food oxygen scavenger* impregnated plastics with chemically reactive additives that absorb oxygen, ethyl, and other agents of spoilage inside the package once it has been sealed; *grocery bag; hot fill package* injection & blow molded bottles, thermoformed containers, etc.; *modified atmosphere packaging* (mat) is a packaging method that uses special mixtures of gases (carbon dioxide, nitrogen, oxygen, or their combinations; hermetically sealed mat extends the shelf life of red meat, skinless turkey breast, chicken, half-baked bread, pizza's crust, bagels, etc.; *peelable film* for case-ready ground beef package add color and shelf life; *pouch* heat-sealed, wrap, and reusable container help keep food fresh and free of contamination; and *retortable pouch* has superior flavor retention and long shelf life.

11.2.20.1 Permeability

There are different approaches for providing permeability resistance in plastic packaging. The more extensively used is barrier plastics with nonbarrier types to meet cost-to-performance requirements. Most plastics can be considered barrier types to some degree, but as barrier properties are maximized in one area (as gases such as O_2, N_2, or CO_2), other properties as permeability and moisture resistance diminish. This is achieved through coextrusion, coinjection, corotation, and other such processes. The other chemically modifies the plastics' surfaces.

There are plastics that have different behaviors to protect and preserve products in storage and distribution. They provide different diffusion (transport) action of gases, vapors, and other low molecular weight species through the materials. Important selection in packaging (food, etc.) is based on the permeability of the materials to factors such as oxygen and water vapor. There are special applications such as packaging bananas where an ethylene gas remains in a package is used to artificially ripen the bananas. There is other industrial gas separation systems that use the selective permeability of plastics to separate their constituents.

This diffusion rate is related to the resistance, within the plastic wall, to the movement of gases and vapors. Important aspects of the diffusion process are permeability and migration of additives. Possible migrants from plastics can include the many different additives and fillers used (Chapter 11.1).

The conditions inside and outside packages relate to the gas and vapor pressure forces penetrating or diffusing through permeable packages. As the engineering handbooks report diffusing substance's transmission rate is expressed by mathematical equations commonly called Fick's First and Second Laws of Diffusion very popular in the industry:

$$F = D(dC/dX) \tag{11.2.52}$$

$$dC/dt = D(d^2C/dX^2) \tag{11.2.53}$$

where F is the flux (the rate of transfer of a diffusing substance per unit area); D the diffusion coefficient; C the concentration of diffusing substance; t the time, and X the space coordinate measured normal to the section.

Different test methods are used to conduct permeability behaviors of plastics to measure gas, water, and other material vapor permeability. Permeability test procedures used to measure the permeability of plastic films include those identified as the absolute pressure method, the isostatic method, and the quasi-isostatic method. Basic approach mounts a plastic film sample between two cell chambers of a permeability cell (two piece closed container). One chamber holds the gas or vapor to be used as the permeant. The permeant then diffuses through the film into a second chamber, where a detection method such as optical devices, infrared spectroscopy, a manometric, gravimetric, or coulometric method; isotopic counting; or gas–liquid chromatography provides a quantitative measurement.

The absolute pressure method (ASTM D1434, Gas Transmission Rate of Plastic Film and Sheeting) is used when no gas other than the permeant in question is present. In this test the change in pressure on the volume of the low-pressure chamber measures the permeation rate. Between the two chambers a pressure differential provides the driving force for developing permeation.

Isostatic testing equipment has been used for measuring the oxygen and carbon dioxide permeability of both plastic films and complete plastic packages. Pressure in each test chamber are held constant by keeping both chambers at atmospheric pressure. With gas permeability measurement, there must be a difference in permeant partial pressure or a concentration gradient between the two cell chambers. Gas that permeates through the film into the lower-concentration chamber is then conveyed to a gas sensor or detector by a carrier gas for quantification.

A variation of the isostatic method is the quasi-isostatic method where at least one chamber is completely closed and no exposure to atmospheric pressure. However, there is a difference in penetrant partial pressure or a concentration gradient between the two cell chambers. The concentration of permeant gas or vapor

that has permeated through into the lower-concentration chamber can be quantified by a technique such as gas chromatography.

There is the chemically modifying approach. It changes the plastic's surface during or after fabrication permiting the control of the permeation behavior of such products as film, sheets, diaphragms, and containers. These techniques are becoming increasingly important. There is an endless search for better barrier materials for packaging applications. As an example in blow-molded gasoline containers/tanks, the amount of gasoline permeation through HDPE even though it is very low, is still excessive per the standard requirements, thus has required some type of barrier. Including a barrier in a multilayer construction has created a satisfactory solution. The approach to functionalized PE formed tanks on the inside wall is by a chemical reaction, mostly by exposing the surface to sulfonation or fluorination.

Oxifluorination has the fluorine gas combined with nitrogen to which several percent of oxygen by volume have been added. Subjecting PE to fluorine and oxygen at the same time leads to functionalization of the PE making it impermeable. This method permits substantially reducing the required amount of fluorine, resulting in a cost-to-performance improvement. Barrier plastics using oxifluorination are used for foods. They provide barriers that are needed to protect them against spoilage from oxidation, moisture loss or gain, and changes or losses in favor, aroma, or color.

11.2.20.2 Cushioning

Different plastic foams are used in many different products meeting different performance requirements (Chapter 11.1). In the packaging industry they provide different performances that includes cushioning. They are used in different weights to meet different product requirements. Regarding cost and performance it is sometimes believed that the lower density closed-cell foams that are usually priced lower provide superior cushioning performance. This assumption is usually incorrect as shown in Figure 11.2.34. Even though it contains less plastic, the fabricating rates, the amount and cost of the blowing agent, and the amount and cost of the base plastic all influence the final cost. As a result, very low density foams can actually be more costly to make than others. Cost of a foam is usually not proportional to its density.

The same type and thickness closed-cell polyethylene foams with differing densities have been subjected to dynamic static load cushioning conditions during impact. Curves relate mechanical shock experienced during an impact where the lower the curve goes, the greater the cushioning efficiency. Curves for foam densities above 6 pcf (lb/ft^3) the maximum cushioning efficiency of each

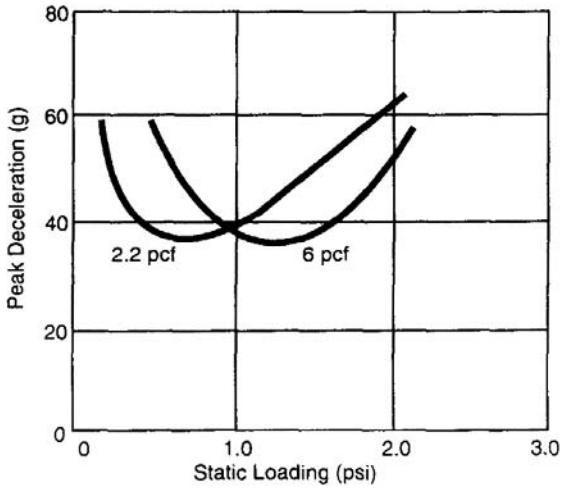


Fig. 11.2.34 Cushioning effect of polyethylene foam density is influenced by loading.

material is not significantly different, however a dramatic change occurs with a change in the applied load.

If a 40 g package were to be designed according to Figure 11.2.34 using a 6.0 pcf foam, the foam would measure 3 in. thick resulting in 40 g shock. If a similar package were then produced using a 2.2 pcf foam, its shock performance would not go as low as 40 g but would instead produce about 60 g's, or 50% more shock or a 50% loss in shock efficiency.

To meet 40 g's, the 2.2 pcf package would need to be redesigned. Greatly increasing the thickness of the pads constructed from the lower density foam can be used to provide adequate protection. However the result would of increasing the package size, impair handling and shipping efficiency, and possibly result in higher costs. The 6.0 pcf foam could, however, be reliably used.

To be at 40 g level and keep the 2.2 pcf foam thickness the same, reduce loading from 1.35 to 0.87 psi can be used. Although this approach keeps the package size the same, nearly twice as much foam must be used to meet the lower loading. Lowering the density produces a considerably higher deceleration and reduces cushioning performance. Figure 11.2.35 compares the effect of density. The lower density foam cost less than half on a cost-per-unit volume basis resulting in a cost savings. Below a density of about 2.2 pcf the cushioning efficiency can begin to change with the density.

Significant in this figure is the narrower range of usable static loadings at the bottoms of the curves that resulted when the density is reduced. Important consideration in comparing foams of different densities is their compressive creep resistance, and their ability to resist undergoing a permanent thickness loss during their time under load. As the density decreases, so does the creep resistance.

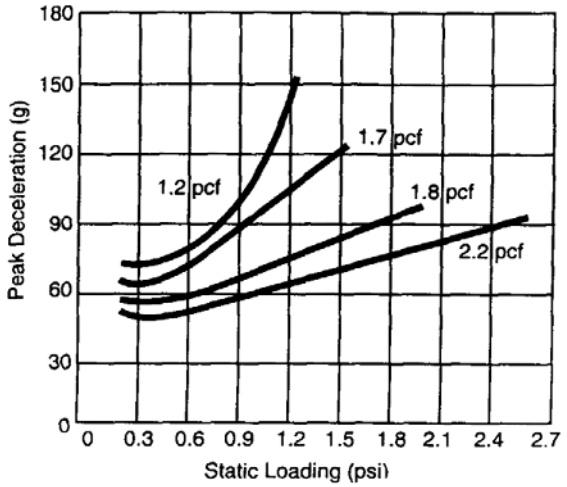


Fig. 11.2.35 Comparison of different foam densities.

11.2.20.3 Building

There are many applications of structural and nonstructural plastics being used in the building and construction industry worldwide. Plastic use continues to expand in present and new products in homes, bridge structures, etc. They continue to be based on a combination of factors that include understanding building requirements, reliable static and dynamic testing of products, material reliability, public acceptability, feasibility, keeping up-to-date on customer problems, product identification, quality control establishment of engineering standards, approval of regulatory agencies, supervise installations, accurate cost and time estimation, organizational responsibility defined, meeting delivery schedule, development of proper marketing and sales approach, resolution of profit potential based on careful selection of application, economics, and acknowledge competition exists. The trend is to resist change in appearance so that plastic originally had to look like something else. With time passing the beauty and aesthetic of plastic was accepted and at the same time usually at lower costs and more benefits developed to the consumer (Figure 11.2.36).

This industry consumes about 20 wt% of all plastics produced. It is the second-largest consumer of plastics following packaging. This amount of plastics only represents about 5% of all materials consumed. Plastics will eventually significantly expand in this market. Its real growth will occur when plastic performance is understood by the building industry (meeting their specifications, etc.) and more important when the price is right in order to compete with other materials. Interesting that due to the destruction caused by wars, particularly WW II, use of plastics in most of those countries significantly escalated due to their desire to

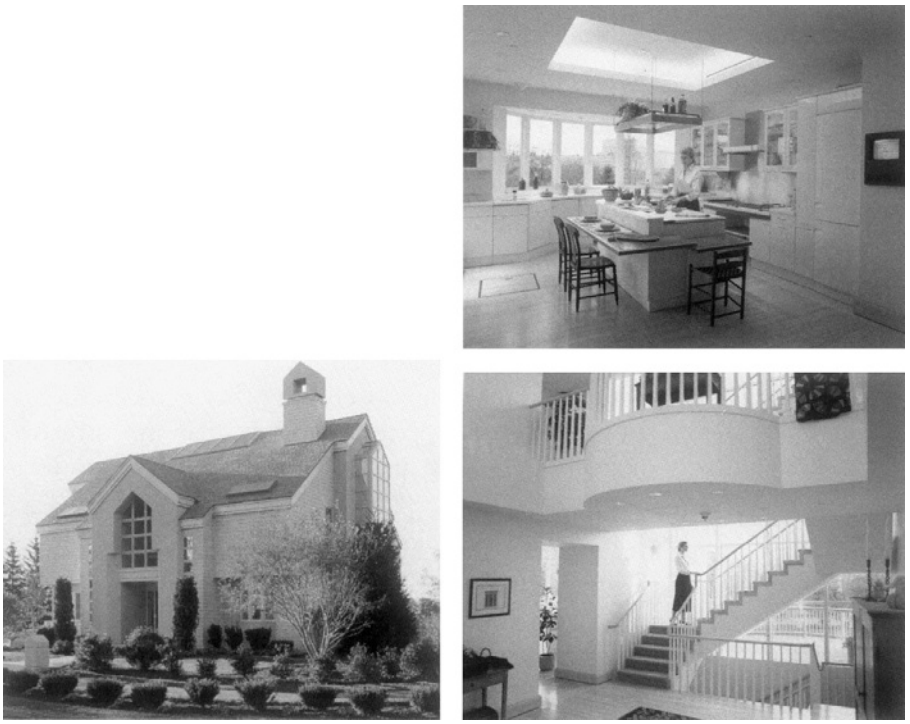


Fig. 11.2.36 All-plastic GE house in Pittsfield, MA. (Courtesy of GE.)

change specifications and standards allowing the use of plastics where it was technically beneficial.

The present and growing large market for plastics in building and construction is principally due to its suitability in different internal and external environments. The versatility of different plastics to exist in different environments permits the ability for plastics to be maintenance-free when compared to the more conventional and older materials such as wood. So it is said (factual) if wood did not have its excellent record of performances and costs for many centuries, based on present laws and regulations they could not be used. They burn, rot, etc. Regardless it would be ridiculous not to use wood. The different plastics inherently have superior properties that make them useful in other markets such as high strength and stiffness, durability, performances, insulation, cosmetics, etc. so eventually their use in building and construction will expand.

The plastic industry understands that the subjects important to the architect and builder includes information such as code requirements, setting up building standards and logical identification of over 25,000 plastics, static and dynamic load capabilities in structural parts, use of multifunctional parts, products provide new or better solutions to the myriad problems that plaguing the construction industry, performance data rather than sales type data, realistic and understandable technical data, limits on service life, fire safety, educating labor on benefits in using plastics, consumer acceptance, legal risk, competitive products, and costs.

Regarding codes unfortunately, particularly in the past, USA codes and other standards would contain requirements that certain plastics could meet with flying colors except those that would specify a specific material such as steel could only be used. So plastics could not be used until other materials such as plastics would be included or no specific material was specified. Outside USA the changes in most cases were immediate.

11.2.20.4 House of the future

Engineering ideas using unconventional approaches to building houses have evolved since at least the 1940s. Designers have applied spray- and solid-foamed homes that were originally introduced by the US Army during the 1950s. At the October 1965 annual National Decorative and Design Show in New York City products were on display of entire rooms and furnishings molded-in-place were found appealing and could survive the environment. Latter all structures (wall, roof, etc.) buildings were successfully built from interlocking extruded PVC hollowed sandwich-ribbed panels that were filled with insulation material (PUR foam, etc.), concrete, or other materials. Also building structures were made from extruded PS foam logs that were heat-bonded wrapped in dome shaped structures, RP filament wound room structures and support pillars, and so on Figures 11.2.37 and 11.2.38.

Monsanto's house of the future (Figure 11.2.37) located in Disney Land, FL was built during 1957 of practically all glass fiber-TS polyester RP; other plastics also used and reinforced concrete (foundation). MIT professors designed it. View (a) shows the four U-shaped monocoque structural cantilever beams each 16 ft from the square center foundation. View (b) shows load

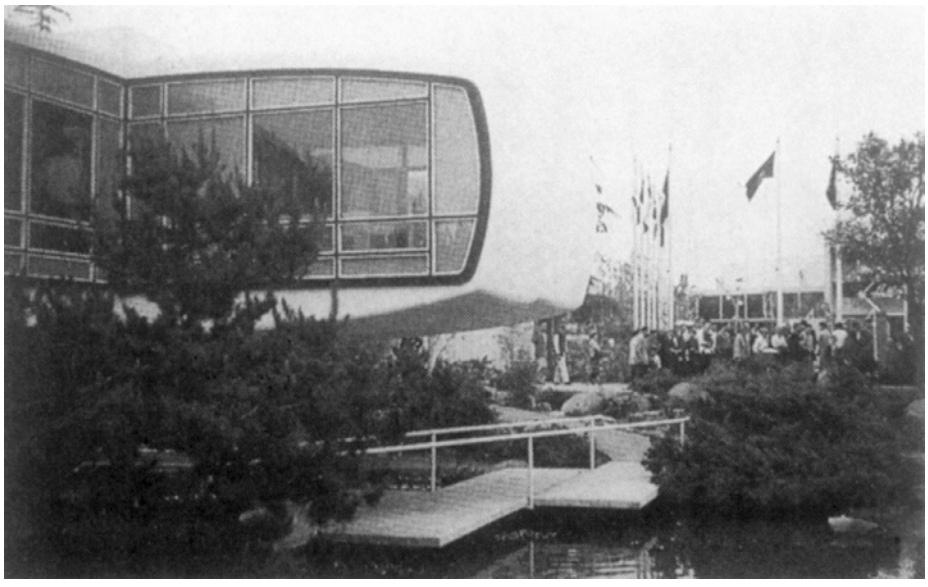


Fig. 11.2.37 Monsanto house in Disney land, California designed by MIT. (Courtesy of Monsanto.)

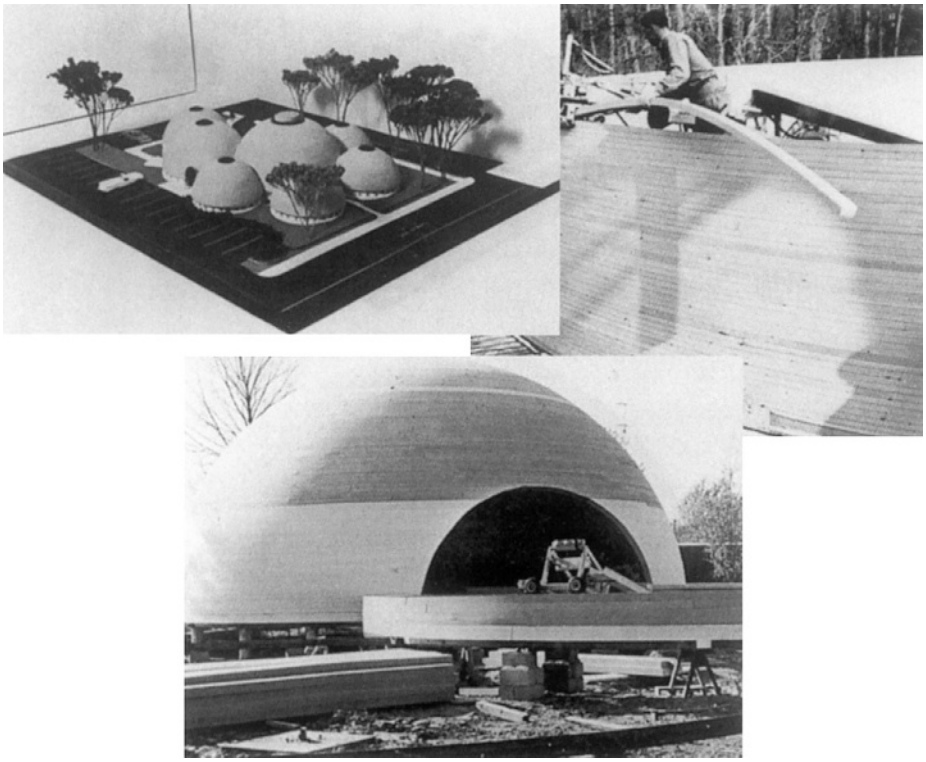


Fig. 11.2.38 Dow house of the future.

bearing requirements of the cantilever beams. View (c) is a sectional view of the house.

This main attraction was subjected to winds, tornadoes, earthquakes, and subjected to the movement of people that simulated a century of use. After two decades it was replaced to locate another scene. At that time it was determine the structure remained intact as it was produced initially. It basically did not deflect. It was almost impossible to destroy by the conventional wrecking procedures. They had to actually cut it up into small sections and then use wrecking balls, etc.

The DOW'S house of the future (Figure 11.2.38) located in Lafayette, IN was built during 1966 of practically all polystyrene foam board patented spiral generation technique producing high structurally loaded self-supporting domes. They provided their own insulation. View (a) is a model of what the finished building complex resembled. View (b) shows the layup procedure where the PS boards are heated on site using a heating device. The operation starts on a support circular steel foundation room. These layers bonded to each other with the applied heat formed the required contours with a continuous board structure.

These self-supporting domes required no internal or external support during or after manufacture. View (c) shows the ease in cutting windows, doors, connecting hallways from dome to dome, etc. The exterior of each dome was covered with steel wire mesh and concrete followed with waterproofing. Unfortunately its limitation in building structures was its potential fire hazard. However this basic concept with appropriate plastic foam material protection provides another approach to the low cost plastic house, etc.

11.2.20.5 House top

House top and top of other buildings is designed to take static loads and endure the outside environment such as sunlight, rain, wind, falling objects, temperature changes, hurricanes, and/or snow. It must also support loads imposed by people walking on it. This house top structure represents a simple type of a design project in static loading in that the loads are clearly long term and well defined. Creep effects can be easily predicted and the structure can be designed with a sufficiently large safety factor such as 3 to avoid the probability of failure. These type data (static, creep, etc.) are available of the material suppliers.

A roof design could use a translucent corrugated RP panel structure. This construction material has been used for roofs to admit daylight on a porch, shanty, canopy, and patio, shelter tops such as those used at a bus stop, industrial building, etc. Also used on transportation

vehicles such as a boat, truck, airplane, bus, and train. Corrugated materials are available in sheets of different sizes such as from 4 ft × 8 ft to as large as 10 ft × 20 ft. A typical material is 0.1 in. thick with 2–4 in. corrugations and a corrugation depth of 1–4 in. The usual material of construction is glass fiber-TS polyester RP. Standardized panels are used that have specific physical and mechanical properties. Usual attachment is by using nails or screws to wooden or pultruded plastic supports.

Design approach takes into account the weight of the material that is the major static load applied to the structure itself. Other weight such as snow is available from experience obtained in the area where the structure is to be used. Similarly, weights due to wind load and people load can be determined from experience factors that are generally known.

Main stress caused by the wind is at the anchorage points of the roof to the rest of the structure. They should be designed to take lifting forces as well as bearing forces. By using a lower angle of top structure, the less wind lifting stress. Proper anchorage of the support structure to the ground is also essential.

Localized loads represented by people walking on the roof can be solved by assuming concentrated loads at various locations such as between supports and by doing a short time solution to the bending problem and the extreme fiber stress condition. The local bearing loads and the localized shear should also be examined since it may cause possible local damage to the structure.

Requirements for the design of the structure included factors such as building and fire codes, and those that are reported by the owner of the property. As an example the material selected is to take the required loads without severe sagging (amount to be specified) for a 20 year period with no danger that the structure will collapse due to excessive stress on the material. One typical way in which excessive loading for a single section is handled is to bond two layers of the corrugated panel together with the corrugations crossed. This approach result in a very stiff section capable of substantially greater weight bearing than a single sheet and it will meet the necessary requirements. It also provides thermal insulation because of the trapped air space between the two layers particularly if they are edged sealed.

When using a large surface area of the RP panels expansion with a temperature rise occurs. Recognize that very few of the traditional building materials, such as wood, have significant expansion under normal temperature shifts. The RP materials, when compared to wood, does not have a problem because of its low thermal coefficients of expansion. What also keeps the dimensional changes low is that the corrugated shape can flex and accommodate the temperature changes. When compared to plastic materials such as polyvinyl chloride (PVC) siding, the expansion factor becomes significant and is an important consideration in the PVC method of fastening.

Data on the material (from the material supplier and/or data determined) will have been determined that it meets the static and dynamic loads required based on engineering analysis. Exposure to the environment (includes water and sunlight) can have a significant effect on the properties of the materials. In this example it is assumed that a 50% or more drop in the mechanical properties is expected to occur in 10 years. Based on knowledge of the materials behavior in service this loss of property levels off to a low rate of deterioration so that any potential failure will not occur in 20 years. If necessary increasing the panel strength requirements can compensate loss of properties by a suitable factor of safety that could be 3 in this example. However one knows that with a protective coating on the panel will minimize the effects of weathering so that only 20% loss in properties occurs after 10 years so that initial requirements for the panel performance will be reduced accordingly. The preferred type of coating could be a fluoropolymer that has the best resistance to sunlight and other weathering factors of all of the plastics and will last 20 years. If a coating is applied there usually is a requirement that the panel bc recoat after a specified period in service.

11.2.21 Transportation

Plastics and RPs offer a wide variety of benefits to the different transportation vehicles (automobiles, trucks, motorcycles, boats, airplanes, and so on). They range from assembling separate parts to providing safety (impact, etc.), aesthetics, lightweight, durability, corrosion resistance, recyclability, and fuel savings. Growth in the use of plastics tends to be a continuous process.

Structural analysis methods coupled with composite mechanics play a unique role during design in that they are used to formally evaluate alternative design concepts prior to committing some of these design concepts to costly testing or fabrication. Important material properties are not known during design in many cases. Frequently, some of these properties are assumed and structural analyses are performed to assess their effects on the design as well as their effects on the structural performance of the designed vehicle component. Also structural analysis methods are used to establish acceptable ranges of key factors (parameters) in a given design.

With RPs there can be the intraply hybrid composites have two kinds of fibers embedded in the matrix in general with the same ply providing different orientation performances. They have evolved as a logical sequel to conventional composites and to interply hybrids. Intraply

hybrid composites have unique features that can be used to meet diverse and competing design requirements in a more cost effective way than either advanced or conventional composites. Some of the specific advantages of intraply hybrids over other composites are balanced strength and stiffness, balanced bending and membrane mechanical properties, balanced thermal distortion stability, reduced weight and/or cost, improved fatigue resistance, reduced notch sensitivity, improved fracture toughness and/or crack-arresting properties, and improved impact resistance. With intraply hybrids, it is possible to obtain a viable compromise between mechanical properties and cost to meet specified design requirements.

Structural mechanics analyses are used to determined design variables such as displacements, forces, vibrations, buckling loads, and dynamic responses, including application of corresponding special areas of structural mechanics for simple structural elements. General purpose finite element programs such as NASTRAN are used for the structural analysis of complex structural shapes, large structures made from simple structural elements, and structural parts made from combinations of simple elements such as bars, rods, and plates.

Composite mechanics in conjunction with structural mechanics can be used to derive explicit equations for the structural response of simple structural elements. These explicit expressions can then be used to perform parametric studies (sensitivity analyses) to assess the influence of the hybridization ratio on structural response. For example, the structural response (behavior variables) equations for maximum deflection, buckling load, and frequency of a simply supported beam made from intraply hybrids are summarized in Figure 11.2.39. Flexural modulus is used to determine the maximum deflection, buckling load, and frequency of a simple supported beam made from intraply hybrids.

The notation in these equations and others that follow are as follows:

a_1, a_2	correlation coefficients for longitudinal compressive strength
B_{c1}, k	Buckling limit (buckling behavior constraint) due to loading condition k
B'_{c2}, k	strength limit (strength behavior constraint) due to loading condition k
B'_{c3}, k	interply delamination limit (delamination constraint) due to loading condition k
E_F	flexural modulus
E_{f11}, E_{f22}	longitudinal and transverse fiber moduli
E_{m11}, E_{m22}	in situ matrix longitudinal and transverse moduli
E_{pc}	modulus primary composite
E_{sc}	modulus secondary composite
G_{f12}	fiber shear modulus
G_{m12}	matrix shear modulus
H_j	matrix interply layer effect
I	moment of inertia
$K_{l12\alpha\beta}$	correlation coefficients in the combined-stress failure criterion: $\alpha, \beta = T$ or C denoting stress direction
k_f	fiber volume ratio
k_v	void volume ratio
k	$K = 1, 2, 3$, denotes loading condition index
l	length
N	number of plies
N_{xk}, N_{yk}	inplane loads – x and y directions corresponding to k

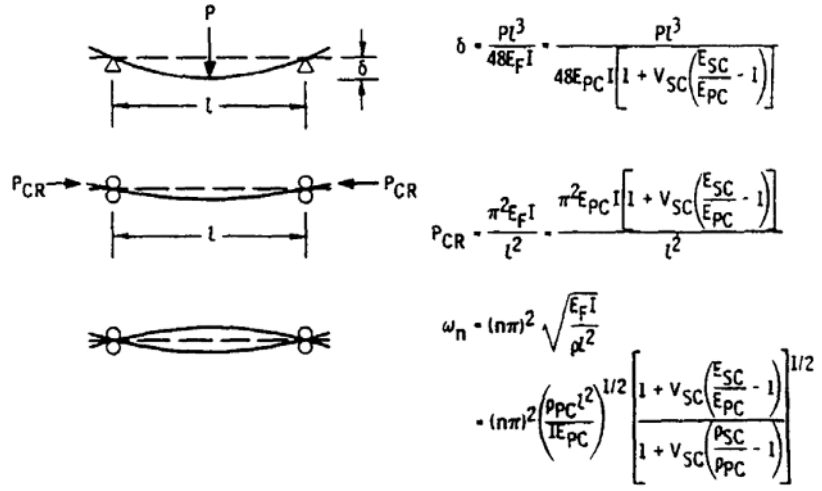


Fig. 11.2.39 Loads related to flexural modulus.

P	load
P_{cr}	buckling load
P_{cro}	buckling load of reference composite
S_{ft}	fiber tensile strength
S_{pc}	strength primary composite
S_{sc}	strength secondary composite
V_{sc}	volume fraction secondary composite
W	panel cost units per unit area
β, β'	correlation coefficients in composite micromechanics to predict ply elastic behavior
β_{del}	interply delamination factor
δ	displacement
δ_0	displacement of reference composite
ε_{mpc}, S, T	in situ allowable matrix strain for compression, shear and tension
θ	ply angle measured from x-axis
ν_f	fiber Poisson's ratio, numerical subscripts denote direction
ν_m	matrix Poisson's ratio, numerical subscripts denote direction
ω_n	frequency of the nth vibration mode
ω_{n0}	frequency of the nth vibration mode of the reference composite.

The equations are first expressed in terms of E_f, the equivalent flexural modulus, and then in terms of the moduli of the constituent composites (E_{pc} and E_{sc}) and die secondary composite volume ratio (V_{sc}). These equations were used to generate the parametric nondimensional plots shown in Figures 11.2.40–11.2.42.

The nondimensionalized structural response is plotted versus the hybridizing ratio V_{sc} for four different intraply hybrid systems. These figures show that small amounts of secondary composite (V_{sc} 0.2) have negligible effect on the structural response. However, small amounts of primary composite (V_{sc} 0.8) have a substantial effect on the structural response. A parametric plot of Izod-type, impact energy density is illustrated in Figure 11.2.43. This parametric plot shows also negligible effects for small hybridizing ratios (V_{sc} 0.2) and substantial effects for hybridizing ratios (V_{sc} 0.2).

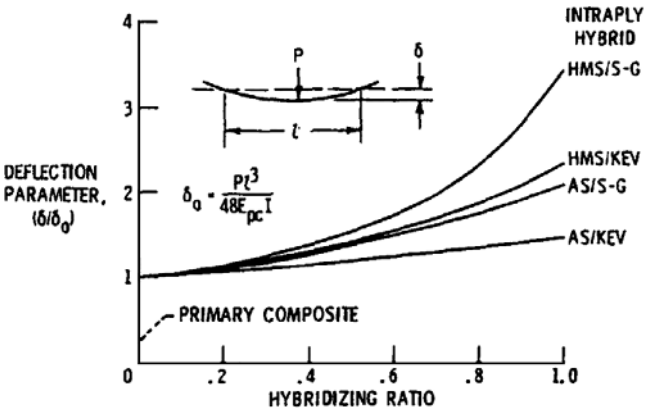


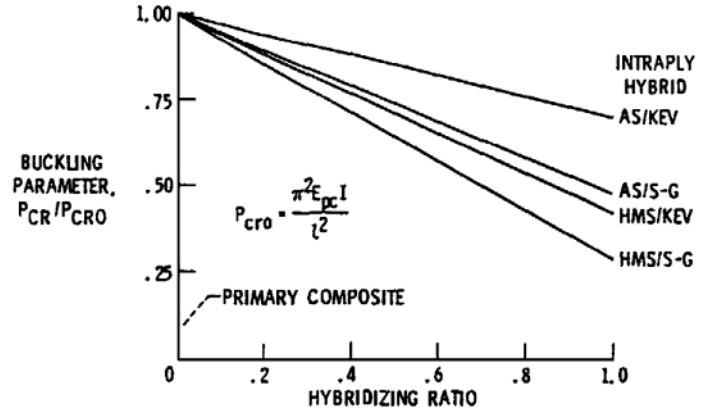
Fig. 11.2.40 Effects of hybridizing ratio and constituent composites on center deflection of intraply hybrid composite beams.

The parametric curves in Figures 11.2.40–11.2.43 can be used individually to select hybridization ratios to select a particular design requirement or they may be used jointly to satisfy two or more design requirements simultaneously, for example, frequency and impact resistance. Comparable plots can be generated for other structural components, such as plates or shells. Also plots can be developed for other behavior variables (local deformation, stress concentration, and stress intensity factors) and/or other design variables (different composite systems). This procedure can be formalized and embedded within a structural synthesis capability to permit optimum designs of intraply hybrid composites based on constituent fibers and matrices.

Low-cost, stiff, lightweight structural panels can be made by embedding strips of advanced unidirectional composite (UDC) in selected locations in inexpensive random composites. For example, advanced composite strips from high modulus graphite/resin, intermediate graphite modulus/resin, and Keviar-49 resin can be embedded in planar random E-glass/resin composite. Schematics showing two possible locations of advanced UDC strips in a random composite are shown in Figure 11.2.44 to illustrate the concept. It is important to note that the embedded strips do not increase either the thickness or the weight of the composite. However, the strips increase the cost.

It is important that the amount, type and location of the strip reinforcement be used judiciously. The determination of all of these is part of the design and analysis procedures. These procedures would require composite mechanics and advanced analyses methods such as finite element. The reason is that these components are designed to meet several adverse design requirements simultaneously. Henceforth, planar random composites reinforced with advanced composite strips will be called strip hybrids. Chamis and Sinclair give a detailed description of strip hybrids. Here, the discussion is limited to some design guidelines inferred from several structural responses obtained by using finite element structural analysis. Structural responses of panels structural components can be used to provide design guidelines for sizing and designing strip hybrids for aircraft engine nacelle, windmill blades and auto body applications. Several examples are described below to illustrate the procedure.

The displacement and base material stress of the strip hybrids for the concentrated load, the buckling load, and the lowest natural frequency are plotted versus reinforcing strip modulus in Figure 11.2.45. As can be seen



ν_f	Fiber Poisson's ratio, numerical subscripts denote direction.
ν_m	Matrix Poisson's ratio, numerical subscripts denote direction.
ω_n	Frequency of the nth vibration mode.
ω_{no}	Frequency of the nth vibration mode of the reference composite.

Fig. 11.2.41 Effects of hybridizing ratio and constituent composites on buckling load of intraply hybrid composite beams.

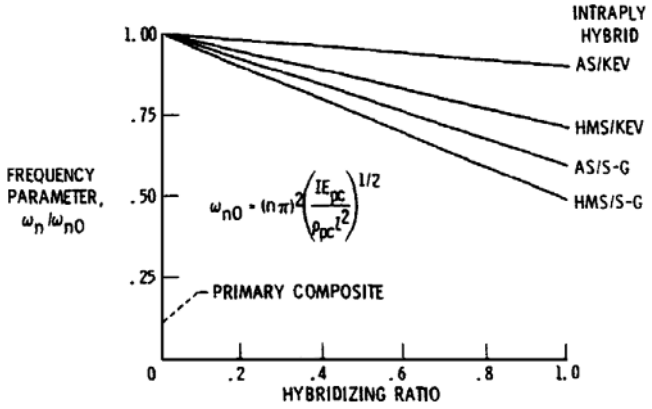


Fig. 11.2.42 Effects of hybridizing ratio and constituent composites on frequencies of intraply hybrid composite beams.

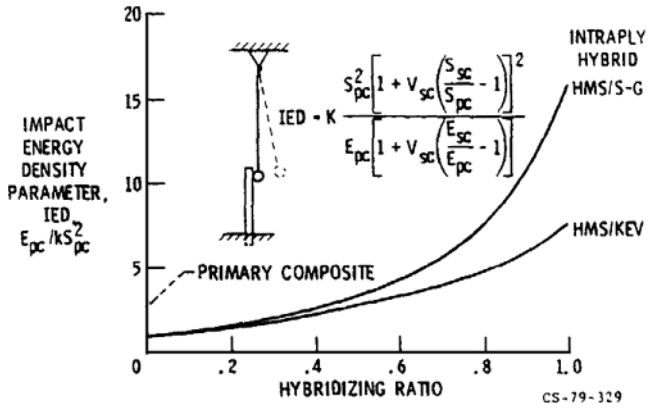


Fig. 11.2.43 Effects of hybridizing ratio and constituent composites on Izod-type energy density of intraply hybrids.

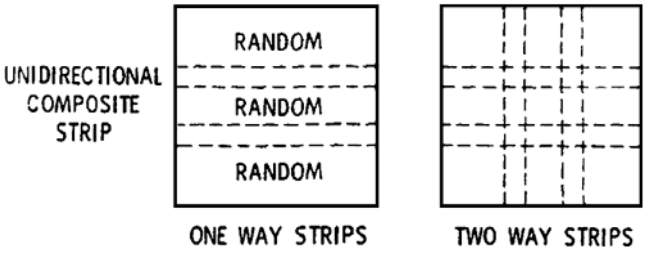


Fig. 11.2.44 Schematic of strip hybrids.

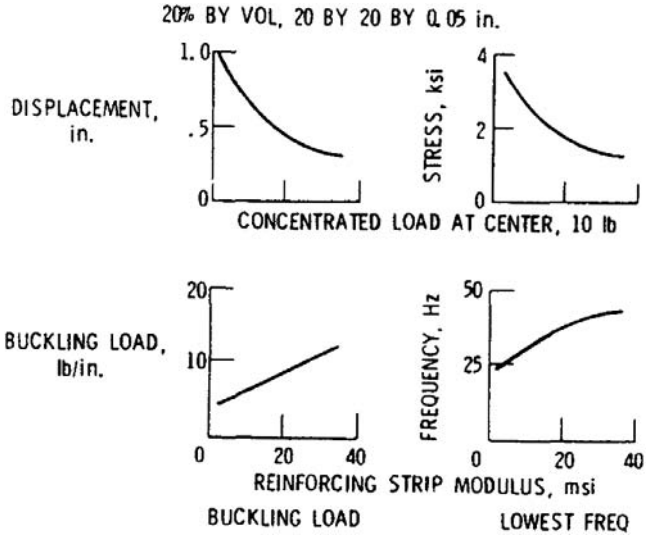


Fig. 11.2.45 Structural responses of strip hybrid plates with fixed edges.

the displacement and stress and the lowest natural frequency vary nonlinearly with reinforcing strip modulus while the buckling load varies linearly. These figures can be used to select reinforcing strip moduli for sizing strip hybrids to meet several specific design requirements. These figures are restricted to square fixed-end panels with 20% strip reinforcement by volume. For designing more general panels. suitable graphical data has to be generated.

The maximum vibratory stress in the base material of the strip hybrids due to periodic excitations with three different frequencies is plotted versus reinforcing strip modulus in Figure 11.2.46. The maximum vibratory stress in the base material varies nonlinearly and decreases rapidly with reinforcing strip modulus to about 103 GPa (15×10^6 psi). It decreases mildly beyond this modulus. The significant point here is that the modulus of the reinforcing strips should be about 103 GPa (15×10^6 psi) to minimize vibratory stresses (since they may cause fatigue failures) for the strip hybrids considered. For more general strip hybrids, graphical data with different percentage reinforcement and different boundary conditions are required.

The maximum dynamic stress in the base material of the strip hybrids due to an impulsive load is plotted in Figure 11.2.47 versus reinforcing strip modulus for two cases: (1) undamped and (2) with 0.009% of critical damping. The points to be noted from this figure are: (a) the dynamic displacement varies nonlinearly with reinforcing strip modulus and (b) the damping is much more effective in strip hybrids with reinforcing strip moduli less than 103 GPa (15×10^6 psi). Corresponding displacements are shown in Figure 11.2.48. The behavior of the dynamic displacements is similar to that of the stress as would be expected. Curves comparable to those in Figures 11.2.46 and 11.2.47 are needed to size and design strip hybrid panels so that

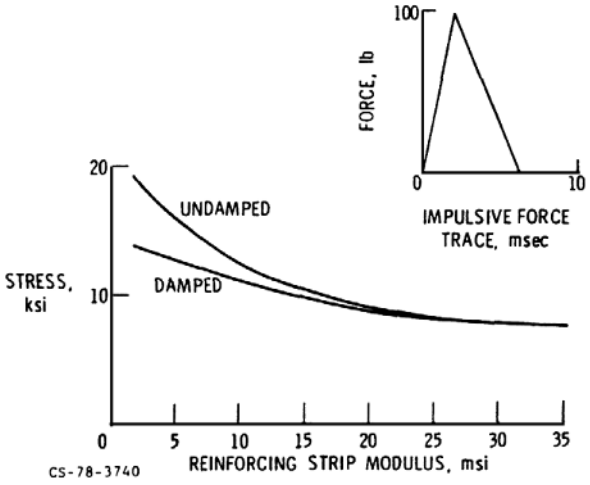


Fig. 11.2.47 Maximum impulse stress at center.

impulsive loads will not induce displacements or stresses in the base material greater than those specified in the design requirements or are incompatible with the material operational capabilities.

The previous discussion and the conclusions derived therefrom were based on panels of equal thickness. Structural responses for panels with different thicknesses can be obtained from the corresponding responses in Figure 11.2.47 as follows (let t = panel thickness):

1. Displacement due to a concentrated static load varies inversely with t^3 and the stress varies inversely with t^2.
2. The buckling load varies directly with t^3.
3. The natural vibration frequencies vary directly with t.

No simple relationships exist for scaling the displacement and stress due to periodic excitation or impulsive loading. Also, all of the above responses vary inversely

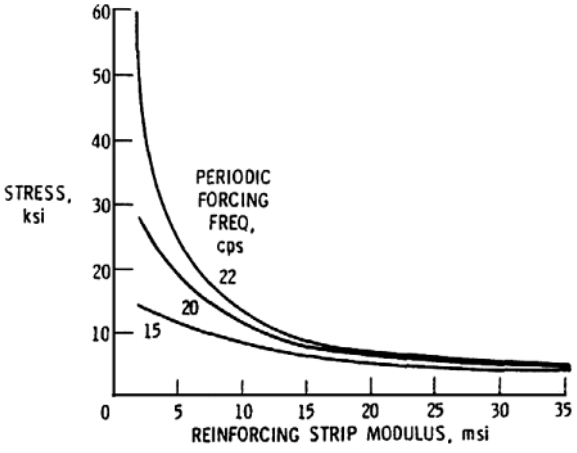


Fig. 11.2.46 Maximum stress in base material die to periodic vibrations.

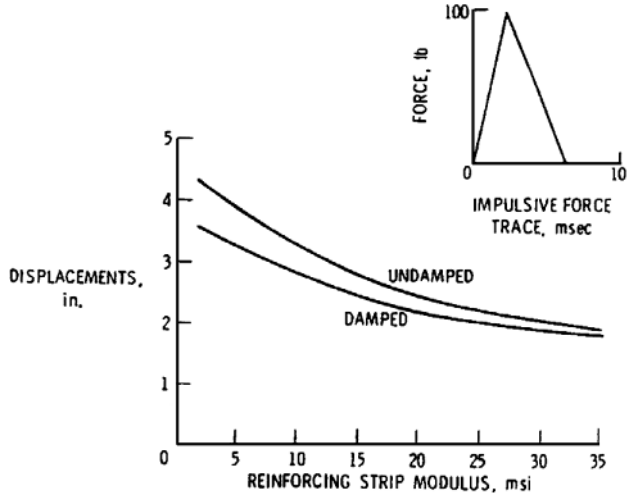


Fig. 11.2.48 Maximum impulse displacement.

with the square of the panel edge dimension. Responses for square panels with different edge dimensions but with all edges fixed can be scaled from the corresponding curve in Figure 11.2.45. The significance of the scaling discussed above is that the curves in Figure 11.2.45 can be used directly to size square strip hybrids for preliminary design purposes.

The effects of a multitude of parameters, inherent in composites, on the structural response and/or performance of composite structures, and/or structural components are difficult to assess in general. These parameters include several fiber properties (transverse and shear moduli), in situ matrix properties, empirical or correlation factors used in the micromechanical equations, stress allowables (strengths), processing variables, and perturbations of applied loading conditions. The difficulty in assessing the effects of these parameters on composite structural response arises from the fact that each parameter cannot be isolated and its effects measured independently of the others. Of course, the effects of single loading conditions can be measured independently. However, small perturbations of several sets of combined design loading conditions are not easily assessed by measurement.

An alternate approach to assess the effects of this multitude of parameters is the use of optimum design (structural synthesis) concepts and procedures. In this approach the design of a composite structure is cast as a mathematical programming problem. The weight or cost of the structure is the objective (merit) function that is minimized subject to a given set of conditions. These conditions may include loading conditions, design variables that are allowed to vary during the design (such as fiber type, ply angle and number of plies), constraints on response (behavior) variables (such as allowable stress, displacements, buckling loads, frequencies, etc.) and variables that are assumed to remain constant (preassigned parameters) during the design.

The preassigned parameters may include fiber volume ratio, void ratio, transverse and shear fiber properties, in situ matrix properties, empirical or correlation factors, structure size and design loads. Once the optimum design for a given structural component has been obtained, the effects of the various preassigned design parameters on the optimum design are determined using sensitivity analyses. Each parameter is perturbed about its preassigned value and the structural component is reoptimized. Any changes in the optimum design are a direct measure of the effects of the parameter being perturbed. This provides a formal approach to quantitatively assess the effects of the numerous parameters mentioned previously on the optimum design of a structural component and to identity which of the parameters studied have significant effects on the optimum design of the structural component of interest. The sensitivity

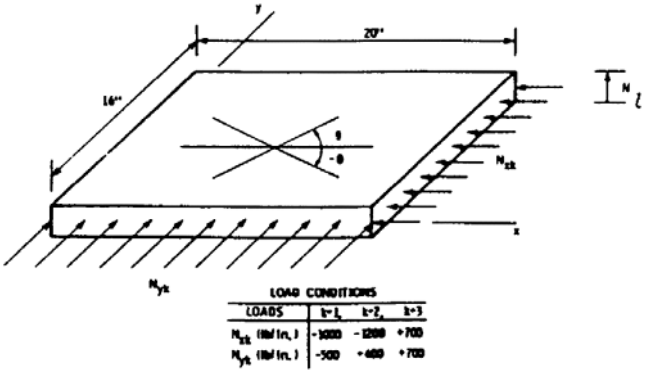


Fig. 11.2.49 Schematic of composite panel used in structural synthesis.

analysis results to be described subsequently were obtained using the angle plied composite panel and loading conditions as shown in Figure 11.2.49.

Sensitivity analyses are carried out to answer, for example, the following questions:

1. What is the influence of the preassigned filament elastic properties on the composite optimum design?
2. What is the influence of the various empirical factors/correlation coefficients on the composite optimum design?
3. Which of the preassigned parameters should be treated with care or as design variables for the multilayered-filamentary composite?
4. What is the influence of applied load perturbations on the composite optimum design?

The load system for the standard case consisted of three distinct load conditions as specified in Figure 11.2.49. The panel used is 20 in. × 16 in. made from an $[(+\theta)_n]_s$. angle plied laminate. The influence of the various preassigned parameters and the applied loads on optimum designs is assessed by sensitivity analyses. The sensitivity analyses consist of perturbing the preassigned parameters individually by some fixed percentage of that value which was used in a reference (standard) case. The results obtained were compared to the standard case for comparison and assessment of their effects.

Introductory approaches have been described to formally evaluate design concepts for select structural components made from composites including intraply hybrid composites and strip hybrids. These approaches consist of structural analysis methods coupled with composite micro-mechanics, FEA in conjunction with composite mechanics, and sensitivity analyses using structural optimization. Specific cases described include:

1. Hybridizing ratio effects on the structural response (displacement, buckling, periodic excitation and

impact) of a simply supported beam made from intraply hybrid composite.

2. Strip modulus effects on the structural response of a panel made from strip hybrid composite and subjected to static and dynamic loading conditions.
3. Various constituent material properties, fabrication processes and loading conditions effects on the optimum design of a panel subject to three different sets of biaxial in-plane loading conditions.

11.2.21.1 Automobile

Plastics play a very important role in vital areas of transportation technology by providing special design considerations, process freedom, novel opportunities, economy, aesthetics, durability, corrosion resistance, lightweight, fue1 savings, recyclability, safety, and so on. Designs include lightweight and low cost principally injection molded TP car body to totally eliminate metal structure to support the body panels such as the concept in Figure 11.2.50. Other processes include thermoforming and stamping. With more fuel-efficiency regulation new developments in lightweight vehicles is occurring with plastics. Plastics used include ABS, TPO, PC, PC/ABS, PVC, PVC/ABS, PUR, and RPs.

Different cars, worldwide have been designed and fabricated such as those that follow. (1) Chrysler's lightweight (50 wt% reduction) Composite Concept Vehicle (CCV) uses large injection molded glass fiber-TP structural body panels with only a limited amount of metal underneath/assembled by adhesive bonding or fusion welding. (2) Ford has plastic parts in its 2001 Explorer Sport Trac sport utility vehicle replaces the steel open cargo area with RP (SMC), and other cars. (3) Daimler-Benz's (Stuttgart, Germany) light-weight 2-seat coupe, called the Smart car, has injection molded outer body panels/unitizes TP body ties together the front fender, outer door panels, fiont panels, rear valence panels, and wheel arch in one wraparound package. (4) GM focusing with plastics in their electric vehicle. (5) Asha/Taisun of Singapore producing taxi cabs for China with thermoformed body panels mounted on a tubular stainless steel space frame. NA Bus Industries of Phoenix is delivering buses in USA and Europe with all RP bodies. Brunswick Tech. Inc. of Brunswick, ME produces-weight 30 ft RP buses except for the metallic engine. Sichuan Huatong Motors Group's (Chengdu, China) 4-door/5-passenger midsize vehicle all-plastic car, called Paradigm, has glass fiber-TS polyester RP sandwich chassis and thermoformed coextruded ABS body panels/chassis features single thermoformed lower tub and an upper skeleton X-brace roof/monocoque structure where body panels are stitched-bonded to the chassis, forming a unitized structure.

11.2.21.2 Truck

Since mid-1040s plastics and RPs have been used in trucks and trailers. In use are long plastic floors, side panels, translucent roofs, aeronautical over-the-cabin structures, insulated refrigerated trucks, etc. (that were initially installed on Strick Trailers by DVR during the late 1940s). The lighter weight plastic products permitted trailers to carry heavier loads, conserve fuel, refrigerated trucks traveled longer distance (due to improved heat insulation), etc. Different plastics continued to be used in the

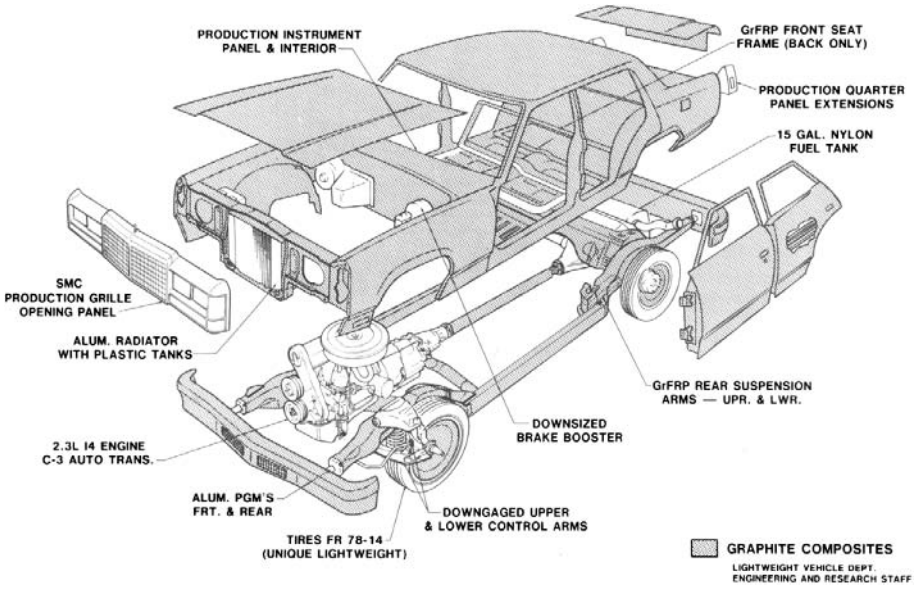


Fig. 11.2.50 Graphite fiber RP automobile. (Courtesy of Ford Co.)

different truck applications to meet static and dynamic loads that includes high vibration loads. Pickup trucks make use of 100 lb box containers using TPs and for the tougher requirements RPs are used.

11.2.21.3 Aircraft

Plastics continue to expand their use in primary and secondary aeronautical structures that include aircraft, helicopters, and balloons, to missiles space structures. Lightweight durable plastics and high performance RPs save on fuel while resisting all kinds of static and dynamic loads (creep, fatigue, impact, etc.) in different and extreme environments. Certain military planes contain up to 60 wt% plastics. Other airplanes take advantage of plastics performances such as the McDonald-Douglas AV-8B Harrier with over 26% of this aircraft's weight using carbon fiber-epoxy RPs; other plastics also used (Figure 11.2.51). Aircraft developments at the present time are extensively using cost-effective RPs and hybrid composites.

A historical event occurred during 1944 at U.S. Air Force, Wright-Patterson AF Base, Dayton, OH with a successful all-plastic airplane (primary and secondary structures) during its first flight. This BT-19 aircraft was designed, fabricated, and flight-tested in the laboratories of WPEFB using RPs (glass fiber-TS polyester hand lay-up that included the use of the lost-wax process sandwich constructions for the fabrication of monocoque fuselage, wings, vertical stabilizer, etc. Sandwich (cellular acetate foamed core) construction provides meeting the static and dynamic loads that the aircraft encountered in flight and on the ground. This project was conducted in case the aluminum that was used to build airplanes became unavailable. The wooden airplane, the Spruce Goose built by Howard Hughes was also a contender for replacing aircraft aluminum.

Extensive material testing was conducted to obtain new engineering data applicable to the loads the sandwich structures would encounter; data was extrapolated for long time periods. Short-term creep and fatigue tests conducted proved to be exceptionally satisfactory. Later 50 of this type of aircraft were built by Grumman Aircraft that also resulted in more than satisfactory technical performance going through different maneuvers.

In order to develop and maximize load performances required in the aircraft structures, glass fabric reinforcement laminated construction (with varying thickness) was oriented in the required patterns. Figure 11.2.52 shows an example of the fabric lay-out pattern for the wing structure. It is a view of a section of the wing after fabrication and ready for attachments, etc.

Developments of aircraft turbine intake engine blades that started during the early 1940s may now reach fulfillment. Major problem in the past has been to control the expansion of the blades that become heated during engine operation. The next generation of turbine

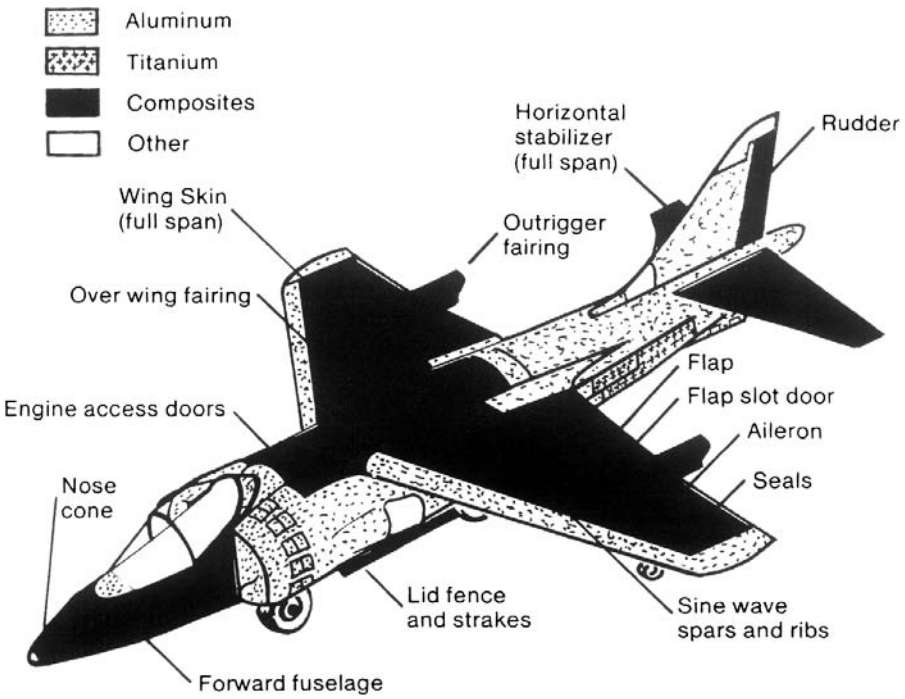


Fig. 11.2.51 McDonald-Douglas AV-8B Harrier plastic parts. (Courtesy of McDonald-Douglas.)

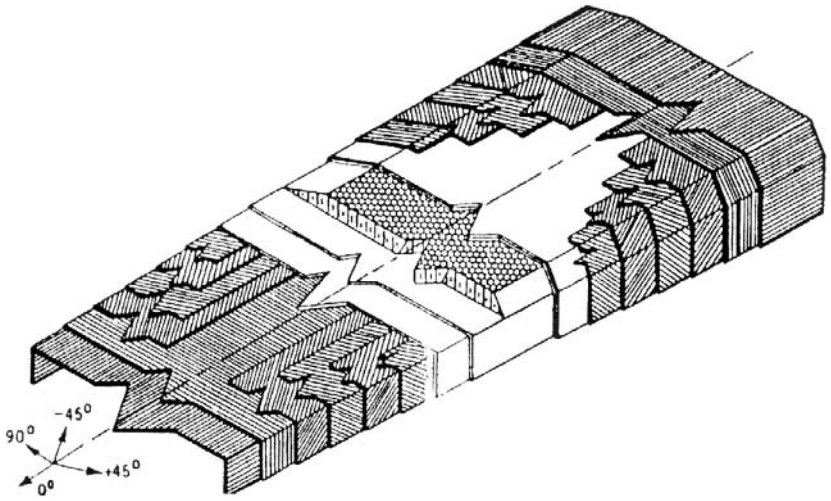


Fig. 11.2.52 Example of orientation of fibers (fabrics) in the all-plastic airplane wing construction.

fan blades should significantly improve safety and reliability, reduce noise, and lower maintenance and fuel costs for commercial and military planes because engineers will probably craft them from carbon fiber RP composites. Initial feasibility tests by University of California at San Diego (UCSD) structural engineers, NASA, and the U.S. Air Force show these carbon composite fan blades are superior to the metallic, titanium blades currently used.

Turbine fan blades play a critical role in overall functionality of an airplane. They connect to the turbine engine located in the nacelle, a large chamber that contains wind flow to generate more power. These usually 6 ft long blades create high wind velocity and 80% of the plane's thrust.

It is reported that the leading cause of engine failure is damaged fan blades. Failure may occur from the ingestion of external objects, such as birds, or it may be related to material defects. If it's a metallic blade and it breaks, it can tear through the nacelle as well as the fuselage and damage fuel lines and control systems. When this happens, the safety of the aircraft and its passengers is threatened, and the likelihood of a plane crash increases.

In contrast, if an RP blade breaks, it simply crumbles to bits and does not pose a threat to the structure of the plane. However, breakage is less likely because composite materials are tougher and lighter than metallic blades and exhibit better fatigue characteristics. A multiengine plane can shut down an engine and continue to fly if a blade is lost and no other damage has occurred. A composite blade disintegrates into many small pieces because it is really just brittle graphite fibers held together in a plastic. A titanium blade, however, will fail at the blade root, causing large, 4- to 6-foot blades to fly through the air.

As designed, the RP blades are essentially hollow with an internal rib structure. These rib like vents direct, mix, and control airflow more effectively which reduces the amount of energy needed to turn the blades and cuts back on noise. Most engine noise actually comes from wind turbulence that collides with the nacelle. By directing air out the back of the fan blades, the noise can be reduced by a factor of 2. And by drawing more air into the blades, engine efficiency is improved by 20%.

There also exists embedded elastic dampening materials in the blades, which minimizes vibrations to improve resiliency. Because the blade is lighter and experiences lower centrifugal forces, further enhanced the blade's durability occurs. Small-scale wind tunnel tests show they last 10–15 times longer than any existing blade. The No. 1 maintenance task is the constant process of taking engines apart to check the blades.

These new blades should lend themselves to more efficient production techniques. If you use titanium, you need to buy a big block of it and machine it down to size, wasting a lot of material. As reported, this is very time consuming, and one has to worry about thermal warping. The RP allows for mass production. It is fabricated into a mold, making the process more precise and ensuring the blades are identical. NASA will test the new blades in large-scale wind tunnels at the NASA Glenn Research Center in Cleveland. If successful, they could see installation by 2004.

Over the years innovations in aircraft designs have given rise to more new plastic developments and have kept the plastics industry profits at a higher level than any other major market principally since they can meet different load and environmental conditions. Virtually all plastics have received the benefit of the aircraft industry's uplifting influence.

11.2.21.4 Marine application

From ships to submarines to mining the sea floor worldwide, certain plastics and RPs can survive the sea environment. This environment can be considered more hostile than that on earth or in space. For water surface vehicles, many different plastics have been used in designs in successful products in both fresh and the more hostile seawater. Boats have been designed and fabricated since at least the 1940s. Anyone can now observe that practically all boats, at least up to 9 m (30 ft) are made from RPs that are usually hand lay-up moldings from glass rovings, chopper glass pray-ups, and/or glass fiber mats with TS polyester resin matrices. Because of the excellent performance of many plastics in fresh and seawater, they have been used in practically all structural and nonstructural applications from ropes to tanks to all kinds of instrument containers.

11.2.21.4.1 Boat

In addition to their use in boat hull construction, plastics and RPs have been used in a variety of shipboard structures (internal and external). They are used generally to save weight and to eliminate corrosion problems inherent in the use of aluminum and steel or other metallic constructions.

Plastic use in boat construction is in both civilian and military boats (28–188 ft. (8.5–55 m)). Hulls with nontraditional structural shapes do not have longitudinal or transverse framing inside the hull. Growth continues where it has been dominating in the small boats and continues with the longer boat boats. The present big boats that are at least up to 188 ft long have been designed and built in different countries (USA, UK, Russia, Italy, etc.). In practically all of these boats low pressure RP molding fabrication techniques were used.

Examples of a large boat are the U.S. Navy's upgraded minehunter fleet, the "Osprey" class minehunter that withstands underwater explosions. Design used longitudinal or transverse framing inside the piece hull. It has a one piece RP super structure. Material of construction used was glass fiber-TS polyester plastic. The designer and fabricator was Interimarine S.P.A., Sarzana, Italy. The unconventional, unstiffened hull with its strength and resiliency was engineered to deform elastically as it absorbs the shock waves of a detonated mine. Its design requirements included to simplify inspection and maintenance from within the structure.

11.2.21.4.2 Underwater hull

On going R&D programs continue to be conducted for deep submergence hulls. Materials of construction are usually limited to certain steels, aluminum, titanium, glass, fiber RPs, and other composites (Figure 11.2.53). There is a factor relating material's strength-to-weight characteristics to a geometric configuration for a specified design depth. Ratio showing the weight of the pressure hull to the weight of the seawater displaced by the submerged hull is the factor referred to as the weight displacement (W/D) ratio. Submergence materials show the variation of the collapse depth of spherical hulls with the weight displacement of these materials. All these materials, initially, would permit building the hull of a rescue vehicle operating at 1800 m (6000 ft) with a collapse depth of 2700 m (9000 ft).

When analyzing materials for an underwater search vehicle operating at 6000 m (20,000 ft) with collapse depth of 9000 m (30,000 ft), metals are not applicable. Materials considered are glass and RP. The strength-to-weight values for metals potentially are not satisfactory. One of the advantages of glass is its high compressive strength; however, one of its major drawbacks is its lack

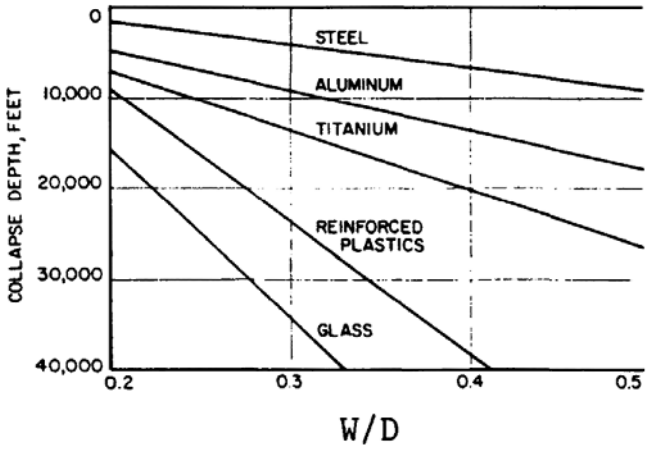


Fig. 11.2.53 Examples of materials for deep submergence vehicles.

of toughness and destructive effect if any twist, etc. occurs other than the compression load. It also has difficulty if the design requires penetrations and hatches in the glass hull. A solution could be filament winding RP around the glass or using a tough plastic skin.

These glass problems show that the RP hull is very attractive on weight-displacement ratio, strength-weight ratio, and for its fabrication capability. By using the higher modulus and lower weight advanced designed fibers (high strength glass, aramid, carbon, graphite, etc.) additional gains will occur.

Depth limitations of various hull materials in near-perfect spheres superimposed the familiar distribution curve of ocean depths. To place materials in their proper perspective, as reviewed, the common factor relating their strength-to-weight characteristics to a geometric configuration for a specified design depth is the ratio showing the weight of the pressure hull to the weight of the seawater displaced by the submerged hull. This factor is referred to as the weight displacement (*W/D*) ratio. The portions the vehicles above the depth distribution curve correspond to hulls having a 0.5 *W/D* ratio; portion beneath showing the depth attainable by heavier hulls with a 0.7 *W/D*.

Based on test programs the ratio of 0.5 and 0.7 is not arbitrary. For small vehicles they can be designed with *W/D* ratios of 0.5 or less, and vehicle displacements can become large as their *W/D* approach 0.7. By using this approach these values permits making meaningful comparisons of the depth potential for various hull materials. With the best examination data reveals that for the metallic pressure-hull materials, best results would permit operation to a depth of about 18,288 m (20,000 ft) only at the expense of increased displacement. RPs (those with just glass fiber-TS polyester plastic) and glass would permit operation to 20,000 ft or more with minimum displacement vehicles.

The design of a hull is a very complex problem. Under varying submergence depths there can be significant working of the hull structure, resulting in movement of the attached piping and foundation. These deflections, however slight, set up high stresses in the attached members. Hence, the extent of such strain loads must be considered in designing attached components.

11.2.22 Missile and rocket

Different plastics, particularly high-performance plastics and RPs are required in missiles (Figure 11.2.54) and rockets as well as outer space vehicles. Parts in a missile are very diverse ranging from structural and non-structural members, piping systems, electrical devices, exhaust insulators, ablative devices, personnel support equipment, etc.

11.2.23 Electrical/electronic

With the diverse electrical properties of plastics, extensive use of plastics has been made since the first plastics was produced. Plastics permits the operation of many electrical and electronic devices worldwide. As it has been said many times most of the electrical/electronic equipment and devices used and enjoyed today would not be practical, economical, and/or some even possibly exist without plastics. Plastics offer the designer a great degree of freedom in the design and particularly the fabrication of products requiring specific electrical properties and usually requiring special and accurately fabricated products. Their combination of mechanical and electrical properties makes them an ideal choice for everything from micro electronic components and fiber optics to large electrical equipment enclosures.

Development of many different polymers and plastic compounds (via additives, fillers, and reinforcements) continues to expand the use of plastics in electrical applications. By including fillers/additives, such as glass in plastics, electrical properties can considerably extend performances of many plastics (Figure 11.2.55).

The electrical propertics of plastics vary from being excellent insulators to being quite conductive in different environments. Depending on the application, plastics may be formulated and processed to exhibit a single property or a designed combination of electrical, mechanical, chemical, thermal, optical, aging properties, and others. The chemical structure of polymers and the various additives they may incorporate provide compounds to meet many different performance requirements.

Plastic provides ideas for advancing electrical and electronic systems from conducting electricity to the telephone to electronic communication devices. Thousands of outstanding applications use plastics in electrical products. The users' and designers' imaginations have excelled in developing new plastic products.

11.2.23.1 Shielding electrical device

With the extensive use of plastics in devices such as computers, medical devices, and communication equipment the issue of electromagnetic compatibility (EMC) exists that in turn relate to electromagnetic interference (EMI) and radio-frequency interference (RFI).

EMC identifies types of electrical device's capability to function normally without interference by any electrical device. These devices are designed to minimize risks associated with reasonably foreseeable environmental conditions. They include magnetic fields, external electrical influences, electrostatic discharge, pressure,

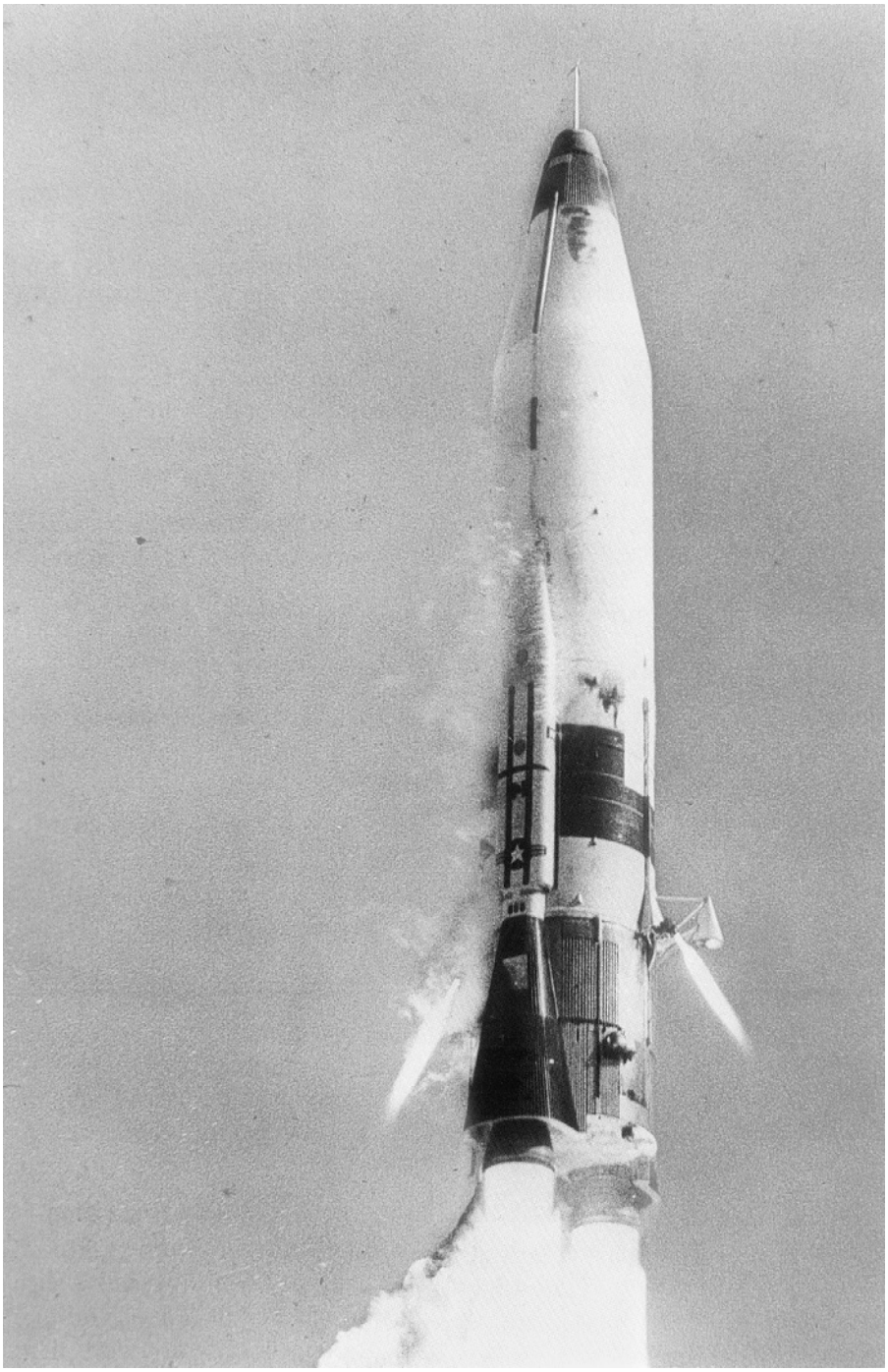


Fig. 11.2.54 Missile in flight includes the use of plastics.

temperature, or variations in pressure and acceleration, and reciprocal interference with other devices normally used in investigations or treatment.

EMI or RFI as well as static charge is the interference related to accumulated electrostatic charge in a nonconductor. As electronic products become smaller and more powerful, there is a growing need for higher shielding levels to assure their performance and guard against failure. From the past 40 dB shielding, the 60 dB is becoming the normal higher value. There is EMI shielding-effectiveness that defines the ratio of the incident electrical field strength to the transmitted electrical field strength. Frequency range is from 30 MHz to 1.5 GHz (ASTM D 4935-89).

Many plastics are electrical insulators because they are nonconductive. They do not shield electronic signals

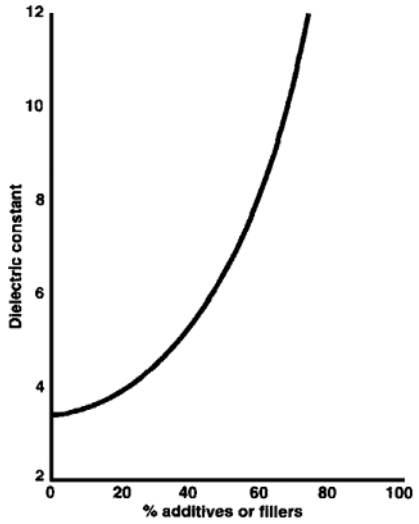


Fig. 11.2.55 Dielectric constant.

generated by outside sources or prevent electromagnetic energy (EME) from being emitted from equipment housed in a plastic enclosure. Government regulations have been set up requiring shielding when the operating frequencies are greater than 10 kHz.

The plastic shielding material used may include the use of additives. Designs may include board-level shielding of circuit, bondable gaskets, and locating all electrical circuits in one location so only that section requires appropriate shielding. Designers of enclosures for electronic devices should be aware of changes in EMC that tend to continually develop worldwide.

Conductive plastics provide EMI/RFI shielding by absorbing EME and converting it into electrical or thermal energy. They also function by reflecting EME. This action ensures operational integrity and EMC with existing standards. Conductive plastics are generally designed to meet specific performance requirements (physical, mechanical, etc.) in addition to EMI/RFI or static control. Often these plastics have to perform structural functions, meet flammability or temperature standards, and provide wear or corrosive resistant surfaces, etc.

The usual plastics alone lack sufficient conductivity to shield EMI and RFI interference. Designers can reduce or eliminate sufficiently electromagnetic emissions from plastic housings like those of medical devices and computers just by shielding the inner emission sources with metal shrouds in the so-called tin can method. The same effect can be obtained by designing electronics to keep emissions below standard limits or by incorporating shielding into the plastic housing itself. Designers will often employ all these strategies in a single design. What is most important is to attempt to locate all the shielding in a relatively small volume within the larger housing and then tin can it to provide a simplified solution rather than spreading it out.

Every electronic system has some level of electromagnetic radiation associated with it. If this level is strong enough to cause other equipment to malfunction, the radiating device will be considered a noise source and usually subjected to shielding regulations. This is especially true when EMI occurs within the normal frequencies of communication. When the electronic noise is sufficient to cause malfunctioning in equipment such as data processing systems, medical devices, flight instrumentation, traffic control, etc. the results could prove damaging and even life threatening. Reducing the emission of and susceptibility to EMI or RFI to safe levels is thus the prime reason to shield medical devices (and other devices) in whatever type of housing exist, including plastic.

In addition to compounding additives for shielding, there is the technology of applying conductive coatings, such as vacuum systems or paint systems (sprays, etc.). Other methods include the use of conductive foils or molded conductive plastics, silver reduction, vacuum metalization, and cathode sputtering. Although zinc-arc spraying once accounted for about half the market, others have surpassed it. Other conductive coatings are also used. Unlike other shielding methods, conductive coatings are usually applied to the interiors of housings and do not require additional design efforts to achieve external aesthetic goals. All systems offer trade-offs in shielding performance, the physical properties of the plastics, ease in production, and cost.

Designers have to confirm the suitability of a material's shielding performance for each system through such conventional means as screen-room or open-field testing. Each approach to shielding should also be subjected to simulated environmental conditions, to determine the shield's behavior during storage, shipment, and exposure to humidity. Sometimes comparison of shielding materials becomes difficult. ASTM has a standard that defines the methods for stabilizing materials measurement, thus allowing relative measurements to be repeated in any laboratory. These procedures permit relative performance ranking, so that comparisons of materials can also be made.

Organizations involved in conducting and/or preparing specifications/standards on the electrical properties on plastics include the Underwriters Laboratories (UL), American Society for Testing and Materials (ASTM), Canadian Standards Association (CSA), International Electrotechnical Commission (IEC), International Organization for Standardization (ISO), and American National Standards Institute (ANSI).

UL has a combination of methods for environmental conditioning and adhesion testing to evaluate various approaches to shielding and to determine the plastic

types that are suitable. The primary concern is safety. Should a metalized plastic delaminate or chip off, an electrical short is formed that could cause a fire.

11.2.23.2 Radome

Radome (radiation dome) is used to cover a microwave electronic communication antenna. It protects the antenna from the environment such as the ground, underwater, and in the air vehicles. To eliminate any transmission interference, it would be desirable not to use a radome since transmission loss of up to 5% occurs with the protective radome cover material. The radome is made to be as possibly transparent to electromagnetic radiation and structurally strong. Different materials can be used such as plastics, wood, rubber-coated air-supported fabric, etc. To meet structural load requirements such as an aircraft radome to ground radomes subjected to wind loads, use is made of RPs that are molded to very tight thickness tolerances. Figure 11.2.56 shows a schematic of a typical ground radome that protects an antenna from the environment (withstand over 150 mph winds and temperatures from arctic to tropical conditions, sand/dirt, etc.) using RP-honeycomb sandwich curved panels. This schematic represents protecting in service 150 ft (46 m) antennas. Since that time the most popular is the use of glass fiber-TS polyester RPs. The shape of the dome, that is usually spherical, is designed not to interfere with the radiation transmission.

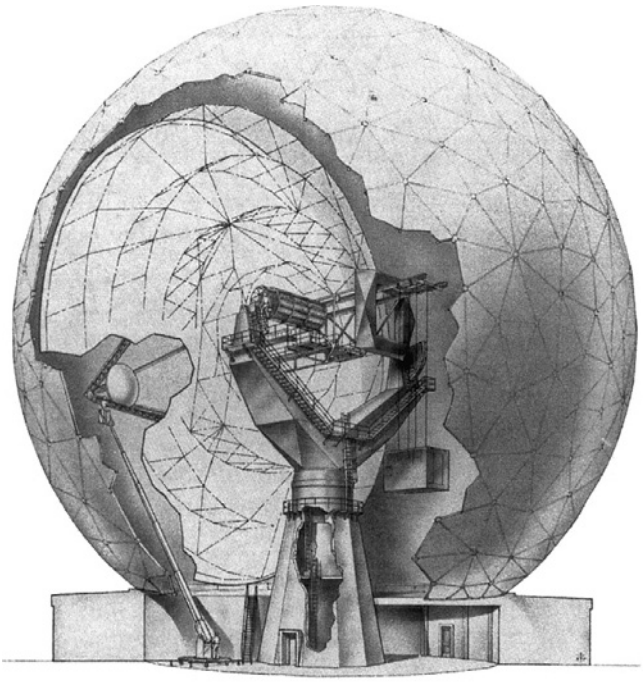


Fig. 11.2.56 Antenna (150 ft) protected by a plastic radome.

The use of the secondary load structure RP aircraft radomes have been used since the early 1940s. At that time the problem of rain erosion developed on their front of the radome. It first appeared on the RP "eagle wing" radome located below the B-29 bomber aircraft. It had an airfoil-shaped radome that was 6 m (20 ft) long located about 0.5 m (11 ft) below its wing. On its initial flight over the Pacific Ocean upon encountering rain, the RP radome (and its radar capacity) was completely destroyed. This introduced the era of rain erosion damage to plastics in using a rain erosion elastomeric plastic coating.

11.2.24 Medical

Plastic devices of all types have become vital in the medical industry. Products range from disposables (medical supplies, drug delivery devices, ointments, etc.) to non-disposables and packages to containers to body parts. Packaged drugs include premeasured single-dose disposable units. The diversified properties and behaviors of plastics have developed into an important market for plastics.

Plastic applications are very diversified ranging from band-aids to parts of the heart. Many examples exist in addition to those being reviewed.

1. Abiomed (Danvers, MA) is the manufacturer of the self-contained artificial heart. Of the six patients who received the grapefruit-sized plastic and titanium heart, three remain alive (2003).
2. Developments have found certain plastics existing in the environment of living tissues.
3. The heart valve that is often used in surgery to correct heart deficiencies was a contribution to medicine. In order for it to be successful it required ingenuity in design that would function as a replacement for the mitral valve and to perform as well as the one replaced long enough to justify the risk involved in the operation. It also included using a bioplastic material that would function in the highly complex environment of the human circulatory system without being degraded and without causing harm to the circulatory system.
4. Many developments occur in the area of implants that include the use of plastics. Examples include pacemaker, surgical prosthesis devices to replace limbs, use of plastic tubing to support damaged blood vessels, and work with the portable artificial kidney.
5. There are applications based on the membrane qualities of plastics. They can control such things as the chemical constituents that pass from one part of

a system to another, the electrical surface potential in a system, the surface catalytic effect on a system, and in some cases the reaction to specific influences such as toxins or strong radiation.

6. Polyelectrolytes plastics are chemically active. They have been used to make artificial mechanical power muscle materials. They create motion by the lengthening and shortening of fibers made from the chemically active plastic by changing the composition of the surrounding liquid medium, either directly or by the use of electrolytic chemical action. It is no a competitor to thermal energy sources, but it is potentially valuable in detector equipment that would be sensitive to the changing composition of a water stream or other environmental flow situation.
7. There is the application of extruded high- and low-pressure plastic balloons used in angioplasty catheters. Use of these balloons has extended into many applications such as other catheters (dilatation, heat transfer, laser, cryogenic, etc.), photodynamic therapy devices, drug delivery devices, etc.
8. Plastics in bioscience have potential to be used for mechanical implants in living systems (includes animals and plants) where they can serve as repair parts or as modifications of the system.
9. Kidney applications involve more than the mechanical characteristics of potential plastic use. The kidney machine consists of large areas of a semi-permeable membrane, a cellulosic material in some machines, where the kidney toxins are removed from the body fluids by dialysis based on the semi-permeable characteristics of the plastic membrane. Different plastics are being studied for use in this area, but the basic unit is a device to circulate the body fluid through the dialysis device to separate toxic substances from the blood. The mechanical aspects of the problem are minor but do involve supports for the large amount of membrane required.
10. Surgical implants are essentially plastic repair parts for worn out parts of the body. It is possible to conceive of major replacements of an entire organ such as a kidney or a heart by combining the plastic skills with tissue regeneration efforts that may extend life. This is used to time the heart action. Extensively used are plastic corrugated, fiber (silicone or TP polyester) braided aortas.
11. Different customized developments exist and are being used. An example is a porous (foam type) ultrahigh-molecular-weight poly-ethylene (UHMWPE) that is an FDA-compliant material. Its porosity and pore size can be adjusted per the end-users' requirements. The porosity is uniform in all three (X, Y, Z) axes, which is vital to constant liquid flow in filtration and separation. It is already used in a wide range of medical and laboratory filtration and separation applications by providing customizing processing where chemical purity of the material is maintained (no additives are used). UHMWPE is a chemically resistant, long-chain polymer of ethylene with an extremely high molecular weight of 3.1 million amu or above. Because of its high molecular weight, the polymer maintains abrasion resistance and strength even when it is made porous for filtration or separation applications. The porous form, which can be pleated, is easily handled in manufacturing. It can be precision skived into films as thin as 0.002 in. (skiving consists of shaving off a thin film layer from a large block of solid plastic, usually a round billet).

Doctors with long, intensive training as basic scientists make them uniquely suitable as product designers. They become involved in designing new products that in turn could require the plastic industry in developing new/modified plastics. With all this action in developing medical products and devices the FDA usually requires approval that takes time.

11.2.24.1 Surgical product

The wide range of forms (film, tube, or fiber) and mechanical properties available in plastics continues to make them attractive candidates for such uses (Figure 11.2.57). These plastics are required to possess desired physical/mechanical properties and the assurance that they may be successfully utilized in the body. To be successful plastic implants involve the cross-fertilization of different disciplines (chemists, designers, physician, fabricator, etc.).

Developing of surgical implants confronts major problems because human bodies have extremely complex environments. They could be identified as having the most horrible environmental situation on earth. In addition different human bodies have different environmental requirements. Thus, what can survive in one body usually does not survive in other bodies. This type of reaction requires extensive R&D to ensure that a medical product can survive and meet its requirements in all human or specific bodies.

11.2.24.2 Dental product

Dentures continue to be an important and major market for plastics. Use is made of acrylics (PMMAs) that includes full dentures, partial dentures, teeth, denture reliners, fillings and miscellaneous uses. PMMAs provide

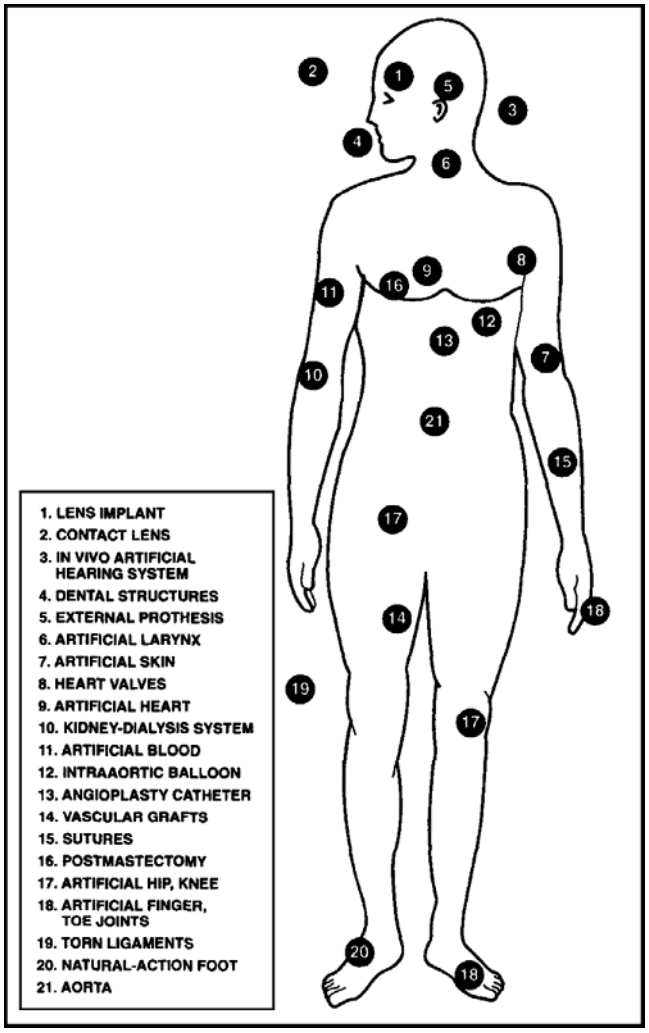


Fig. 11.2.57 Examples of surgical implants. (Courtesy of Plastics FALLO.)

strength, exceptional optical properties, low water absorption and solubility, and excellent dimensional stability. In the past plastics used included nitrocellulose, phenol-formaldehyde, and vinyls as denture base materials. Results, however, were not satisfactory because these plastics did not have the proper requisites of dental plastics. Since then, PMMAs have kept their lead as the most useful dental plastics, although many new plastics have appeared and others are being tested.

Todate plastics are not very useful as filling materials with only about 2 wt% of all fillings using plastics. The low mechanical properties of plastics in comparison with metals limit their application where stresses are great. It is interesting to note development efforts has taken place in the use of whiskers for reinforcing dental plastic, metal, and ceramic fillings. Some preliminary test results on the addition of randomly distributed chopped, short whiskers to a coating plastic have reversed the previous proportional loss of strength with powder additions. Although this is far from theoretical, it is already quite significant in that it allows the addition of pigment for coloring purposes and a restoration of the loss of strength with whisker additions.

11.2.24.3 Health care

Here they have made many major contributions to the contemporary scene. Health-care professionals depend on plastics for everything from intravenous bags to wheelchairs, disposable labware to silicone body parts, and so on. The diversity of plastics allows them to serve in many ways, improving and prolonging lives (such as a braided, corrugated (Du Pont's Dacron polyester) TP aorta tube).

Thousands of biodegradable plastics have been analyzed. An example is at the Massachusetts Institute of Technology (MIT) test program that may save lives in the form of a medical implant. Tests include its effectiveness as a drug-releasing implant in brain cancer patients, These implants, roughly the size of a quarter, are being placed in patients' brains to release the chemotherapy drug BCNU (Carmustine). Todate these biocompatible implants have been found to be safer than injections, which can cause the BCNU to enter the bone marrow or lungs, where the drug is toxic.

Use is being made of the polyanhydride plastic that was designed shapewise so that water would trigger its degradation but would not allow a drug to be released all at once. The implant degrades from the outside, like a bar of soap, releasing the drug at a controlled rate even as it becomes smaller. The rate at which the drug is released is determined by the surface shaped area of the implant and the rate of plastic degradation that is customized to release drugs at rates varying from one day to many years. This design approach also holds promise for use with different drugs for various other-medical systems. Interesting that in the past these plastics were used with explosives for use in a war. When the plastic degraded by water or sunlight, the device would explode.

11.2.25 Recreation

Plastics provide structurally sound, durability, and safe equipment for sports and recreation facilities. The broad range of properties available from plastics has made them part of all types of sports and recreational equipment worldwide for land, water, and airborne activities. Roller-skate wheels have become abrasion- and wear-resistant polyurethane, tennis rackets are molded from specially RPs (using glass, aramid, graphite, and/or hybrid fibers),

skis are laminated with RPs, RP pole-vaults extensively used to go "higher", and many more.

11.2.26 Appliance

By far most appliances use plastics to take advantage of the properties they inherently provide ranging from aesthetics to structural integrated (corrosion resistance, tough, etc.) to insulation. When examining the plastics in an appliance in addition too different plastics being used, there are the many different fabricating methods used to produce simple to complex parts.

Many success stories exist that include the benefits gained in the household and other places. In the past washing, drying, and ironing clothes was a rigorous, two-day affair involving the filling of metal tubs, scrubbing by hand, hanging clothes to dry, and heating cast-iron flat irons on a stove. With new technology and "plastics" laundry rooms and kitchens worldwide are operating in relatively a few minutes and looking better than ever before.

As one of many thousands of success examples, Milwaukee Electric Tool Corp. in the past found itself on the short end of the age-old supply-and-demand equation. That is, it was unable to keep up with demand for its heavy-duty electric power tools. Problem was that their machining operations could not turn out enough aluminum die-cast motor housings to keep up with market demand. The firm briefly considered what would have been a long-term solution; a state-of-the-art machining center. But a feasibility study showed that capital costs for such a facility would run into hundreds of thousands of dollars, while resulting savings would amount to a few cents per part.

Fortunately, there was the other option of using plastic motor housings. Du Pont agreed to produce plastic prototypes of the housing in Zytel nylon 82G plastic. Prototypes were quickly assembled; then they passed demanding drop tests and other field tests that are standard for Milwaukee Electric tools. When the housings of impact-resistant Zytel passed the tests with no problem, the firm had a new, lower-cost solution to its machining problem: a plastic housing, produced from a production mold that required no machining.

Redesign presented several additional opportunities. Initial target was to replace aluminum die-casting, and thereby eliminate machining as well as deflashing, trimming, and spadoning (a surface treatment that imparts a matte finish). But they also wanted to eliminate as many parts as possible, simplify the assembly, and use a product that worked as well or better than aluminum. Achieving these goals produced some spectacular benefits; parts costs dropped by two-thirds while manufacturing throughput rates increased. Savings in labor, machining, and assembly operations were augmented by lower capital and maintenance costs. As many as one million plastic housings were injection molded without major tool repair or replacement, vs. 100,000 parts for the die-casting operation.

Six parts in the housing were eliminated. Because plastic used is not conductive, designers were able to do away with insulating parts, such as a coil shield that separated the electric brush holder area from the aluminum, and the cardboard insulating sleeve that went between the copper wiring of the field core and the housing. Removing the sleeve had the added benefit of creating better airflow inside the housing, so the motor ran cooler under load. Press-fitting a rear ball bearing into the housing and keeping the bearing securely in place proved to be a major obstacle. The solution was to use eight small ribs inside the rear-bearing pocket. The ribs increase the amount of interference that can be overcome when press-fitting the ball bearing, and keep the bearing in its pocket with a strong, uniform force.

Another concern was achieving overall perpendicularity of the housing face where it fitted with a mating gear case. A molder solved this problem by repeatedly adjusting molded housing dimensions by a few thousands of an inch. Key to this fine-tuning was to establish three adjustment spots; one at each screw hole location. Thus it was much easier to design mating parts so that they sat on the lands, specific points, rather than trying to align a complete surface. Accurately repeating such minute dimensions required batch-to-batch plastic consistency and process control.

This example has been repeated many times over in many electrical devices. Many different electric appliance devices have been designed and put in long production runs worldwide.

11.2.27 Furniture

Plastic office furniture, principally just for chairs, panels, and laminates for desks, represents about a $10 billion market or just 3½ wt% of all plastics. Unreinforced and glass reinforced PPs are principally used. Lawn and garden furniture consumes about 2½ wt% plastics also principally using PPs. All these products as well as others have a good growth rate.

An exercise follows in designing and fabricated furniture racks. Their function is to serve as a support that can hold several objects in a desired location for storage or for display. Unless a more specific function is defined, the rack can take on a wide range of possible shapes, structures, and materials subjected to very little damaging loads; racks can fall, be twisted during relocations, and so on. One type of rack is a bookshelf compatible with the environment found in a library, suitable for

holding five or more books per foot of shelf in a certain location, and for constant reference use rather than for storage.

In setting requirements for this type rack factors to consider includes defining the environment, set load levels and type of loading situation, and with the library environment provide some idea of the possible aesthetics of the rack. A wide range of design choices remain such as to the material to be used, structure, and shape but they would be limited to those normally used in a library environment. The more accurately and completely the function is defined, the more restricted are the design possibilities and the more detailed the specification requirements become.

This product has to fit its function within the confines of the space in which it is used that could involve how many books it will hold or by stating the size of the supporting rack that will be used. The size can then be decided either by burden or by space restrictions. There will be those where one or the other of these considerations will apply. Also both may apply.

The width of the shelf may be determined by the width of a book. In this example it ranges from 6 to 11 in. (15–28 cm). The typical bookshelf is supported at 3 ft (0.9 m) intervals so that the shelf would be sufficiently wide to fit a typical book rack in the library. The shelf will hold about five books per foot with each having an average weight of 2 lb (0.91 kg). Maximum load on the shelf becomes 30 lb (13.6 kg). If the shelf were completely filled, it could be considered a distributed load, or it can be considered as a set of discreet loads.

The type of shelf design is another consideration. The shelf can be a solid plate of plastic material, an inverted pan-like structure with reinforcing ribs, a sandwich-type structure with two skins and an expanded core, or even a lattice type sheet that has a series of openings. The choice is usually dictated by a number of factors. An important one is appearance or aesthetics.

The design includes developing its appearance that is to match its performance. There are lattice-type shelf, hanging solid structures from the ceiling, etc. They may all be functionally but may not look appropriate for a bookshelf in a library. A simple plastic beam that will function adequately in terms of strength and stiffness may be rather thin. A shelf of this type can look weak even if it is more than functional. However, the solid plate approach results in an uneconomical use of material.

One of the requirements included that it look like a wood shelf to match the atmosphere of the library. To produce the desired thickness appearance using plastics either a lipped edge with internal reinforcement can be used or a sandwich-type structure (Figure 11.2.10). In either case the displacement of the material from the plane of bending will improve the stiffness efficiency of the product.

In the materials selection properties to be considered or analyzed are static tensile and compressive strength, stiffness (modulus), impact properties, temperature resistance, expansion during temperature changes, and even dynamic properties. In this study the only major concern is the long-term stiffness since this is a statically loaded product with minimum heat and environmental exposure. While some degree of impact strength is desirable to take occasional abuse, it is not really subjected to any significant impact. Using several materials such as PP, glass-filled PS, and PS molded structural foam sandwich type panel material, the design procedure follows to determine the deflection and stress limitations of the material in the several designs.

From a load carrying requirement the rack is to meet a maximum deflection and relate it to the cost of fabrication (material, labor, amortization costs of equipment including tools, and the usual plant costs of overhead, etc.). If the designer does not have this type experience it is best done with the assistance of a reliable fabricator who potentially will make the product.

The various designs and costs can be tabulated and the ones that are the most economical can be determined. It may become evident that the design life can be long and the cost of increasing the design life small, or, alternatively, it may be that the cost of small increments in the design life are quite costly. In the latter case, the design life should be limited to the acceptable minimum.

A value judgment can be made as to the product quality requirements and the final design made to meet requirements. It may included reinforced areas, coating or plating, inserts, and so on. With the product dimension a performance evaluation can be conducted. The colored plastic and fabricating process is selected.

In this evaluation what is being presented appears unwanted complications that are not needed for such a simple design. However when going into a complicated design one can apply this basic approach to the more complicated designs. Thus, the next move is to produce prototypes to ensure that the product will meet performance requirements. Prototypes made by machining or other simplified model making techniques do not have the same properties as the product made by molding or extrusion or whatever process is to be used. Simplified prototype procedures may reduce trial mold cost and produce adequate test data in some cases. Main value in this study is appearance and feel to determine whether the aesthetics are correct. Any testing has to be done with considerable reservation and caution.

Following testing of the prototypes for appearance and structural integrity field is made to ensure all is well. Locations and individuals selected require closely monitored or controlled environment. Perhaps

a requirement was overlooked such as learning that the library shelves are often cleaned with a solution that disintegrates the plastic being used requiring a new analysis. Unless there has been a serious misjudgment by the product development designers, the field tests do not lead to redesign. The results of the field-testing can help define the basis for labeling and the instructions to be used with the product so that it can properly operate in service.

11.2.28 Water filter

Worldwide everybody desires pure drinking water. Pure water is obtainable and literally all around us if we want to convert the ocean waters with their dissolved salts. Areas such as the Middle East, where fresh water is very scarce, ocean salt water has been filtered by different techniques for the past many decades. PVC and/or RP piping is used to direct the flow of the salt and filter water.

Todate conventional converting systems are wasteful energywise and costwise. Plastic membranes are being used in systems that could go into large-scale water recovery systems from the sea. There is the process called reverse osmosis (RO) where water is separated from a concentrated solution of a salt by a semi-permeable membrane. As an example the Du Pont Permasep permeators (Figures 11.2.58 and 11.2.59) has pressure that drives the pure water into the solution for dilution. The driving force is the concentration gradient and it is in the form of a pressure that is related to the difference in the vapor pressure of the water and the vapor pressure of the solution at the temperature at which the process takes place. By applying a pressure greater than the osmotic pressure to a solution, the direction of flow of the water is reversed and pure water is removed from the solution.

These permeators come in different product types that are used according to the type of water to be treated. There is seawater and brackish water. Brackish water is lower in salinity than normal salt water and higher in salinity than fresh water, ranging from 0.5 to 30 parts salt per 1000 parts water located between seawater and fresh water.

11.2.29 Lumber

Different recyclable scrap plastics (PE, PP, etc.) can be made in what is usually called plastic lumber that is take off from wood lumber. To increase their mechanical properties short glass fibers (as low as 3 wt%) are added to the recycled plastic resulting in doubling strength. Most are extruded into boards and other shapes. Another process used is injection molding. Fabricated products compete with wood lumber on land and particularly in the water (fresh, brackish water, and salt water). Uses includes bridge decks, boat docks, pallets, outdoor furniture, and so on. Compared to wood advantages exist such as being relatively maintenance-free for at least half a century, as opposed to 15 years for treated wood and 5 years for untreated wood. Todate the major disadvantage has been cost. Extensive use is made in applying plastics in wood to improve their structural and decorative properties (compreg, etc.).

Plastic lumber are now available in stores such as Home Depot Inc., USA Plastic Lumber Co. (Boca Raton, Fl), Loewes, and Menard who are large home improvement retailers.

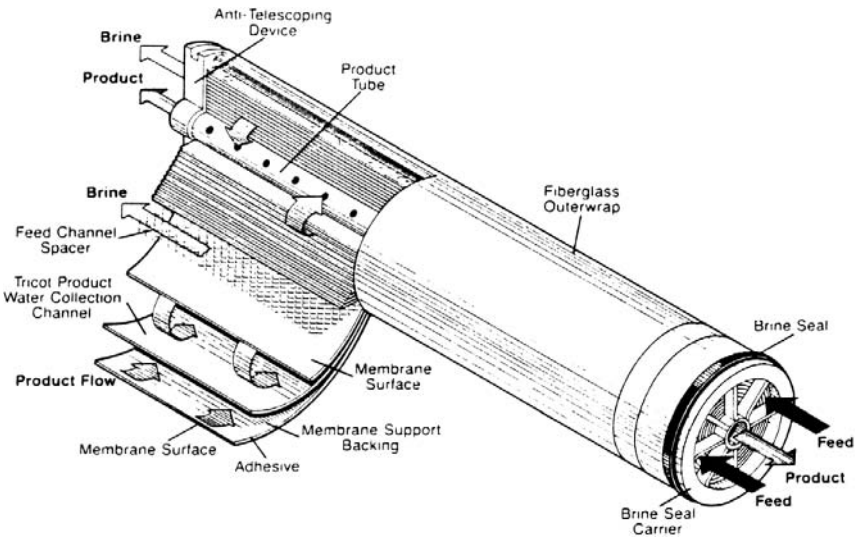


Fig. 11.2.58 Water filter through system. (Courtesy of Du Pont.)

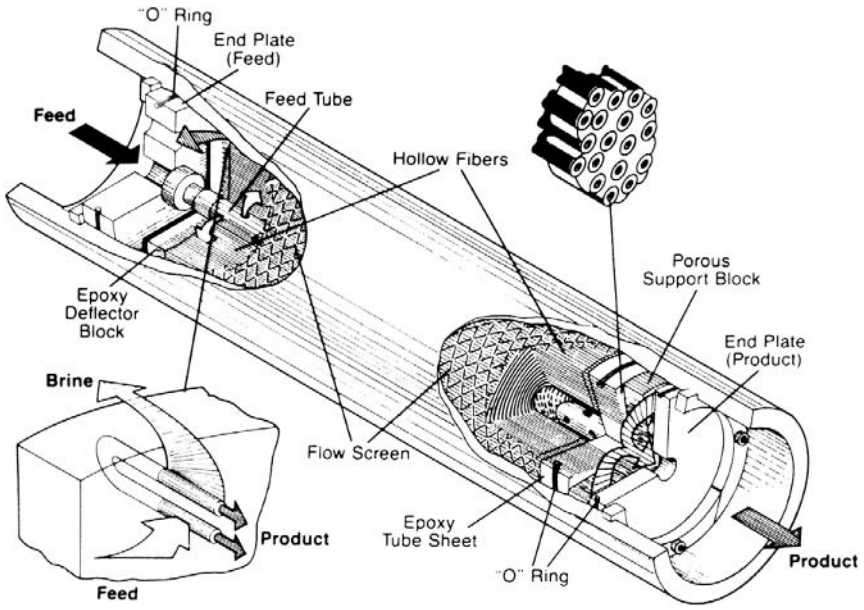


Fig. 11.2.59 Water filter deflector block system. (Courtesy of Du Pont.)

Throughout the years there have been used different forms and/or terms of plastic lumber that usually provide significant increases in physical, mechanical, chemical resistant, long life, and other properties. Compressed wood (also called densified wood or laminated wood) is wood that has been subjected to high compression pressure to increase its density with or without plastics. Composition boards are small particles of wood bonded together that usually use a plastic to bond them. Laminated wood is a high pressure bonded wood product composed of layers of wood with plastic such as phenolic as the laminating agent. Modified wood is any wood where its properties have been improved by plastics. Compreg (also called impregnated wood-plastic) is a contraction of compressed plastic impregnated wood.

11.2.30 Metal

The mechanical behavior of metals in service can often be assumed to be that of a linear, isotropic, and elastic solid. Thus, design analysis can be based on classical strength of materials theory extensively reviewed in textbooks and literature. Such uses are most appropriate to components of simple geometric shapes for which standard solutions exist, or for more complex shapes where they can possibly be used for initial approximate design calculations.

For the more complex, and shapes that do not exist, the solution of the applicable elasticity equations may require some form of numerical procedure, such as FEA or finite difference analysis (FDA). If design analysis involves frequent consideration of similar problems, then the burden on the designer can be reduced by generating a set of solutions presented as a set of design charts.

An alternative is to provide a method in the form of a computer program for which die design analyst requires minimal familiarity with the design method. However, in critical situations, there may be no alternative to a detailed FEA with evaluation of the manufactured product to ensure meeting performance requirements.

Under appropriate conditions, metal design involves plasticity, creep, and geometric nonlinearity. These topics are treated in standard texts and have been put into computer software. However, such complexities are necessarily modeled in a simple technical format.

11.2.30.1 Metal replacement with plastic

This review shows how to design a plastic to replace a metal in such a way as to optimize weight and cost reductions. The plastic used is a stampable reinforced thermoplastic sheet (STX) material (STX is a registered tradename). It is composed of a glass fiber-nylon RTP. Replacing metal with plastic while maintaining material efficiency and achieving cost reduction requires a degree of ingenuity and practicality in design. Direct replacement of 21 gauge (0.032 in.) sheet steel having a modulus of 30×10^6 psi would require that an STX material with a modulus of 1×10^6 psi (1/30th that of steel) have a thickness of 0.099 in. (three times that of steel). These figures are derived as follows:

$$E_1 l_1 = E_2 l_2 \tag{11.2.54}$$

where E is the flexural modulus and l the moment of inertia,

$$E_1 = 30,000,000 \qquad E_2 = 1,000,000$$

$$l_1 = \frac{bh^3}{12} = 2.7 \times 10^{-6} b, \qquad l_2 = \frac{h^3}{12} b$$

$$(30,000,000)(2.7 \times 10^{-6})b = 1,000,000(h^3/12)b$$

$$81.92 = (h^3/12) \times 10^6$$

$$h^3 = 9.83 \times 10^{-4}$$

$$h = 0.099$$

While this increase in thickness may have some disadvantages (space limitations and cost among them), plastics can still be weight and energy effective. To achieve comparable deflection, a plastic beam would require three times the height of a steel beam, but importantly, the weight of the plastic would be only 62% of the metal.

To use material effectively and, at the same time, overcome the space utilization of packaging problem that probably would be encountered with a large section beam, the designers can resort to the extremely simple addition of ribbing to a plastic part. What would be a very complex operation in sheet metal is an attractive option in plastic because the fabrication method permits incorporation of ribs in the molding process. Ribbing increases part section modulus with minimum weight increase. In most cases, ribbing can be very simply and easily incorporated into an STX or other RP constructions part with minimal weight addition and without a penalty in molding cost.

This principle of ribbing can be applied to other mechanical strengthening and stiffening requirements of an STX component. For example, to replace a normal sheet metal flange, a plastic flange would be designed, incorporating several modifications that are easy to make in plastic. The flange is thickened and an outer, down or upturned lip is added, increasing the moment of inertia and thus limiting torque type racking; in addition, gussets are added to distribute the bolt loading. Corrugations in the wall are valuable to provide stiffness to reduce vibration noises or tin canning, and in flange design, to increase the effectiveness of bolt loading by moving the wall to the bolt centerline, thereby utilizing the wall section's stiffness to distribute the bolt loading between bolt holes.

In contrast to sheet metal, where the thickness is uniform, STX can be molded with sections of varying thickness. While it is always recommended in design of plastic parts that section thicknesses be uniform, it is always the first rule to be violated. It is this, very design flexibility that becomes advantageous, and designers are usually unable to resist the ability to strategically modify section thickness to meet the load or stress requirements. If stress, temperature, creep, impact or any other environmental conditions require a heavier or specific section thickness, it can be done within prevalent design parameters.

11.2.30.2 Performance behavior

All structures under load will deflect in some manner and the resultant moment will create bending and shear stresses. Therefore, it is important that the load analysis be executed correctly. There are several major considerations that must be addressed before any mathematical analysis can be performed. The following conditions directly affect the design stresses that will be obtained by computations.

11.2.30.2.1 Moisture effect

Nylon, the resin base for the first series of STX composites, is a hygroscopic material. Therefore, STX materials exhibit the normal property and dimensional modifications characteristic of nylon as they absorb moisture. Dimensional changes in improperly designed rigidly affixed parts can develop excessive stresses. (Note that the effects of temperature are similar.)

11.2.30.2.2 Long-term vs. short-term loading

This is a very important consideration since all materials will exhibit creep or relaxation deformation: with plastic materials, it occurs at lower temperatures and stress levels. Hookean stresses, which are instantaneous stresses, are shown in Figure 11.2.60 for reference levels only, or for very short-term loadings. However, it is the flexural creep characteristics that govern the usable design stresses, especially when dealing with continuous loading such as is encountered in beams and spring members. The specific curves representing percentage strain vs. time at a given stress of 7000 psi, is shown in Figure 11.2.61.

Once these stresses are allowed by the design, the proper geometry (moment of inertia) and deflection can be obtained from the known equations of mechanics of solids. Since stress and deflections are related to both flexural modulus and moment of inertia, the design can now be optimized. It is recommended to use a value for E (modulus) of about 700,000 psi at room temperature and 50% R.H. For other conditions, review Figures 11.2.62 and 11.2.63. The moment of inertia (I) can be changed substantially by the addition of ribs or gussets or a combination of both.

The effectiveness of gussets is very easily analyzed considering that the deflection of any beam is a function of the unsupported length (L) at the third power. By incorporating a gusset at the midpoint of the span, the unsupported length is reduced by half, therefore, the deflection is now decreased by a factor of 8 $[(L/2)^2]$.

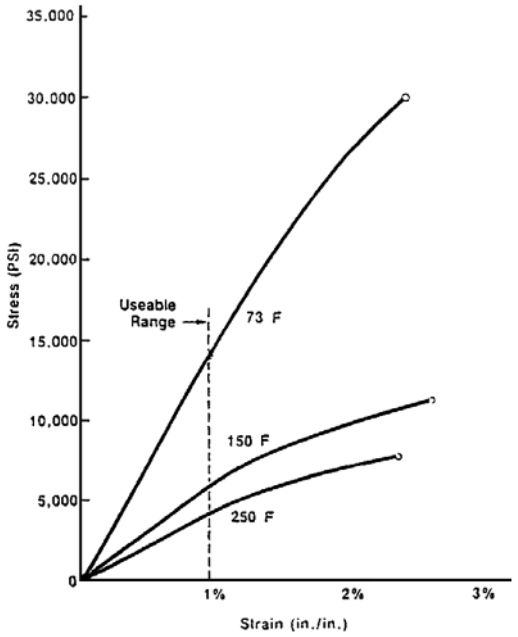


Fig. 11.2.60 Stress–strain relations.

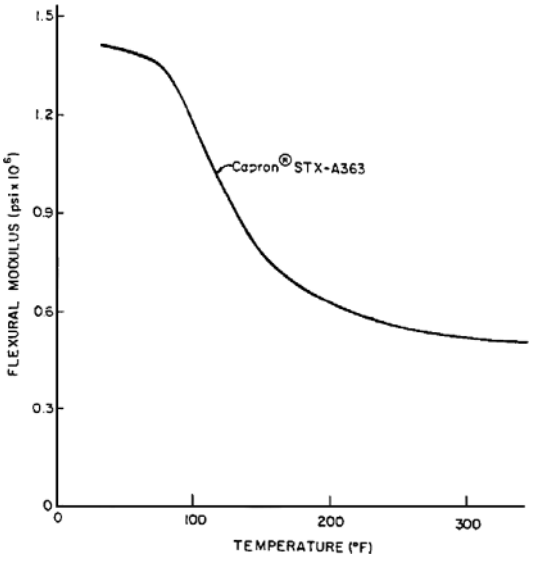


Fig. 11.2.62 Effect of temperature on the flexural modulus.

These simple examples show how both stress and deflection can easily be minimized by taking advantage of the material manufacturing characteristics that allow the forming of ribs and gussets during the molding cycle. Although beam problems were discussed (they are the most commonly used structural members), the same reasoning can be applied to other structural members such as plates or shell sections. Flanges of covers for example, can be handled in that way.

11.2.30.2.3 Stress concentration

The addition of ribs and gussets while reducing deflection can also increase the bending stress if care is not taken in the design stage. Since the ribs or gussets are always in a plane perpendicular to the original surface, the intersection of the planes must comprise a radius to avoid stress concentration. Consequently, it is suggested that a minimum radius of 1/8 in. be incorporated in all designs. With specific reference to ribbing, the shape and size of the ribs must be carefully analyzed to prevent an increase in the bending stress (machine design books review this subject).

11.2.30.2.4 Coefficient of expansion

Since the coefficient of expansion of metals and plastics is different, stresses and deformation can result. A given plastic beam clamped to a steel support at both ends by fasteners, will elongate about three times more than the steel support: α steel $= 6.5 \times 10^6$ in./in./°F, α STX $= 16.2 \times 10^6$ in./in./°F. If the plastic beam is not allowed to

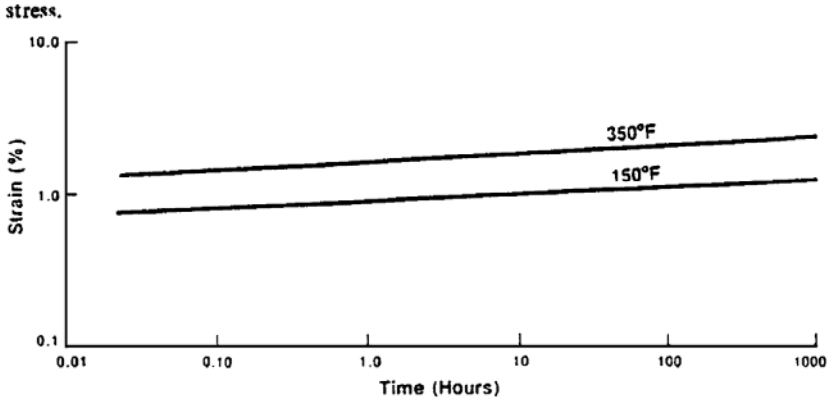


Fig. 11.2.61 Flexural creep resistance at 7000 psi initial stress.

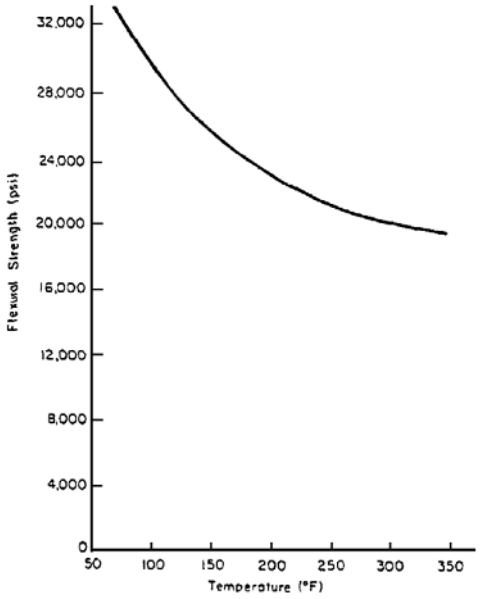


Fig. 11.2.63 Example of temperature effect on flexural properties.

slide under the fasteners then end stresses will occur. These stresses can be computed simply as follows:

$$\Delta L = L \times (\alpha_p - \alpha_s) \times \Delta T \tag{11.2.55}$$

$$\varepsilon = \frac{\Delta L}{L}$$

$$S = \varepsilon \times E_p$$

$$P_{cr} = \frac{4 \times \pi^2 \times E_p \times l}{L^2} \text{ (Euler equation)}$$

where L is the distance between bolts; E the flexural modulus; L and S, expansion stress. The subscripts p and s are plastic and steel, respectively.

Since this stress is directed in the plane parallel to the beam axis, the beam becomes a column fixed at both ends. Applying the standard Euler equation, the critical buckling load (P_{cr}) and corresponding stress can now be computed ($S_{cr} = P_{cr}/A$). The Euler equation comprises several parameters: E, L, I. Of these three, E is fixed, and L (distance between bolts) is usually fixed, (in most cases more bolts means additional costs). I is therefore the only possible variable. Adding one or more ribs will change I and substantially reduce the induced expansion load so that the beam will remain flat and free of distortion since the expansion stress is now much smaller than the critical stress.

11.2.30.2.5 Bolt torque effect

Since there be STX and other plastic components that are bolted to other components, it is important to consider the effects of the bolt torque, namely the compressive and shear stresses under the head of the bolt. To retain the maximum torque, the following parameters must be considered: (1) STX material, when loaded will tend to relax (long-term effects) and (2) based on the fastener requirements, there is a minimum torque suggested to stretch the bolt (bolt preload).

When the torque drops below this minimum value, the fastener tends to loosen. If the STX part is a vessel for containment of fluids, leakage could result at the flanges even with a gasket. The load due to the torque is computed by

$$F = T/(0.2 \times \text{dia. of bolt}) \tag{11.2.56}$$

where T is the torque.

This load (F) is then used to calculate the compressive and shear stress as shown by the following equations, respectively:

$$S_c = \frac{F}{\text{net } A \text{ washer}} \tag{11.2.57}$$

$$S_s = \frac{F}{\pi \times t \times \text{O.D. washer}} \tag{11.2.58}$$

where t is the thickness of material; and $A = 0.785$ (O.D.^{2}) – (I.D.^{2}).

These equations show that the use of a flat washer to distribute the load and reduce the respective stresses is recommended (permitted S_c design = 7000 psi, and permitted S_s design = 6000 psi). To illustrate the torque and relaxation relationship, see Figure 11.2.64. The ordinate represents the ratio of original stress to the actual or time-dependent stress and the abscissa is the time cycle. The relaxation curve shows that there is no difference in relaxation between stress levels of 5000 psi and 15,000 psi and that most of the relaxation takes place during the first twenty hours.

Incorporating a gasket between the two surfaces changes the shape of these curves depending upon the thickness and hardness of the gasket material. The analysis is now further complicated and must be handled by the theory of beams on elastic foundations to obtain moment distribution and deflection. Good results can be obtained by using an elastomer such as RTV as gasketing material. The effects of gasket behavior are shown in Figure 11.2.65.

11.2.30.2.6 Impact barrier

The glass mat at the core of the STX sheet was shown to impart some very unusual impact properties to the material. The testing performed in the laboratory (Gardner drop test) and the field results from production components have shown that, in contrast to glass reinforced plastics, the failure mode is elastic rather than brittle. The resultant damage is a hole with little or no crack

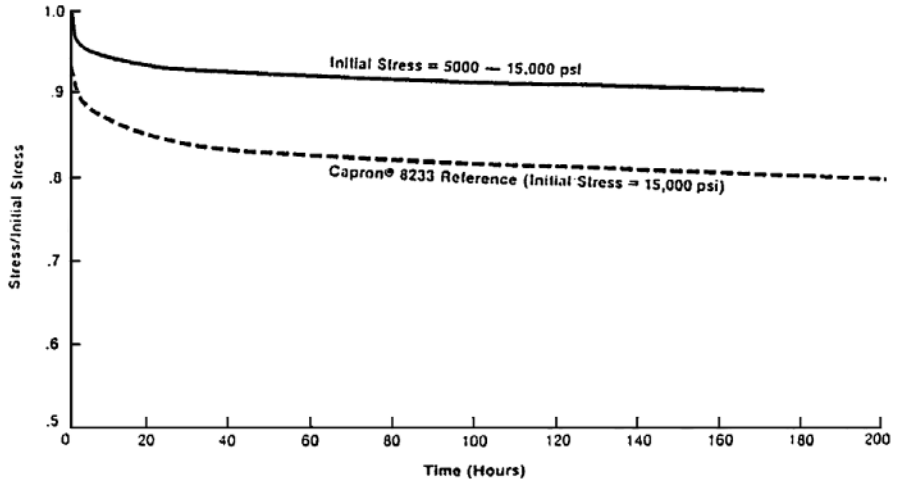


Fig. 11.2.64 Compressive relaxation at 73 °F (0.2655 in.^{2} compressed area).

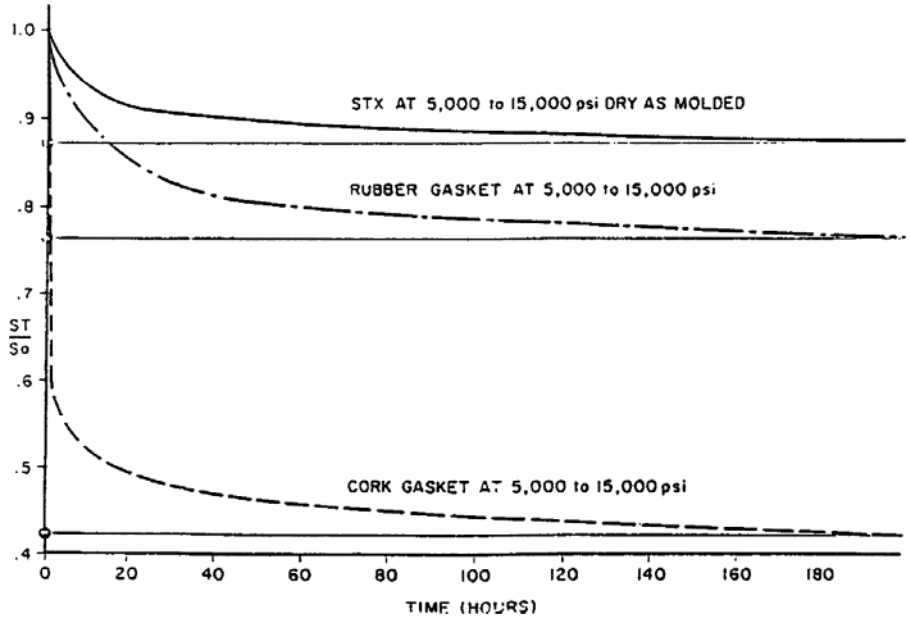


Fig. 11.2.65 Compressive relaxation at 73 °F between STX without and with gasket.

propagation around the impact area whereas the other reinforced materials will shatter under similar conditions. The STX failure mode applies not only at room temperature, but is equally valid at −40 °F. STX impact resistance increases with the thickness of the material, i.e., ¼ in. thick plate will resist much higher impact loads than ⅛ in. thick plate.

However, there is a better solution that result both in lower material costs and lower weight. By keeping the impact area to a minimum thickness and adding ribs, strength is dramatically improved. As an example of this approach, a 0.10 in. thick plate was pierced when impacted by a ½ in. steel ball (1.85×10^{-2} lb) travelling at 164 mph (198 in-lb of K.E.). By adding small ribs, 0.10 in. wide × 0.10 in. high spaced 0.5 in. apart (the plate thickness remained at 0.10 in.), the steel ball had to travel at 272 mph (552 in-lb of K.E.) to duplicate the break resulting in a substantial improvement.

11.2.30.2.7 Vehicle oil pan

The oil pan is a critical component in auto and truck applications because any failure would be catastrophic. It must withstand high temperature and impact, keep the bolt torque intact for good sealing, resist vibration and abuse, especially during assembly, and in many instances, functions as a structural member. The pan is divided into basic elements and analyzed as follows: (1) sides are

considered to be flat plates uniformly loaded with rigid supports, (2) flanges are analyzed as beams on elastic foundations because of the gasket/metal interfacing, and (3) bottom section is considered as a plate but needs to withstand high impact loads (dynamic loading).

After the bolt torque is established, the washer size must be determined by computation based upon compressive and shear stresses and flange deflection. The flange thickness will also be determined from calculations. After it has been established that these values are safe, the remaining sections such as the sides and bottom can now be designed for thickness and shape based on the values of the static and/or dynamic loads (impact) supplied as part of the input data. If some elements show either high stress or excessive deflection (1% elongation in the elastic range) ribs or gussets can be added and/or the wall thickness increased in selective areas only. History has shown that theoretical analysis yields a design very close to the final manufactured product and meets intended performance requirements dictated by laboratory as well as field testing.

11.2.30.2.8 Attachment

There will be cases where attachments become necessary. One of the most frequent types is threaded parts. In most instances, these fasteners are used repeatedly to attach or remove components. To ensure that the reliability and the life of the frequently used joint are maximized, a threaded insert is used rather than threading the plastic material. In this manner, the wall thickness of the insert and plastic part is reduced, or, if a boss is used, it is smaller in diameter and shorter in height compared to its plastic counterpart. Since E_{metal} is much larger than $E_{plastic}$ higher torque can be applied (shear stress is also much larger).

Extremely close tolerances such as might be necessary for a seal or a bearing may not be within the capabilities of the STX molding process. Machining is not recommended because it would break the nylon surface and expose the glass fibers that then act as wicks for fluid. An aluminum insert that has been finish machined is used as a substitute. Since the ratios of the moduli are quite large (about 20:1 for steel and 10:1 for aluminum), there will be no deformation of the metal insert and only the plastic will be highly stressed. It is most important that these stresses are calculated so that the boss does not split during assembly or that the metallic ring does not become loose because of long-term relaxation effects.

11.2.31 Design limitation and constraint

Designing acceptable products requires knowledge of the different plastics and their processing limitations such as individual advantages and disadvantages. Although there is no limit theoretically to the shapes that can be created, practical considerations must be met such as available and size of processing equipment and cost. These relate not only to the product design, but also the mold or die design, since they must be considered as one entity in the total creation of a usable, economically feasible product.

One of the earliest steps in product design is to establish the configuration that will form the basis on which strength calculations will be made and a suitable material selected to meet the anticipated requirements. During the sketching and drawing phase of working with shapes and cross sections there are certain design features with plastics that have to be kept in mind to obtain the best cost-performances and avoid degradation of the properties. Such features may be called property detractors or constraints. Most of them are responsible for the unwanted internal stresses that can reduce the available stress level for load-bearing products. Other features may be classified as precautionary measures that may influence the favorable performance of a product if they are properly incorporated.

As an example a weld line(s) can exist in a product that could have met design requirements if the weld line(s) did not develop. The designer did not contemplate the potential for weld line(s). However the person designing the mold took a logical approach to simplify its construction and reduce cycle time to mold the product based on the requirements specified for the product. Result was in reducing the cost of the mold and fabrication time. To meet this objective the design of the mold caused one or more weld lines to develop in the product. With conventional injection molding, molded products can be designed that create unwanted weld line(s). The so-called line forms when two melt flow fronts meet during the filling of an injection mold cavity. This action can also occur during extrusion through a die, etc. Depending on how the weld line forms it could have very little strength and under the most ideal molding conditions it may obtain up to possibly 85% strength retention. To eliminate any problem the product requirements have to account for loss in properties if weld line(s) occur or specie that no weld lines are to occur.

Weld lines are also called knit lines. During processing, such as by injection molding and extrusion, weld lines can occur. They can form during molding when hot melts meet in a cavity because of flow patterns caused by the cavity configuration or when there are two or more gates. With extrusion dies, such as those with "spiders" that hold a center metal core, as in certain pipe dies, the hot melt that is separated momentarily produces a weld line in the direction of the extrudate and machine direction. The results of these weld lines could be a poor bond at the weld lines, dimensional changes, aesthetic damages, a reduction of mechanical properties, and other such conditions.

The top set has a single gate for each specimen, the center set has double gates that are opposite to each other for each specimen, and the bottom set has fan gates on the side of each specimen. The highest mechanical properties come with the top set of specimens, because of its melt orientation being in the most beneficial direction. The bottom set of specimens, with its flow direction being limited insofar as the test method is concerned, results in lower test data performance. With the double-gated specimens (the center set) weld lines develop in the critical testing area that usually results in this set's having the potential lowest performance of any of the specimens in this diagram.

Fabricating techniques can be used to reduce this problem in a product. However, the approach used in designing the product, particularly its mold (relocate gates), is most important to eliminate unwanted orientation or weld lines. This approach is no different from that of designing with other materials like steel, aluminum, or glass.

With moldings that include openings (holes), problems can develop. In the process of filling a cavity the flowing melt is obstructed by the core, splits its stream, and surrounds the core. The split stream then reunites and continues flowing until the cavity is filed. The rejoining of the split streams forms a weld line. It lacks the strength properties that exist in an area without a weld line because the flowing material tends to wipe air, moisture, and/or lubricant into the area where the joining of the stream takes place and introduces foreign substances into the welding surface. Furthermore, since the plastic material has lost some of its heat, the temperature for self-welding is not conducive to the most favorable results.

A surface that is to be subjected to load bearing should be targeted not to contain weld lines. If this is not possible, the allowable working stress should be reduced by at least 15%. Under the ideal molding conditions up to about 85% of available strength in the solidified plastic can be developed. At the other extreme where poor process controls exist the weld line could approach zero strength. In fact the two melt fronts could just meet and not blend so that there is relatively a microscopic space. Other problems occur such as influencing aesthetics.

Prior to designing a product, the designer should understand such basic factors as those reviewed in this book. Recognize that success with plastics, or any other material for that matter, is directly related to observing design details.

The important factors to consider in designing can be categorized as follows: shape, part thickness, tolerances, ribs, bosses and studs, radii and fillets, drafts or tapers, holes, threads, colors, surface finishes and gloss levels, decorating operations, parting lines, shrinkages, assembly techniques, production volumes, mold or die designs, tooling and other equipment amortization periods, as well as the plastic and process selections. The order that these factors follow can vary, depending on the product to be designed and the designer's familiarity with particular materials and processes.

Chapter 11.3

11.3

Material and shape selection

Nigel Mills

11.3.1 Introduction

This chapter, covering the materials and shape selection aspects of design, builds on some simple exercises. Ulrich and Eppinger (Further Reading) take a more general approach, showing how materials engineers interact with other disciplines in the design process. Polymer selection packages will be explored, before considering component shapes that maximise bending and torsion stiffness. These component shapes are influenced by the process used for manufacture, and the need to optimise its output. The reader is encouraged to observe the details of a variety of successful products, and to recognise why certain designs can be manufactured efficiently.

11.3.2 Polymer selection

11.3.2.1 Polymer selection packages

Materials selection packages have partly replaced tables of data for particular grades of plastics. They provide basic information. More help is available from the manufacturer's website or technical service department. CAMPUS, freely available from www.campusplastics.com, allows comparison of grades from a single manufacturer. An augmented version, available from www.mbase.de for an annual fee, allows data comparison between different companies. The properties are classified into types: rheological, mechanical, thermal, electrical, and other processing and additives. Each menu contains many items; for instance, *mechanical* contains tensile modulus, charpy impact strength,… with a single value for each property. There are also multi-point data, in graphical form, such as

(a) shear modulus versus temperature;
(b) tensile stress–strain curves at a range of temperatures;
(c) shear viscosity versus strain rate at a range of temperatures;
(d) creep strain versus time at a range of creep stresses.

Not all manufacturers provide the same range of data, so the information for some grades of plastics may be incomplete. The selection process *ranks* polymer grades on the basis of one or more properties. Minimum and maximum values should be specified for the properties felt to be appropriate to the application. The programme then finds how many grades meet these requirements. It is unhelpful if the answer is hundreds of grades, or none. If the latter is the answer, one or more selection conditions must be relaxed. The number of grades manufactured is limited to reduce inventories. Thus, the level of glass fibre reinforcement or rubber toughening is limited to two or three levels. Consequently, there is unlikely to be a perfect match to a set of target specifications.

Single-value mechanical parameters lead to an unambiguous ranking of the polymer grades. However, this can mislead the unwary user into ignoring the time and temperature dependence of properties. A single modulus value relates to a specific timescale, strain rate and test temperature, and to the processing of the test specimen. Injection-moulded tensile bars are usually end-gated, therefore, polymer orientation along the length of the bar gives optimum values of the strength and modulus. Tests on a thickness, which is the same as that of the intended product. However, the data is for standard 3.2 mm (1/8

Plastics: Microstructure and Engineering Applications; ISBN: 9780750651486
Copyright © 2005 Elsevier Ltd. All rights of reproduction, in any form, reserved.

in.) thick bars. The lack of fracture mechanics data in the databases is another drawback. Such data does not appear to be part of the routine industrial characterisation of plastics.

It would be useful to sort grades according to price, so the cheapest grade that meets the requirements heads the list. However, the prices of plastics fluctuate with the price of oil and CAMPUS does not give prices. It indicates whether a grade is suitable for a particular process, but the database neither indicates the cheapest process route, nor gives the cost of the manufactured product.

Polar charts allow the comparison of polymers for multiple selection criteria. The properties for a particular grade are plotted on *n* scales that radiate from a central origin, and the points joined to form a closed polygon. Figure 11.3.1 compares five properties of a grade of polyamide 66, polyamide 612 and polyoxymethylene. There are different rankings for water uptake compared with notched impact strength. If it is possible to specify the minimum property values required for the product, a target polygon can also be drawn; only grades whose polygons enclose this are acceptable.

Plastic products should be easy to identify for recycling—hence, the marking of car components with polymer abbreviations such as 'PP'. Products may be designed for ease of dismantling and separation into their component polymers at the end of their lives. This tends to reduce the number of plastics used, while discouraging the use of composite structures. A PBTP bumper skin, filled with polyurethane foam, attached to a steel subframe is more difficult to recycle than a bumper with a PP skin, PP foam core and a glass-reinforced PP mounting beam at the back. Recycling constraints are not at present part of selection packages.

11.3.2.2 Property combinations for materials selection

Some products or parts of products, such as rectangular cross section beams, are frequently loaded in bending or torsion. If the product must meet a mechanical property target, and its mass must be minimised, property combinations can be used to explain why some materials are feasible and others impractical. It is unlikely that the plastic was initially chosen on the basis of property combinations. It is more likely that a plastic from a similar product was used, with low cost being the most important selection parameter.

Two examples are used to explain where the use of plastics is appropriate in car bodies, and where metals are more appropriate. Comparisons between plastics and wood are difficult, since different processing methods apply, and certain designs are optimum for each material. A design optimised for injection moulding in plastic (Sections 11.3.5 and 11.3.6), is neither suitable for anisostropic materials like wood, nor easy to construct.

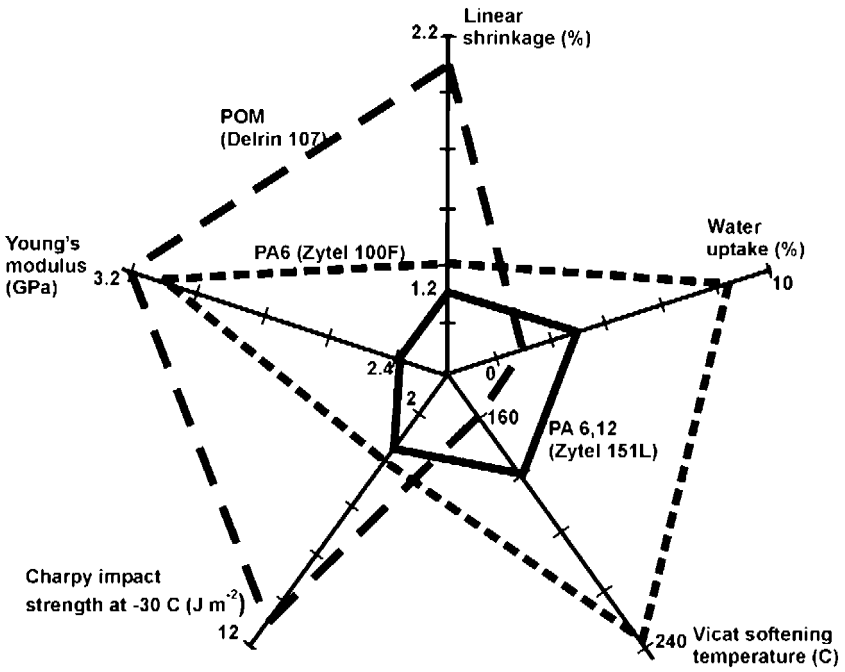


Fig. 11.3.1 Comparison of five properties of un-reinforced grades of polyamide 66, polyamide 612 and polyoxymethylene, on a polar plot (redrawn from Du Pont data from CAMPUS).

11.3.2.3 The nearly flat skin of a car door

A car door skin is loaded in bending when pushed shut with the hand. If it deflects too much, it appears flimsy. To simplify the analysis, the panel deflection is calculated as for a constant cross section beam, with the strain varying linearly through the thickness. In reality, the curvature and styling ridges on metal door panels play an important stiffening role. For a flat panel of width w (treated as a constant) and thickness t, the bending stiffness is

$$\text{MR} = \text{EI} = \frac{Ewt^3}{12} \tag{11.3.1}$$

The skin panel must have a certain bending stiffness; the external loads fix M, while the deflection limit sets a minimum limit for R. The minimum MR value means that beam bending stiffness EI must exceed a minimum value. If the material, hence E, is varied while EI is kept constant, the necessary skin thickness t depends on $E^{-1/3}$. The panel must cover the door, so has a constant width. Hence, its mass m is directly proportional to the material density ρ and to t, so

$$m \propto \rho t \propto \frac{\rho}{E^{1/3}} \tag{11.3.2}$$

To minimise the panel mass, the combination of properties $E^{1/3}/\rho$ must be maximised. Hence, $E^{1/3}/\rho$ is the *selection parameter* for the panel bending stiffness requirement. The data in Table 11.3.1, for two types of fibre-reinforced plastics, polycarbonate and two metals, shows that sheet moulding compound (SMC) and aluminium are the best materials. However, the range of $E^{1/3}/\rho$ values is not large, so the material selection may be based on other factors, such as cost of processing into a slightly curved shape or corrosion resistance. A skin panel's resistance to denting (when another door edge hit it in a car park) is proportional to the material's elongation at yield. RRIM has a marked advantage in this respect.

11.3.2.4 The tubular frame of a car body

The bending and torsional stiffness of a car body is mainly provided by the hollow tubes that surround the passenger compartment. These have complex cross sections, but, to simplify the analysis, can be considered as square-sectioned, thin-walled tubes of width w and wall thickness t. Their bending stiffness is given by the product of Young's modulus with

$$I = \frac{1}{12}\left[(w+2t)^4 - w^4\right] \cong \frac{2}{3}w^4 t \tag{11.3.3}$$

The maximum value of w is fixed by the design requirements that the door pillar does not obstruct vision and the sill does not obstruct entry. In a beam of constant mass, t is inversely proportional to the material density. Hence, the selection parameter for bending stiffness is E/ρ; steel and aluminium have high values (Table 11.3.1), so are used in preference to SMC or thermoplastics.

To provide protection for the occupants in a frontal crash, there are two fore-and-aft tubular 'rails', which support the engine, then make an S bend and connect to the door sills. These rails have a crushing resistance of approximately 200 kN, so that the passenger cage of a 1 tonne car decelerates at approximately 20 g. To minimise the mass of these rails, a material with a high ratio of yield stress σ_y to density is required. Table 11.3.1 shows that high strength steel is optimal for this component. For these two reasons, plastics are not used for the main structure of a car body.

The production engineering complexity of using one process technology for car panels and another for the

Table 11.3.1 Selection parameters for car body materials

Material		**Steel**	**Aluminium**	**PC**	**RRIM**	**SMC**
Density ρ(kg m^{-3})		**7800**	**2800**	**1200**	**1200**	**1800**
Young's modulus E (GPa)	**Select**	**207**	**69**	**2.2**	**1**	**12**
Property	**on**					
Flat panel bending stiffness	$\sqrt[3]{E}/\rho$	0.76	**1.46**	1.08	0.83	1.27
Box. beam bending stiffness	E/ρ	**2.65**	2.46	0.18	0.08	0.67
Resistance to denting	e_{yield} (%)	0.15	0.2	6	**10**	1
Crash energy absorption	σ^*/ρ	**225**[a]	223	95	49	101

[a] Value for high-strength steel (best results given in bold).

tubular steel frame have been overcome for some American sports cars, but the majority of motor manufacturers prefer to continue to use steel throughout. The total production costs for sheet metal panels (tooling and materials) are smaller than for plastics for large volume production, but lower for RRIM and SMC plastics for lower production volumes.

11.3.3 Shape selection to optimise stiffness

11.3.3.1 Corrugations

A common way to increase the bending stiffness of products from melt-inflation processes is to use corrugations. It is not possible to create ribbed shapes, described in the next section, because there is no independent control of the shape of the inner surface of the melt. Corrugations are seen on thermoformed products, blow-moulded bottles and extruded pipes.

The diametral deflection of a buried pipe, due to soil loading, can be reduced by increasing the second moment of area I of the pipe wall. The corrugated wall of the soil drainage pipe can be approximated as a trapezoidal wave, of wavelength L, wave amplitude H, crest length C and constant vertical thickness t. Its second moment of area I_C for a longitudinal section is related to that of the plane-walled pipe by

$$\frac{I_C}{I} = 1 + \left(\frac{H}{t}\right)^2 \left(1 + \frac{4L}{C}\right) \qquad (11.3.4)$$

For the shape shown, the stiffening factor is approximately 450. However, the section perpendicular to the pipe length has a low I; the corrugations deform like a bellows when the pipe is bent.

A twin-walled corrugated pipe is shown in Figure 11.1.14. The external wall is formed by applying an internal air pressure to the molten tubular extrudate

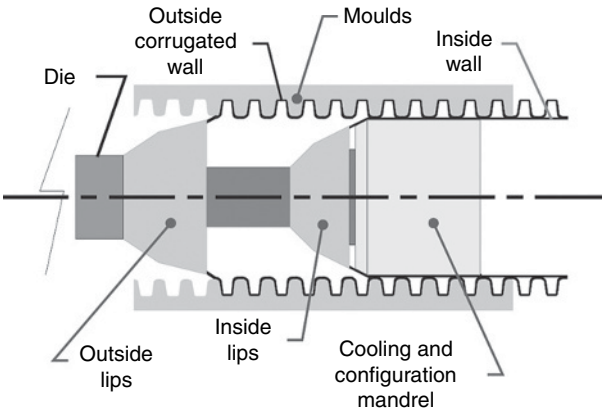


Fig. 11.3.2 Extrusion method of making twin-walled corrugated HDPE pipe (from Diez C, Characteristics of Corrugated HDPE pipe, 7th Ed. 2005, Inc, Quebec).

(Figure 11.3.2), which expands against pairs of mould sections—these move with the cooling pipe for some distance, then return on a caterpillar track to the die.

The second, interior wall is extruded, while the inner surface of outer wall is still molten, creating a twin-walled pipe. The weld line between the two layers is visible in Figure 11.1.14. Such pipes have a high bending stiffness, both longitudinally and for soil loading.

11.3.3.2 Ribs on injection mouldings

The easiest way to stiffen injection mouldings is to use ribs on the hidden surface (Figure 11.3.3a). Ribs correspond to slots in the mould, so straight ribs are easier to machine (using a milling wheel) than those with corners. The rib thickness r should be less than 2/3 of the thickness t of the surface that they support, so they complete solidification first, thereby avoiding sink marks appearing on the product surface.

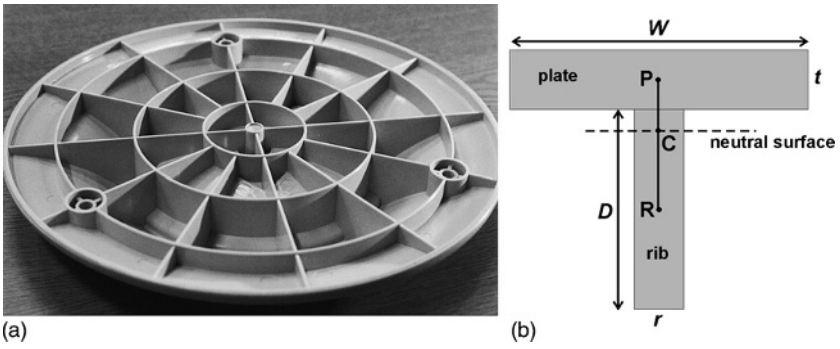


Fig. 11.3.3 (a) Ribbing on the underside of the base of a computer monitor stand, (b) location of the centroid if a section of ribbed plate.

A basic analysis assumes that the neutral surface is planar, so it passes through the centroid or centre of gravity, of the section (Figure 11.3.3b). The section repeats to create a multi-ribbed plate. Each unit of the plate, of area Wt, has a centroid at its centre P, while the rib of area Dr has a centroid at its centre R. The overall centroid is at a position C, where by the lever rule

$$WtRC = DrCR \tag{11.3.5}$$

This calculation neglects the taper angle of 1° or 2°, needed on the rib to allow easy ejection from the mould.

When a ribbed plate is designed, several solutions are possible. Consider the replacement of a solid plate of thickness $s = 5$ mm by a cross-ribbed plate of uniform thickness $t = 2$ mm, having the same bending stiffness (Figure 11.3.4). The chart recommends one of the pairs of parameters (W, t) that meet the bending stiffness requirement. This is not the minimum mass solution. The mass can be reduced further by increasing the rib depth D and spacing W, but such ribs would buckle at a low applied moment. Reading across horizontally (the grey arrow) from $t/s = 2/5$ on the left-hand scale, the number by the curve indicates that the ribbed plate has 0.55 of the solid plate volume, i.e. there is a 45% weight saving. Reading down from this point, to the horizontal scale at the bottom of the chart, $s/W = 0.125$. As $s = 5$ mm, the rib spacing $W = 40$ mm. Finally, reading across from the curve to the right-hand scale, $T/s = 1.95$, so $T = 10$ mm and the rib depth D is 8 mm.

Finite element analysis (FEA) of the design example shows that the stress distribution is more complex (Figure 11.3.5) than that assumed in the simple analysis. The neutral surface is non-planar, dipping towards the midplane of the plate, midway between the ribs. The transverse ribs also have a local effect on the stresses in the

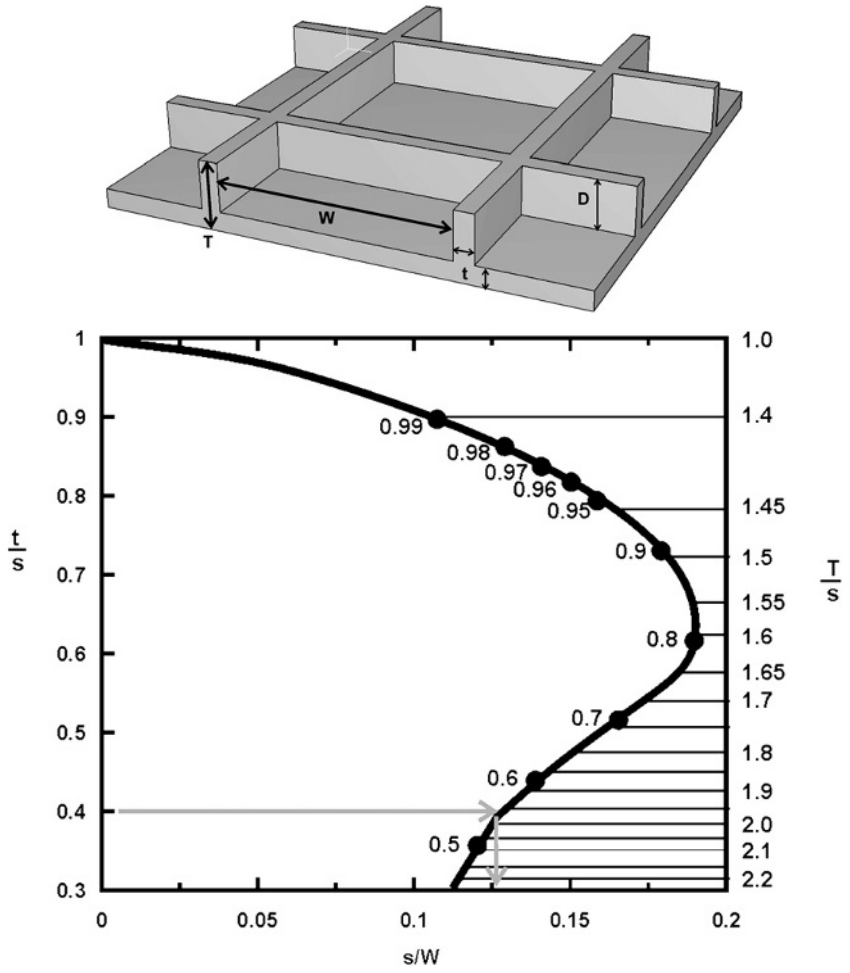


Fig. 11.3.4 (a) Replacement of a flat plate with a cross-ribbed plate of the same bending stiffness. (b) The reduction in mass achieved by the change, and the rib dimensions required (redrawn from *Delrin Design Handbook,* Du Pont, 1980).

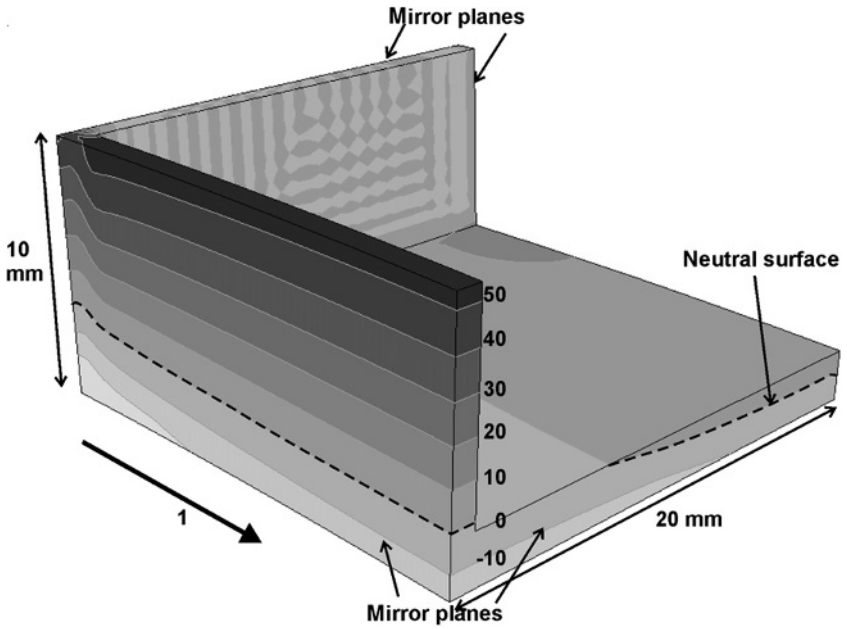


Fig. 11.3.5 FEA of a repeating unit of the recommended ribbed plate design: contours of longitudinal stress (MPa) in the 1 direction for polystyrene, when the mean radius of curvature is 0.4 m.

longitudinal ribs. Consequently, the computed bending stiffness is only 78% of that predicted by Figure 11.3.4.

A cross-ribbed design only provides stiffening in the two rib directions. There is a low bending stiffness along a direction at 45° to the ribs, and a low torsional stiffness. An isotropic stiffened plate needs ribs in at least three directions, so the mass saving will not be quite as high as suggested by Figure 11.3.4. If the direction of the main bending moments is known, ribs can be placed in the appropriate directions. The edges of a product, such as stacking polypropylene chair, are natural places to incorporate ribs.

Figure 11.3.6 shows the relative efficiency of different beam cross sections. Each has the same cross-sectional area and the neutral surface is horizontal. The *I* beam, of height equal to twice its width, is given a second moment of area of 100%. The C beam, while apparently as efficient, will twist when placed under a bending moment. A ribbed plate, with a rib depth equal to 2/3 of the rib spacing, has a relative efficiency of only 9%. Although *I* beams and hollow tubes are efficient, they are difficult to mould in one piece. Therefore, it is common to use ribbed casings for injection-moulded products.

11.3.3.3 Buckling of ribs

The bending moment in service should not cause ribs put into compression, to buckle. As one side of the rib is supported by the plate, only the free side can buckle. FEA is necessary to predict the shape. Figure 11.3.7

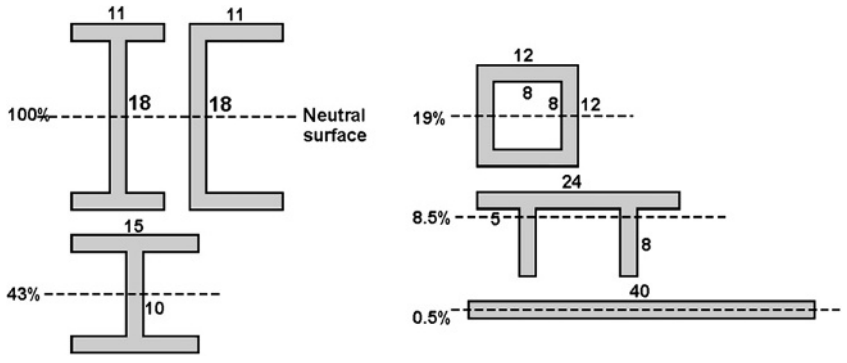


Fig. 11.3.6 Relative bending stiffness of beams, of the same cross-sectional area, about the neutral axes shown (beam dimensions in millimeters, with 2 mm wall thickness).

shows a unit of a cross-ribbed plate, with deeper and longer ribs than those considered in the last section. The free sides of the longitudinal ribs are initially straight, but buckle when the plate curvature becomes large. The bending moment versus curvature relationship becomes non-linear when buckling commences, but the slope remains positive.

If FEA shows that some ribs buckle, the product can be redesigned with thicker or shorter ribs. The latter can be achieved by having cross-ribs at closer intervals. These add to mould construction costs, but add very little to the materials costs or the process cycle time. The productivity of an injection mould for the ribbed plate will be controlled by the solidification time. It will be ~3 s for the ribbed plate and ~19 s for the 5 mm thick original design. The increase in mould productivity is usually more important than the materials saving from product mass reduction.

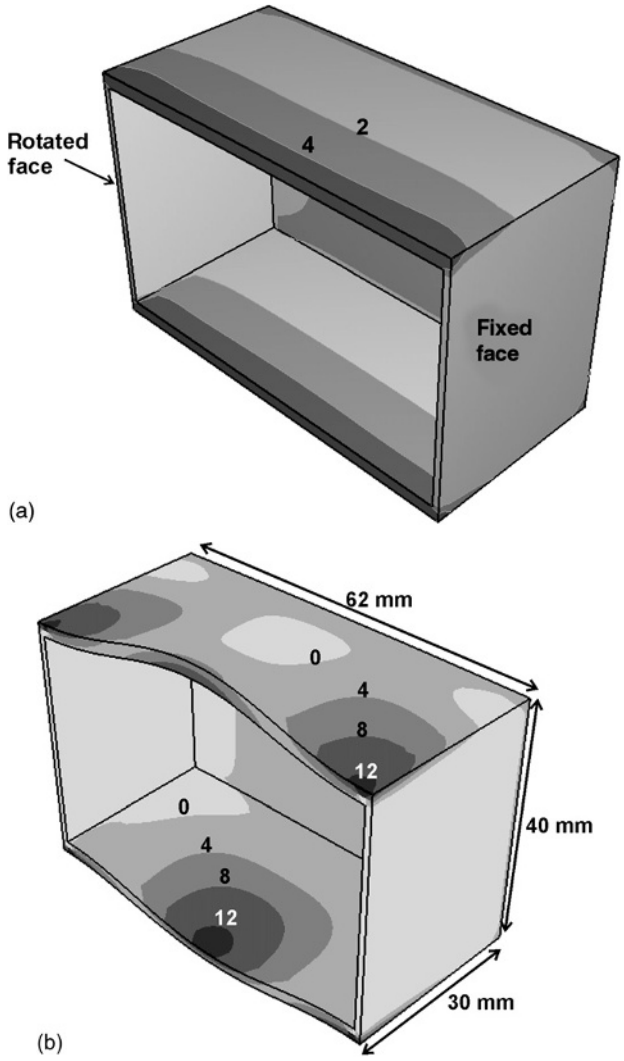


Fig. 11.3.7 FEA of bent section of cross-ribbed PP plate: (a) relatively undeformed when the radius of curvature $R = 4.2$ m, (b) some bulked ribs when $R = 1.24$ m. The rib and plate thickness is 2 mm. Contours of von Mises stress (MPa).

11.3.3.4 Torsion of beams of constant cross section

In order to find a beam's torsional stiffness, the shear stress distribution in the cross-sectional plane (the x, y plane) must be determined. This requires the solution of Poisson's equation for the cross section

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 2G\theta \tag{11.3.6}$$

where ϕ is a potential function, G is the material shear modulus and θ is the angle of twist per unit length of beam. The shear stresses in the xz and yz planes are the partial differentials of z with x and y, respectively. This equation can either be solved by computer methods, or a bubble analog can reveal the stress distribution. The equation for the height z of a soap bubble, supported on a horizontal wire frame is

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = \frac{\Delta p}{\sigma} \tag{11.3.7}$$

where σ is the biaxial membrane stress, and Δp the pressure differential across the bubble. Equations (11.3.6) and (11.3.7) are identical in form, and the analogues at a point on the cross section or bubble are given in Table 11.3.2. A permanent bubble can be made by thermoforming. A hole is cut from a sheet of metal or plywood, with the beam cross section shape. A thin sheet of molten thermoplastic is 'free blown' with a small air pressure through the hole and allowed to cool under pressure. Height contours are then drawn on the bubble.

Many finite element packages allow the solution of a two-dimensional, steady state, heat flow problem in

Table 11.3.2 Analogues between beam torsion, soap bubbles and two-dimensional steady heat flow

Torsion	Soap bubble	Heat flow
Beam boundary shape	Wire frame shape	Material boundary shape
Shear stress magnitude	Maximum slope	Heat flux magnitude
Shear stress direction	Along a height contour	Along a temperature contour
Torsional stiffness	Bubble volume	Total heat generated

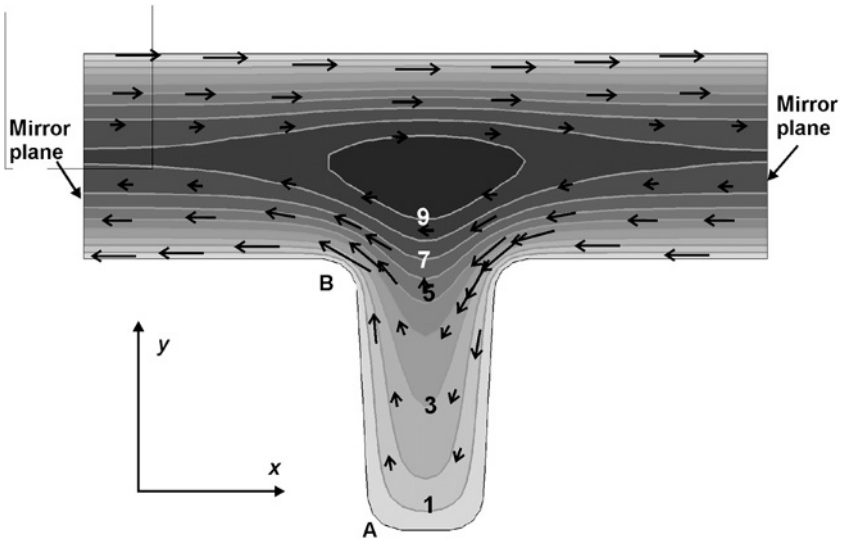


Fig. 11.3.8 Temperature contours (arbitrary units) in a 2-D heat flow, interpreted to show the shear stress direction and magnitude (arrows) for the torsion of a multi-ribbed plate.

a body with a constant heat generation rate A. The differential equation that applies is

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = A \tag{11.3.8}$$

which is another form of Poisson's equation. A boundary at a fixed temperature is the equivalent of the free surface condition in the beam torsion problem. Computed temperature contours (isotherms) can be interpreted like soap bubble height contours (Table 11.3.3), and heat flux magnitude as shear stress magnitude in the twisted beam. Isotherms close to free surfaces (Figure 11.3.8) are parallel to the surface. Hence, the shear stress directions in these locations, are nearly parallel to the boundary. The highest temperature occurs at the midsection, indicating that the shear stress is zero at this location. The shear stresses double back inside the cross section and the total moment of the shear forces about the axis of twist is low. Hence, the torsional stiffness of all open-section shapes is low.

Four different open sections, shown in Figure 11.3.9, all have the same total width $W = 40$ mm, and thickness $t = 2$ mm. As the effects of corners in the I and ribbed sections can be ignored, the torsional stiffness is the same as for a flat plate of width W. As $W \gg t$, the ends of the rectangular section can be ignored; the linear variation of shear stress across the majority of section leads to the torsional stiffness being

$$\frac{T}{\theta} \cong GW\frac{t^3}{3} \tag{11.3.9}$$

To model the torsion of a hollow, thin-walled tube, a bubble is blown through a circular cutout, with a flat horizontal surface, free to rise, representing the interior of the tube. The free part of the bubble has a constant slope in the radial direction, showing that a constant shear stress acts in the tube wall, parallel to the boundary. The bubble volume is proportional to the area A inside the tube, and to the height of the central flat surface. For a constant pressure differential this height is proportional to the wall thickness t. The torsional stiffness of a hollow tube of perimeter P, wall thickness t and enclosed area A is

$$\frac{T}{\theta} \cong 4A^2 G\frac{t}{P} \tag{11.3.10}$$

This reduces to $2\pi R^3 Gt$ for a circular section tube of mean radius R. Figure 11.3.9 ranks the torsional stiffness of some beams of constant cross-sectional area. The preferred designs are hollow tubes; the best has a circular section, as this includes the greatest area A for a given perimeter. The per cent stiffness values only apply to these specific dimensions. If the size of the section is increased, while the thickness t is kept at 2 mm, the advantage of the hollow tubes increases.

Two injection mouldings can be welded together to produce a stiff, hollow section. If the joint is planar, hotplate welding (Section 11.3.2.4) may be possible, but electrothermal welding is preferred for non-planar joints. Figure 11.3.10 shows the section through a car bumper; woven copper braid was placed between three mouldings, of thickness 3.0 mm. When the mouldings are clamped together and a current of hundreds of amperes

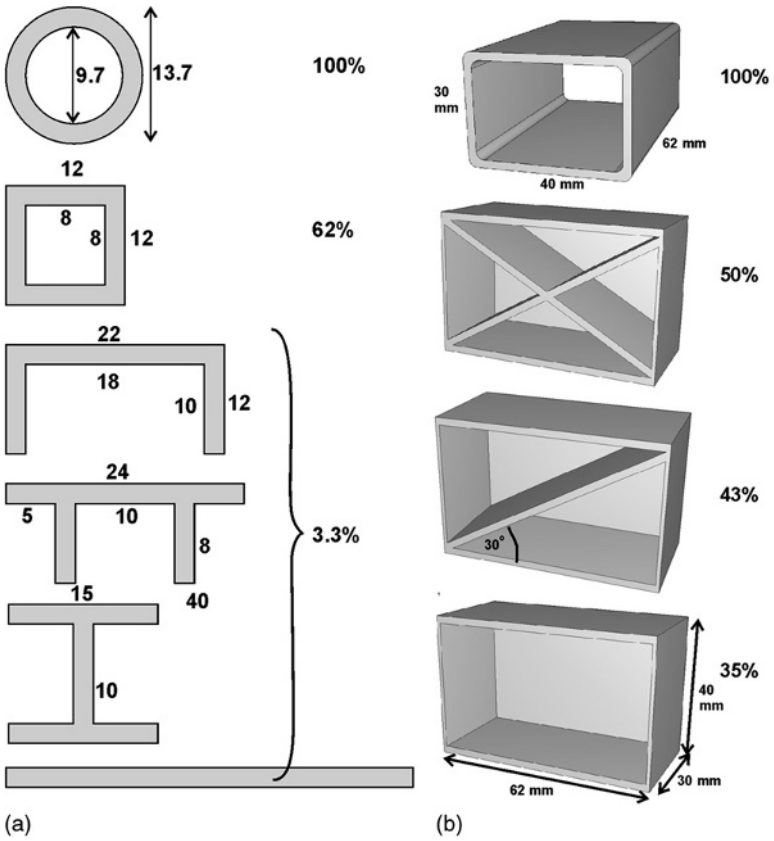


Fig. 11.3.9 Specific torsional stiffness of beams: (a) the same cross-sectional area; (b) cross-ribbed. Dimensions in mm, all wall thicknesses 2 mm wall thickness for torsional stiffness, relative to the thin-walled cylinder.

is passed for a few seconds, a layer of 1 mm of plastic melts, flows through the spaces in the braid and forms a strong weld. The continuous weld along the bumper allows shear stresses to flow uniformly around the section. If the parts were mechanically fastened together, the torsional stiffness would be lower, because shear stresses can only pass through small areas near the fasteners.

11.3.3.5 Torsion of beams of non-constant cross section

It is possible to design injection mouldings with higher torsional stiffness, by varying the beam cross section. Diagonal and cross-ribs, on **I** or **U** beams, take tensile or compressive loads when the beam is twisted. They, with ribs at the sides of the beam, act somewhat like a pin-jointed framework. If the geometrically equivalent framework is a *mechanism* capable of motion (Figure 11.3.11), the beam torsional stiffness will be low. Try twisting the ends of the inner part of a matchbox, or a rectangular plastic tray, and note the high deflection. If the equivalent framework is rigid, the ribbed beam has a high torsional stiffness. A matchbox, with a diagonal piece of card glued in place, has a high torsional stiffness.

The specific torsional stiffness of cross-ribbed beams is lower than that of hollow tubular beams (Figure 11.3.9). In the former, some parts of the structure (the cross-ribs and the back of the U—Figure 11.3.12a) are lightly stressed, while in the latter there is a uniform stress. Hollow tubular beams are resistant to buckling at high torques, since they contain no unconstrained free edges. However, Figure 11.3.12b shows that the free side of a diagonal rib has buckled. The twisted cross-rib has also tilted about a vertical axis.

The relationships between torque and angle of twist, for the ribbed beams of Figure 11.3.9, are shown in Figure 11.3.13. The hollow tube has an almost linear response up to a torque of 55 Nm before the polypropylene yields, whereas the ribbed beams have significantly lower torsional strengths, due to rib buckling. Thus, the failure modes are different. These thin-walled ribbed beams should be redesigned so that they fail in torsion by yielding. The cross-ribbed moulding design has 50% of the specific torsional stiffness and 31% of the specific torsional strength of the rectangular hollow tube.

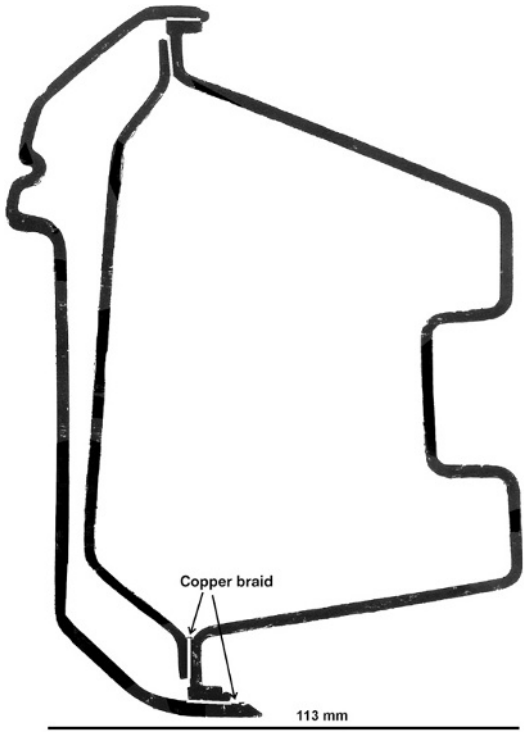


Fig. 11.3.10 Section through a car bumper: three mouldings have been joined with electrofusion welds, to form a hollow beam, with high torsional and bending stiffness.

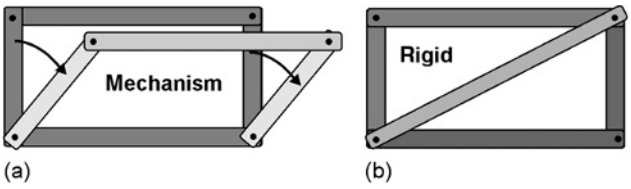


Fig. 11.3.11 Pin-jointed frameworks: (a) mechanism; (b) simply stiff. If the rib pattern on a beam is like (a), the beam will have a low torsional stiffness; if it is like (b), it will be much stiffer.

Car accelerator pedals are non-straight beams, subjected to a bending load when the foot presses down. Although many pedals are stamped from thick steel plate, plastic mouldings are possible. If the moulding lacks torsional stiffness, it can deform by twisting, a form of buckling. It is difficult to injection mould a hollow tube with a bend by using movable cores, that retract along the axis of the tube. One design (Figure 11.3.14) used a beam with a non-constant cross section. A ribbed **U** section was preferred to a ribbed **I** section because of its higher torsional stiffness. The angle between the ribs and the flange was 15°. Half way along the pedal, the direction of the **U** changed, cancelling out the slight twisting effect when the pedal is loaded in bending. Glass-reinforced polyamide was selected for its high strength and toughness.

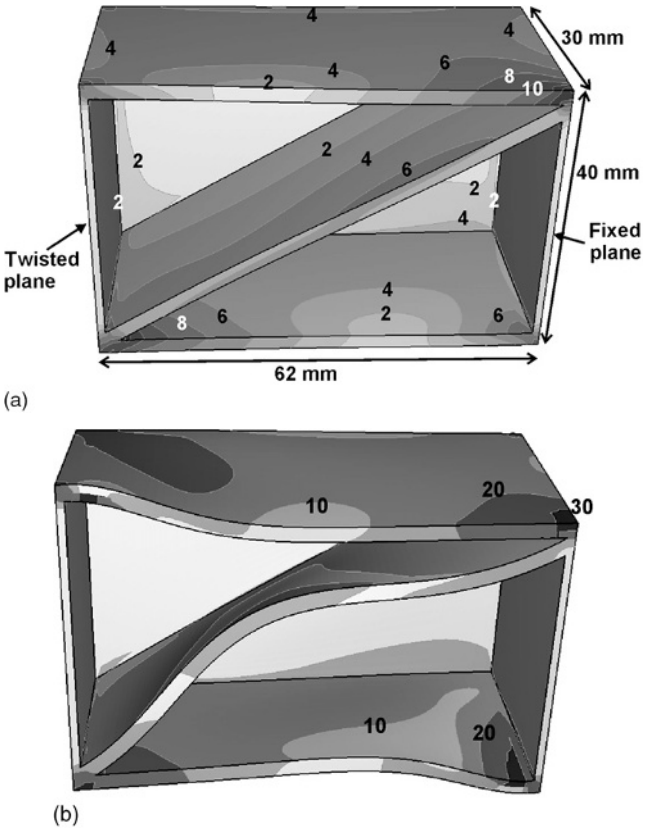


Fig. 11.3.12 Predicted torsion of diagonally ribbed PP beam, with contours of von Mises stress (MPa): (a) at $\theta = 0.035$ rad m^{-1}; (b) $\theta = 0.148$ rad m^{-1}.

The glass fibres orient towards the flow directions as the melt flows along the pedal, increasing Young's modulus and strength along the ribs. The principal stress directions are along the ribs, so the microstructure is optimal.

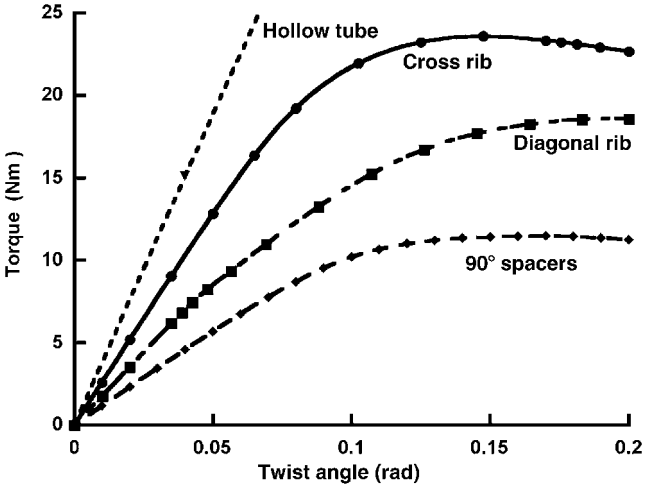


Fig. 11.3.13 Torque vs. twist angle for the ribbed beams shown in the right of Figure 11.3.9.

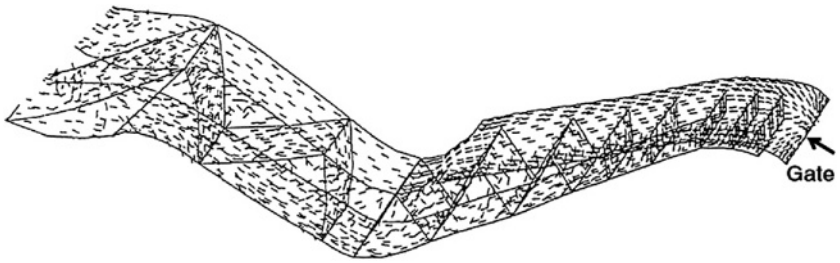


Fig. 11.3.14 Design of a vehicle pedal. The arrows are predicted flow directions for mould filling (from Brünings W.D. *et al.*, *Kunststoffe,* 79, 254, 1979).

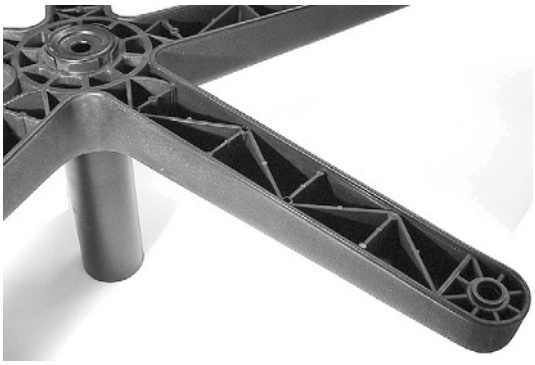


Fig. 11.3.15 Diagonal rib pattern, on the underside of office chair legs, that provides bending and torsional stiffness.

Such ribbing can also be seen on the underside of the starfish-shaped base mouldings of office chairs, in which the seat swivels on a vertical gas-strut suspension (Figure 11.3.15). If the seat is rocked, bending loads are applied to some of the beams; any significant beam torsion could cause the chair to become unstable. There are extra reinforcing ribs where a shaft is inserted for the caster wheel at the beam end, and at the hub around the moulded-in metal suspension tube.

11.3.4 Product shapes for injection moulding

The injection moulding process allows a great variety of possible product shapes. Process optimisation, important for the economics of the product, in part determines the product design. It is important to minimise the cycle time, and to avoid shrinkage flaws, as well as using shapes that maximise the product stiffness. Two simple product design rules, considered in Sections 11.3.1 and 11.3.2, are that the product should either have uniform thickness or the thickness should decrease from the gate. The remaining sections consider how many parts can be integrated in a single moulding.

11.3.4.1 Uniform part thickness

Figure 11.1.1 shows the nearly uniform thickness of the wall of a kettle. This allows all regions of the moulding to complete solidification at nearly the same time. Figure 11.3.16 shows one effect of ignoring this rule. The thin internal strut solidified first, then was put under compressive load by the subsequent shrinkage of the thicker outer region. As the strut had a low bending stiffness, it buckled. When the outer region was cored out to be of uniform thickness, the strut no longer buckled.

Once a mould is full, there is effectively one-dimensional heat flow in the direction perpendicular to the mould wall, while the plastic solidifies. The moulding must solidify all the way through, before it can be ejected, so the cycle time is determined mainly by this 'through-thickness solidification time' t_S. If the plastic is solid at $T - T_0 = 0.4$, the melt is fully solid

Fig. 11.3.16 Buckled internal strut, due to shrinkage of a thin part of a moulding (Hasenauer, J. *et al.*, *Top 10 Design Tips,* Du Pont, UK, © reprinted with permission).

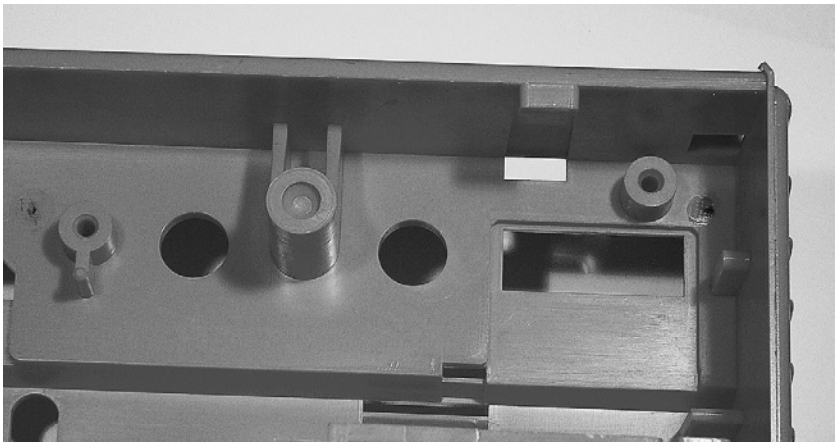


Fig. 11.3.17 Bosses for screw assembly, seen inside a cassette player; the central one is for a screw from the product exterior.

when the Fourier number $Fo = 0.5$. Hence, the solidification time is given by

$$t_s = \frac{FoL^2}{\alpha} \tag{11.3.11}$$

where L is the part half thickness in mm, and thermal diffusivity $\alpha \cong 0.1\,\text{mm}\,\text{s}^{-1}$. For a 4 mm thick part, $t_s = 20$ s, whereas it is 5 s for a 2 mm thick moulding. The thickness chosen is a compromise between mould productivity, which is proportional to L^{-2}, and mechanical properties such as bending stiffness, which increase with L^3.

11.3.4.2 Part thickness that decreases away from the gate

Thin regions cool more rapidly than do thick regions, according to Eq. (11.3.11). Therefore, if the moulding is thicker, in a region remote from the gate, than near the gate, the remote region will have a molten core, cut off from the melt supply, at some stage during cooling. This region cannot be fed by the holding pressure, so shrinks more than other regions. Consequently, either the soft, solid skin is sucked in, leaving an unsightly sink mark, or an internal void forms.

Convex regions of the product surface (A in Figure 11.3.8), cool more rapidly than do internal corners (B) due to the difference in heat flow geometry. Hence, the solid skin is thinnest at concave locations. This can be partly rectified by radiusing both internal and external corners, and keeping the part thickness uniform. The stress concentration factor at rounded corners is lower than at sharp corners when the product is bent. Rounded corners make flow into the mould easier, and are unlikely to cause injury to the user of the product.

Figure 11.3.17 shows that, when bosses are provided for self-tapping screws, they should be separated from the main wall of the product, to avoid the creation of thick regions. Note the number of sharp internal corners in the moulding. The bosses may need to be supported by buttresses that connect to other parts of the product. An extreme example occurs in stacking chairs (Figure 11.1.5).

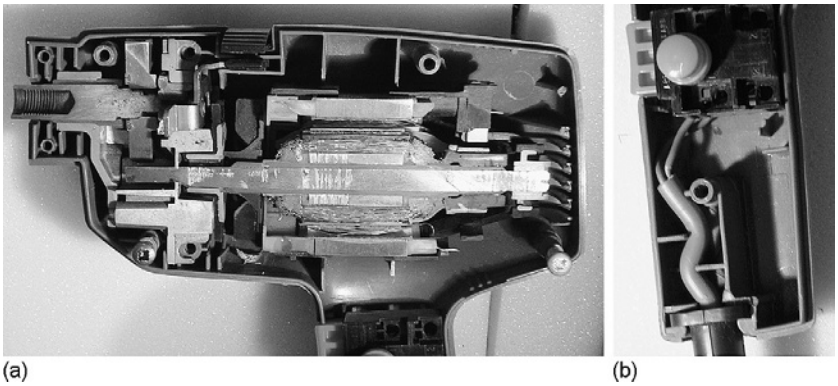


Fig. 11.3.18 A sectioned electric drill, showing internal ribs that locate: (a) the motor and gearbox, (b) the mains cable.

11.3.4.3 Product casings that locate components

Hand-held electric tools, such as electric drills or hot-air blowers, usually have 'clam shell' mouldings. When the two casing halves close, tongues or bosses on one half locate in grooves or recesses in the other. The bending and torsional stiffness of the screwed together casing is reasonably high. Internal ribs in each half casing locate items such as an electric motor and gearbox (Figure 11.3.18a). A maze of partial ribs can prevent the PVC-coated mains cable from pulling from the tool (Figure 11.3.18b).

11.3.4.4 Integral springs and snap joints

Figure 11.1.4 showed separate plastic springs in a video cassette. They use the high elastic tensile strains and the low Young's modulus, of plastics to replace metal coil springs. Thin cantilever springs can also be integrated into injection mouldings. Brochures on the 'plastics.bayer.com' website describe typical designs, and give formulae to calculate the spring stiffness. An interactive programme *FEMsnap*, for a simple cantilever catch, computes the beam stiffness and insertion forces, and gives a contour map of the stresses or strains in the beam. The maximum tensile strains are usually kept below 1%. Figure 11.3.19 shows the principal tensile strain contours and deformed shape of such a catch, at maximum deflection during insertion into a slot in another moulding. The peak strain of 2.5% is close to the maximum allowed for polycarbonate. Such a system acts as a *snap joint*, removing the need for screw fixing. If the catch rear face is perpendicular to the direction of assembly, the product assembly is

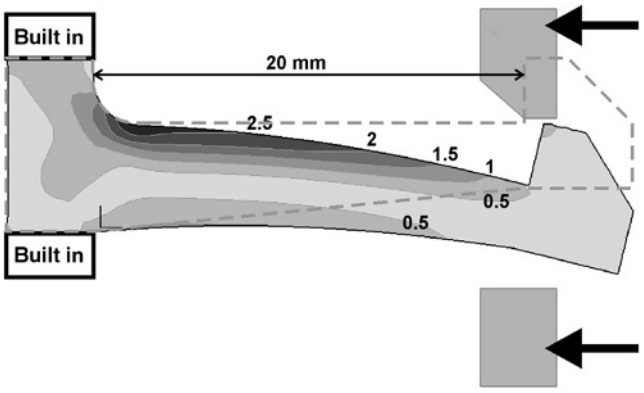


Fig. 11.3.19 Snap joint: an integral cantilever spring at maximum deflection on insertion, compared with the undeformed shape, and principal tensile strain contours (%).

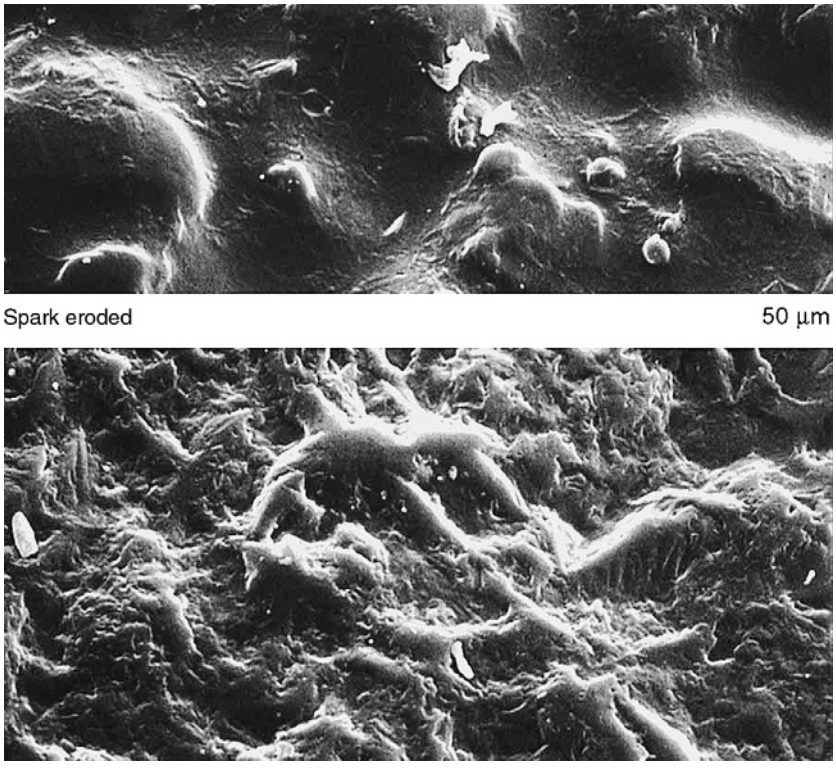


Fig. 11.3.20 Texture on a nylon 6 moulding, due to spark erosion (top) or photo-etching (lower) mould surfaces (from Schauf D, ATI 584e, booklet, © Bayer Material Science AG, 1988).

permanent. However, if angled faces are used, or there is a method of disengaging the catch, the product can be disassembled.

11.3.4.5 Integral hinges

Narrow, thin sections, that run straight across injection mouldings, can act as integral hinges (Figure 11.1.15). These replace the metal pins, pressings and screws of conventional hinges. However, this region is a likely site for failure, unless a highly oriented semi-crystalline morphology can be achieved, with orientation direction across the hinge. Such microstructures are only easy to produce in polypropylene, because of its relatively low melt viscosity, and its propensity to crystallise in an oriented form near the surface. The thickness S of the solid layer that develops during mould filling flow in a time t, as

$$S \cong \sqrt{\alpha t} \tag{11.3.12}$$

As the thermal diffusivity $\alpha \cong 0.1$ mm^2 s^{-1}, a layer $S = 0.2$ mm develops in a mould fill time of 0.4 s. Hence, the mould must be filled fast to avoid the hinge area solidifying and preventing the complete filling of the mould. When the hinge is first flexed, part of the PP stretches and whitens, but it keeps its tensile strength. Even if the PP delaminates parallel to the hinge surface, this does not detract from the hinge performance.

11.3.5 Instrument panel case study

11.3.5.1 Instrument panel shape

A car instrument panel is a large injection moulding with a complex shape. The panel shape was optimised, in ways discussed earlier in this chapter, to increase its bending stiffness. The visual effects of scratching are minimised by having a textured surface on the mould. Figure 11.3.20 shows a matt surface on a similar moulding that scatters the majority of incident light, so there are no direct reflections. The spark-eroded mould surface provides rounded features on the moulding, with a better appearance than those from the photo-etched mould surface. An alternative is to use a 'soft touch' finish as in the Ford Focus—the moulding has a surface layer of a low modulus ethylene copolymer, that is difficult to scratch. The effects of the differential thermal expansion coefficient between the plastic instrument panel and the steel body are disguised, by making sure that any gaps are hidden from view.

11.3.5.2 Free head-form impact tests

Federal Motor Vehicle Standard FMVSS 201 requires that a 4.5 kg head-form is fired at 15 mph at any interior region that could be hit by the head of an unbelted vehicle occupant. There is a triaxial accelerometer inside the head-form, which has a rubbery skin. A weighed integral of the head-form acceleration must not exceed a level that would cause moderate head injury. This is roughly equivalent to the peak acceleration being less than 140 g. Consequently, the instrument panel must absorb a kinetic energy of 100 J and it must not have any rigid projections on its surface. Figure 11.3.21a shows a finite element simulation of the head-form impacting an instrument panel, and Figure 11.3.21b shows more details of the head-form impacting a ribbed plastic moulding, covering the roof rail at the top of the windscreen A pillar. For the impact on the

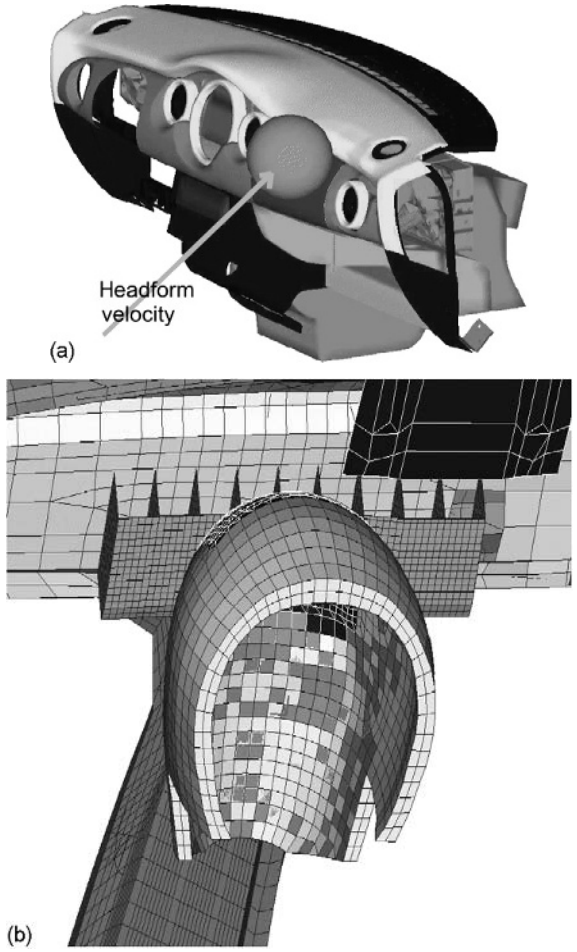


Fig. 11.3.21 FEA of head-form impact on (a) an instrument panel (Silk, G., SAE Technical Paper 2002-01-1270; Haque *et al.*, SAE Paper 2000-01-0624 for an impact) (b) a ribbed moulding at the top of an padded A pillar © SAE International.

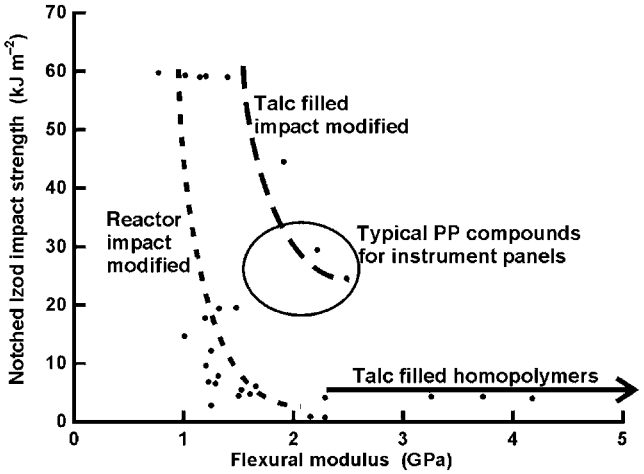


Fig. 11.3.22 Young's modulus vs. charpy impact strength at 23 °C (redrawn from Juan, P. *et al.*, SAE paper 980067).

instrument panel, the peak forces would be about 6 kN, and peak deflections about 40 mm.

11.3.5.3 Grade development

ABS, with an upper layer of plasticised PVC to simulate leather, was initially used. However, by the late 1980s, PP grades were developed to suit the application. Nucleated grades had increased crystallinity, hence increased yield stress (35 MPa at 23 °C) and Young's modulus. The PP was toughened by adding compatible ethylene propylene copolymer rubber, which is well bonded to the PP matrix. Talc was added to increase Young's modulus and the temperature resistance of the instrument panel. Talc acts as a nucleating agent for PP. PP crystals grow epitaxially on the talc platelets, with the **c** spacing of the PP crystal being close to the **b** spacing of the talc crystal.

Due to the head-form impact test in Federal Motor Vehicle Standards, the PP should have a high-impact strength at low temperatures. Figure 11.3.22 shows how the grades used for instrument panels are superior to just rubber-toughened or talc-filled PP in having high values of both Young's modulus and Izod impact strength.

11.3.5.4 Recycling

European regulations about recycling cars at the end of their lives are slated to come into effect in 2007. The instrument panel is an example of a major component that has not been designed for easy dismantling. It is currently time consuming to remove the steering wheel, instruments and the wiring harness from the many places of attachment. In future, redesign will make recycling easier.

This page is intentionally left blank

Index

"1-sigma", 202
5th percentile force, 246
10 × rule, 304
(α + β) alloys, 221
α alloys, 219
α-stabilizers, 219, 220
β alloys, 219, 221
β-eutectoid elements, 219
β-isomorphous elements, 219
β-stabilizers, 219, 220
γ-loop, 208

A
Absolute density, 494
Absolute pressure method, 552
AC small-signal analysis, 392
Accommodation of eye, 234
Acrylic adhesive system (TS 3929-70), 535
Adaptation process, 234–5
Aesthetic attributes, 190–2
  hearing, 192
  sight, 191–2
  touch, 190–1
Alloy steels, 208–9, 213
Alloys, advanced, 207
  aluminum alloys, *see* Aluminum alloys
  cast irons, 213–15, 216
  commercial steels, *see* Commercial steels
  structural intermetallic compounds, *see* Intermetallic compounds, structural
  superalloys, *see* Superalloys
  titanium alloys, *see* Titanium alloys
Alternative design ideas, need for, 59
  choosing best combined set from all alternatives, 59
  choosing best from alternatives, 59
  design optimization, 60
  reducing risk, by use of alternative design, 59
Aluminum alloys:
  aluminum–lithium alloys, 228–9
  applications, 227–8
  designation, 227
  eco-attributes of one grade of, 178
  processing developments:
    mechanical alloys, 230
    rapid solidification processing, 229–30
    superplastic aluminum alloys, 229
Aluminum–lithium alloys, 228–9
Aluminum-structured vehicle technology (ASVT), 228
Ambient illumination, 289
American Association of Railroads (AAR), 534
American Society for Testing and Materials (ASTM), 569
Amorphous TPs, 489
Analogies and chance, in solving problems, 36–8
Analytic methods, 38–9
Anthropometry, 242–5
  body size application principles, 244–5
  body size variability, sources of, 243–4
Apollo program, 108–9
Apparent density, 494
Apparent viscosities, 495
Appraisal costs, 300, 301
Assemblability, 333
Assembly capability, 332–3
  assembly sequence diagrams, 335
  assembly techniques, design for, 334–5
Assembly Evaluation Method (AEM), 334
Astroloy, 217
Atomic weight, 495
Attention process, 237–8
Auditory sense, 237
Austempered ductile iron (ADI), 215
Author, *see* Writer
Automobile, designing, 563
  aircraft, 564–5
  marine application, 566
    boat, 566
    underwater hull, 566–7
  truck, 563–4
Automotive EMC Directive, 433

B
Back compression design limit (BCDL), 256
Back compression upper limit (BCUL), 256
Base-line PIC devices, 402
Bath-tub curve, 305–6
  infant mortality period, 306
  useful life period, 306–7
  wearout period, 307
Bauschinger effect, 211
Beam, 512–14
Bearings, 474–5, 526
Belts, 475–6
Bio-materials, 201
Biomechanical assessment, 263
Black box, 34, 35
Black-heart malleable iron, 214–15
Blackbodies, 276
Blank data summary, 96
Blank editor advice log, 95
Blank master plan, 95
Blank process meeting log, 96
Blender:
  customer needs statement for, 12
  product plan for, 14
Blister carded packaging, 551
Blow moldings (BM), 502
Body size application, principles of, 244–5
Body size variability, sources of, 243–4
Boothroyd–Dewhurst DFMA technique, 334, 335
Brainstorming, 36
Brand image research, 11
Broadband over Power Line (BPL), *see* Power Line Telecommunication
BS5760, 374–5
Bubble pack, 551
Bulk density, 494–5
Bulk molding compounds (BMCs), 493
Business and product evolution flow chart, 148–51
Business development game:
  playing, 145
  winning, 145–7
Business objective, 103
  background:
    new product development and economy, 103–9
    research and development, differentiating, 109–12
  perspectives external to organization, 126
    global product and business development, 128–9
    market investigation, 129–33
    marketplace, defining, 126
    marketplace, understanding, 126–8
  perspectives internal to organization:
    consistency, 120–2
    continuity, 122–3

Business objective (*continued*)
dead-ended single-product programs, 120
flow, 124–5
future, 117
introspection, 115–17
leverage, 123–4
product fit, 112–13
product strategy, 113–15
pursuit, 125
transition, 117–19

C

Calendering, 504
CAMPUS, 583
Cantilever spring, 547
Capable components and assemblies, designing:
assembly capability, 332–3
assembly sequence diagrams, 335
assembly techniques, design for, 334–5
case studies:
electronic power assisted steering hub design, 352–5
solenoid end assembly, 357–63
solenoid security cover, 355
telescopic lever assembly, 355–7
completing conformability matrix, 350–2
component assembly variability risks analysis, 335–7, 348–50
manufacturing capability, 319–20
component manufacturing variability risks analysis, 322–30
cost–tolerance relationships, 321
process capability and tolerances, 321–2
surface roughness chart, 330
validation of component manufacturing variability risks analysis, 330–2
variability factors in, 320–1
non-conformance effects:
design acceptability, 337–41
quality costs, map of, 341–3
objectives, application and guidance for analysis, 344–8
CAPP screening worksheet, 263
Car accelerator pedals, 592
Car body materials, selection parameters for, 585
Car door skin, 585
Carbon and alloy steels:
compositions of, 213
Carbon and carbon–manganese steels, 213
Carbon steels, 207–8
stress–strain curves for, 210
Carpal tunnel syndrome (CTS), 250
Cast irons, 213–15, 216
microstructure of, 214
Castings, 504
CE mark, 436–7
Cementite, 213
Ceramics, 193
Chains, 475–6
Characteristic spectral distribution, 276
Checkers, 85–6
Checking rate, 80
Circuit simulation, 391–2
AC small-signal analysis, 392
DC analysis, 392
logic simulation, 397
netlists and component models, 395–7
noise analysis, 395
pole-zero analysis, 393–4
practical investigation, 397–400
problems, 400
sensitivity analysis, 395
small-signal distortion analysis, 394–5
thermal analysis, 395
transient analysis, 393
Classic Plastics Inc., 55
Clutches and brakes, 477
Coatings, 504
Co-efficients of linear thermal expansion (CLTE), 492
Coextrusions, 500
Coffee maker, automatic:
function analysis for, 35
Cognitive capabilities, 237–40
attention, 237–8
decision making, 240
memory, 238–40
perception, 238
Cognitive performance and noise:
relationship between, 283–4
Cognitive tasks:
effect of cold on, 279
effect of heat on, 278
Coinjections, 501
Cold effects, on performance, 279
cognitive tasks, 279
physical activities, 279
Cold environment, 280
exposure and tolerance to cold exposure, 280
working conditions in, 282
Cold stress, *see* Hypothermia
Color deficiency, 235
Color rendering index (CRI), 276
Color temperature, 276
Color vision, 235
Column formulas, 517–18
Comet G-ALYP, 25
Comet G-ALYU, 26
Comet G-ALYV, 25
Comet G-ALYY, 25
Comfort and discomfort zones, 280–1
Commodity plastics (CPs), 493
Common law, 385
Competent Body, 447
Competitive arena:
competitive game, 151–2
large versus small companies, 152–4
Complex cognitive tasks, 278
Component assembly variability risks analysis, 335–7, 348–50
Component manufacturing variability risks analysis, 319, 322–6, 347–8
Composite Concept Vehicle (CCV), 563
Composite materials, development of, 201
Compression moldings (CM), 504
Computer-generated design drawing, 22
Computer technology:
population and time, 136
wireless technology, 135–6
Concept challenge method, 38
Conceptional design review, 49
Conceptual design, 459, 464
boundary shifting, 464
brain-storming, 464–6
function tree method, 467–8
morphological analysis, 466–7
Conduction, 278
Conduction deafness, 284
Conductive plastics, 569
Conformability analysis (CA) methodology, 304, 344–52, 366–7
Conformability Map, 341–3
Conformability Matrix, 346, 350, 352
Conjoint analysis, 11
Constraints, 581
Consumer products, design spectrum for, 19
Consumer Protection Act 1987, 386, 387–8
Contact stress, 526
Containers' energy content, case study of:
method and results, 183
problem, 182–3
Continuous process improvement, 99–100
Contour packaging, 551–2
Contract clauses:
essential areas, 377
definitions, 377–8
demonstration and prediction, 378–9
design standards, 380

environment, 378
liability, 379
maintenance support, 378
reliability and maintainability analysis, 379–80
reliability and maintainability programme, 379
safety-related equipment, 380
storage, 380
examples, 382–3
penalties, 381, 382
apportionment of costs during guarantee, 381
payment according to downtime, 381
and pitfalls, 380
definitions, 380
environment, 381
liability, 381
quoted specifications, 381
repair time, 380
statistical risks, 380–1
subcontracted reliability assessments, 382
Contract law, 385
Convection, 278
Corporate skills assessment, 154
business capabilities, 154–5
capabilities, partnering for, 155–6
collaborative communications, 157
contingency plans, 156–7
corporate strategy, 158
getting tools and remembering objective, 156
mutual opportunity, focus on, 158
Cost reductions, 106
Cost–tolerance relationships, 321, 322
Coupling multipliers, 255
Crash barriers, case study of:
method and results, 184
problem, 183
Creative design review, 49
Critical path method (CPM), 14
Cross-linked thermoplastics, 490–1
Crystalline plastics, 489, 492
Cumulative cash flow and product life cycle, relationship between, 4
Cumulative trauma disorders (CTDs), 249
physical activity risk factors related to, 249–50
Curiosity-driven research, 200–1
Cushioning, 553–4
Customer needs statement:
for blender, 12
for rechargeable electric toothbrush, 12
Cyanoacrylate, 37
Cylinder, designing, 530

D

D of C, *see* Declaration of Conformity
Dark adaptation, 234
DC analysis, 392
De Havilland Comet 1, 25
Decision making, 240
Declaration of Conformity (D of C), 437
Defect Detection Process (DDP), 80
Defect Prevention Process (DPP), 78, 80
Deformation, 177
Demonstrations, 129–30
Design, meaning of, 19–20
Design and design engineering, difference between, 57
Design brief, 461
Design constraints, 59
Design for Assembly (DFA), 310
Design for Manufacture (DFM) techniques, 311
Design for Manufacture and Assembly (DFMA), 334
Design for Quality (DFQ), 310–13
Design for six sigma (DFSS), 45–6
Design ideas, 58
design constraints, 59
potential design ideas, 59
Design intent, *see* Design brief
Design methods, definition of, 35
Design Optimization Tradeoff, strategy options for, 60
Design paradigms, 21
confirming false hypothesis, avoiding, 25–6
model, need for, 21–2
redundancy, need for, 22–3
scale effect, 23–4
starting problem analysis in middle, avoiding, 24–5
tunnel vision, avoiding, 26–7
Design review process, 49
case study, 53–4
experimental designs, 51–3
failure mode and effects analysis, 51
problems with implementation of, 50
sequence, 50
soft-hard (SH) review, 51
steps in, 49
Design specification:
language core, 63
rules, 63–5
template, 74–5
see also Specification quality control (SQC)
Design specifications masquerading as requirements, case study, 72
DesignforAssembly (DFA) techniques, 334
Detailed design, 459
Deterministic design, 314
versus probabilistic design, 316
DFA/Manufacturing Analysis (MA), 334, 335
Diathermy and electrosurgery, 418
Direct analogies, 37
Direct glare, 276
Disability glare, 277
Discomfort glare, 277
Disconnection index, 136
Dispersion-hardened superalloys, 218
Divided attention, 238
DMADV, 46
Dough molding compounds (DMCs), *see* Bulk molding compounds
Dow house of future, 556
Drying operations, 498–9
Du Pont, 550, 573
Dual-ovenable tray, 552
Dual-phase (DP) steels, 210–12
stress–strain curves for, 210
Ductilities, 492
Dunlop, John Boyd, 36
Dymetrol® elastomeric tape, 550
Dynamic anthropometry, *see* Functional anthropometry
Dynamic dimensions, *see* Functional anthropometry
Dynamic effort, *see* Dynamic work
Dynamic mechanical thermal analysis (DMTA), 496
Dynamic priority evaluation, 71–2
Dynamic work, 241, 242

E

Eco-attributes, of materials:
eco-indicators, 177–8
end of life, 177
material processing energies, at 30% efficiency, 176–7
material production: energy and emissions, 175–6
Economic quality-cost models, 303
Eco-selection, 178
material production phase, 178–82
product disposal phase, 182
product-manufacture phase, 182
use phase, 182
Editor, 86
Editor Advice Log, 86
EI theory, 509–10
Elastomers, 201
Elastomers/rubbers, 493–4
Electric coffee mill, step-by-step addition of subfunctions for, 39
Electrical/electronic device, designing, 567
radome, 570
shielding electrical device, 567–70
Electro-explosive devices (EEDs), 427

Electromagnetic compatibility (EMC), 415–16, 567–8
assessment, 445
compatibility between and within systems, 419
inter-system EMC, 419–20
intra-system EMC, 419, 420
electromagnetic fields and human health, 428
athermal effects, 429
ICNIRP basic restrictions, 428–9
medical devices, interference to, 417–19
portable electronic devices in aircraft, 416–17
quacking duck, 419
scope:
compatibility gap, 427–8
control systems, malfunction of, 420–1
disturbances of mains supply, 423–4
interference with radio reception, 421
issues, 426–7
limits, setting, 421–3
powerline telecoms, 424–6
Electromagnetic energy (EME), 569
Electromagnetic fields and human health, 428
athermal effects, 429
ICNIRP basic restrictions, 428–9
Electromagnetic interference (EMI), 415, 567, 568, 569
Electronic packaging, 552
Electronic power assisted steering hub design, case study, 352–5
EMC directive, 431
apparatus with Directive, compliance of, 444
notified body, 446–7
self certification and internal production control, 444–6
testing, 447
using standards, 448–50
compliance for product manufacturer, 450–1
first EMC Directive, 432
Automotive EMC Directive, 433
Low Voltage Directive (LVD), 433
SLIM, 433
legislation, 432
new approach Directives, 431–2
second EMC Directive, 433
CE mark and paperwork, 436–9
changes, 433–4
components, 436
exceptions, 435–6
fixed installations, 439–42
implementation, enforcement and sanctions, 443–4
manufacturing quality assessment, 439
requirements, 434
sale and use of products, 434–5
system, 443
EMC Test Laboratories Association (EMCTLA), 444
EMCD, 437, 439, 442
Enclosures, 478–9
Energy consumption:
and material consumption, 174–5, 176
Engineering and industrial designs, relative costs of, 20
Engineering design process, 28–9, 30, 31
divergence and convergence in, 32
general approach to, 31
Engineering plastics (EPs), 493
Entrepreneurs, licensing, and technology transfer:
dealing with, 168–9
Environment:
and individual, concern for, 203–4
and materials, *see* Materials and environment
Environmental factors, 273
case study:
final results, 290
method, 288
pilot lighting system, installation of, 289–90
recommendations, 289
results, 288
illumination, 273
age factors, 275
lighting quality, 276–7
lighting quantity, 274
task factors, 274–5
noise, 282
effects of noise on health, 284–5
effects of noise on performance, 283–4
guidelines to control noise, 285–6
temperature, 278
comfort and discomfort zones, 280–1
effects of cold on performance, 279
effects of heat on health, 279–80
effects of heat on performance, 278–9
recommendations to improve working conditions, 282
work tolerance in hot environment, 281–2
vibration, 286
effects of vibration on health, 287–8
effects of vibration on performance, 287
guidelines to reduce or control vibration, 288
Epoxy adhesive, 535
Ergonomics program, 260
evaluating, 263
planning, 260
risk assessment process, 260–2
solutions, 262–3
Euler's formula, 517–18
EuropeanEconomic Area (EEA), 434
Evaporation, 278
Event map, 9
Evolutionary Project Management (Evo) process, 58, 60–1
Existing products, improvements and revisions of, 105–6
External design review, 50
External failure costs, 300, 301
Exteroceptors, 233
Extrusions, 500

F

Fabricating processes, 499
extrusions, 500
injection moldings, 500–2
thermoformings, 502–3
Failure costs, 295, 300
Failure mode and effect analysis (FMEA), 51, 300
in capable and reliable products designing, 309–10
occurrence ratings, 340
severity ratings, 340
FALLO approach, 505, 506
Fantasy analogies, 37
Farsightedness, 234
Fasteners, 478
Fault Tree Analysis (FTA), 313
Federal Works Administration, 26
FEMsnap, 595
Fick's First and Second Laws of Diffusion, 552
Filament winding (FW) shapes, 528
Filled-in data summary, 84
Filled-in editor advice log, 83
Filled-in master plan, 82
Finite element analysis (FEA), 510, 523, 535, 539, 587, 589
Firth of Forth Bridge, 23
Fixing bolt assembly, variability risks table for, 351
Flash program memory, 402
Flowcode, 406
Flux, 273
Foams, 503
Focused attention, 238
Folded plate technology, 511–12

Forces for change, 197, 198
environment and individual, concern for, 203–4
growing population and wealth, and market saturation, 201
market-pull and science-push: curiosity-driven research, 200–1
market, power of, 197–200
miniaturization and multifunctionality, 202–3
product liability and service provision, 201–2
Formal design reviews (FDRs), 49
Fossil-fuel energy, 176
Fowler, Sir John, 23
Free-cutting steels, 213
Free radiation frequencies, 422, 423
Frequency, 282
Frequency multipliers, 254
Function tree method, 467–8
Functional anthropometry, 243
Functional safety capability, 374
Functionally-graded materials, 201
Funds flow model:
manufacturer, 103
service business, 104
Future house, designing, 555–6
FW process, 535

G

Gantt charts, 14
Gap analysis, 9
Gas-assist injection molding (GAIM), 501–2
Gasket, 527–8
Gears, 475–6
designing, 523–6
General Motors Corp., 301, 550
Geometric shape, of plastics, 509
GFRP (glass fiber-TS polyester RP), 532
Glare, 276–7
Glass, 193
Glass-reinforced polyamide, 592
Glass transition temperatures, 496–7
Global product and business development, 128–9
Goodyear, Charles, 36
Graphite, 213
Gray cast irons, 213, 214
Grommet, 527

H

H Factor analysis, 53
Hand-intensive repetitive tasks, 257
hand tools, 258–9
risk factors, 257
survey and observation tools, 258
Hand tools, 258–9
Head-form impact tests, 596–7
Health And Safety at Work Act 1974, 386–7
Hearing, 192
Hearing loss:
aging, 284
noise-induced, 284–5
permanent, 284–5
Heat cramps, 279
Heat effects:
on health:
cold environment, 280
hot environment, 279–80
on performance, 278
cognitive tasks, 278
physical activities, 278–9
Heat exhaustion, 279
Heat rash, 279
Heat stroke, 279, 280
High-alloy steels, 213
High-carbon steels, 208
High-end PIC device, 402
High-strength low-alloy (HSLA) steels, 209–10
stress–strain curves for, 210
Hinge, 548
Hookean stresses, 577
Hoop-stress equations, 548
Hospital and emergency service radio management, 417–18
Hot environment, 279–80
exposure and tolerance to, 280
work tolerance in, 281–2
House of quality (HOQ) matrix, 47, 48
House top, designing, 556–7
Human knowledge, faster doubling effects of:
Human/system interface (HSI) model, 233
Hypothermia, 280

I

Illuminance, *see* Illumination
Illumination, 273
age factors, 275
lighting quality, 276–7
lighting quantity, 274
task factors, 274–5
Impact Estimation (IE), 60
Impact estimation table, 62
Inclusive design, 203
Inco 718, 217
Inco 901, 217
Indirect glare, 276
Industrial designers, 187
Industrial, scientific and medical (ISM) equipments, 422, 423
Information processing, 237, 241
Information Resources, Inc., 12
Injection moldings (IM), 500–2
product shapes for, 593
integral hinges, 596
integral springs and snap joints, 595–6
part thickness, 594
product casings, 595
uniform part thickness, 593–4
Instrument panel case study, 596
free head-form impact tests, 596–7
grade development, 597
instrument panel shape, 596
recycling, 597
Integral skin foaming, *see* Structural foam
Integrated Development Environment (IDE), 406, 408
Intelligent materials, 201
Intensity, 282
Intermediate impossible method, 38
Intermetallic compounds, structural:
general properties of, 222–3
nickel aluminides, 223–4
titanium aluminides, 224–5
Intermetallics, 222
Internal and external balance, 168
Internal design review, 50
Internal failure costs, 300
Internal product development, 158
funding, 160
internal skills development, 159–60
management development, 160
International Commission on Non-Ionising Radiation Protection (ICNIRP), 428–9
International Standards Organization (ISO), 43
Inter-system EMC, 419–20
Interviews, 129
Intra-system EMC, 419, 420
Intrinsic/contained energy, 176
Inverse creep, 224
Isostatic testing equipment, 552
Isotensoid, 529

K

Kin documents, 81
Knowledge gap, 296, 297

L

Large versus small companies, 152–4
Larsen-Nielsen method, 22
Larson–Miller parameter, 219
Laundry detergent purchase trends, in United States, 12
Leonardo da Vinci's flying machine, 20
Liberty ships, 24
Light adaptation, 234
Light stimulus, 273

Lighting:
color, 276
quality, 276–7
quantity, 274
Lighting systems, recommendations for, 289
Liquid crystalline polymers (LCPs), 489–90
Liquid injection molding (LIM), 521
Living hinges, 548
Logic simulation, 397
Logical page, 80
Long-term memory (LTM), 240
Longitudinal Conversion Loss (LCL), 425
Lost opportunities, 300, 301
Low-carbon steels, 207–8
Low Voltage Directive (LVD), 433
Luminance, 274
Luminance ratio, 277

M

Manual handling tasks, 250
evaluation methods, 252
biomechanical models, 253–6
postural observation checklists, 252
weight limit, calculation of, 252–3
general guidelines for, 252
risk factors, 250–2
Manufacturing financial model, 104
Maraging steels, 209
Maritime Commission, 24
Market characterization, 138–9
leading-edge technology, 139
life cycle and market cycle implications of profit, 139
low-cost producer, 139
market maturity, 138–9
Market investigation, 129–33
correlating data, 133
customer, interviewing, 131
customer complaint, extracting opportunity within, 132–3
customer engagement, 131–2
knowledge and experiences, exchange of, 133
primary market research, 129–30
rapport, expertise, assistance, improvement, and authority, establishing, 133
secondary market research, 130–1
surveys and customer feedback, 132
testing data, 133
Market opportunity, 135
change, seasons of, 136–7
competitive arena:
competitive game, 151–2
large versus small companies, 152–4
dynamics, of marketplace:
fast-changing world, 137–8
market characterization, 138–9
faster-changing world, 135–6
product concept synthesis:
business concept embodied on product idea, 139–41
customer's problem, solving, 141–2
idea evaluation within business framework, 142–4
product integration into business:
business and product evolution flow chart, 148–51
business development game, playing, 145
business development game, winning, 145–7
competitive weapon, new product is, 148
product as vehicle to profitability, 144–5
product, pathways to
corporate skills assessment, 154–8
entrepreneurs, licensing, and technology transfer, dealing with, 168–9
external means to product development, 160–7
internal product development, 158–60
keeping balance internally and externally, 168
Market-pull, 197
and science-push:
curiosity-driven research, 200–1
market, power of, 197–200
Market research techniques:
qualitative techniques, 11
quantitative techniques, 11
Market saturation, 201
Marketplace:
defining, 126
understanding, 126–8
Marketplace dynamics, 137–9
fast-changing world, 137
diverse value systems, 138
reduced ability to cope with changes, 138
market characterization, 138
leading-edge technology, 139
life cycle and market cycle implications of profit, 139
low-cost producer, 139
market maturity, 138–9
Massachusetts Institute of Technology (MIT), 572
Material and energy-consuming systems, 174
use-patterns of products, 175, 176
Material and shape selection, 583
injection moulding, product shapes for, 593
integral hinges, 596
integral springs and snap joints, 595–6
part thickness that decreases away from gate, 594
product casings that locate components, 595
uniform part thickness, 593–4
instrument panel case study, 596
free head-form impact tests, 596–7
grade development, 597
instrument panel shape, 596
recycling, 597
polymer, *see* Polymer selection
Material life cycle, 173–4
Material processing energies, at 30% efficiency:
deformation, 177
melting, 176–7
vaporization, 177
Materials and environment, 173
drink containers and crash barriers, case studies, 182–4
eco-attributes:
aggregated measures, 177–8
end of life, 177
material processing energies, at 30% efficiency, 176–7
material production, 175–6
eco-selection, 178
material production phase, 178–82
product disposal phase, 182
product-manufacture phase, 182
use phase, 182
material and energy-consuming systems, 174
use-patterns of products, 175, 176
material life cycle, 173–4
Materials and industrial design, 187
product character, 188–9, 190
product personality creation, 189
aesthetic attributes, 190–2
associations and perceptions, 193–5
requirements pyramid, 187–8
Matrix scoring model, 9
Mechanical alloying (MA), 212, 218
Mechanical design, 455
conceptual design, 464
boundary shifting, 464
brain-storming, 464–6
function tree method, 467–8
morphological analysis, 466–7
process, 456–9
product design specification (PDS), 461–4

technology base, 468, 469–70
total design, 459–61
Mechanical elements, 473
bearings, 474–5
belts, 475–6
chains, 475–6
clutches and brakes, 477
enclosures, 478–9
fasteners, 478
gears, 475–6
seals, 476–7
springs, 477–8
tribology, 474
Mechanical failure, 305
designer's responsibility for, 307
frequency and causes of, 306
Mechanically alloyed (MA) steels, 212–12
Medicines and Healthcare Products Regulatory Agency (MHRA), 418
Medium-carbon steels, 208
Mega-projects, 197
Melt flow index (MFI), 495
Melt index (MI), 495–6
Melt temperatures, 497–8
Melting, 176–7
Memory, 238–40
'Meso-scaling', *see* Quality Loss Function
Metagenes, 21
Metals, 193
MicroElektronika EasyPIC3 PIC development system, 411
Micromoldings, 502
Micron-scale fabrication methods, 201
Micro-scaling/bottom up approach to quality costs, 303–4
Mid-range PIC device, 402
Milwaukee Electric Tool Corp., 573
Mind, and materials, 193–5
expression through material, 193–5
Miniature electro-mechanical systems (MEMS), 201
Miniaturization and multifunctionality, 202–3
Missile and rocket, 567
Mistake proofing (Poka-Yoke), 45, 46–7
Modulus of elasticity, of plastics, 509
Molecular weight (MW), 495
Molecular weight distributions (MWD), 495
Monocoque structure, of plastics, 509
Monolithic ceramics, 201
Motor control, vibration effects on, 287
Motor performance, *see* Musculoskeletal movement
Motorola, 9
MTBF, 371
and failure rates, 377
Muscular endurance and fatigue, 242
Muscular strength and endurance, 241–2
Musculoskeletal disorders (MSDs), 249–50
Musculoskeletal movement, 241
Myopia, *see* Nearsightedness

N

Nano-structured materials, 201
National Aeronautics and Space Administration (NASA), 109
National Institute of Occupational Safety and Health (NIOSH):
equation and figure, 255
multipliers and task variables, 253
Nearsightedness, 234
NEAT plastics, 487
Nerve deafness, 284
Netlists and component models, 395–7
Netting analysis, 528–9
New product development and economy, 103–9
Apollo program, world without, 108–9
cost reductions, 106
economic development, 103–4
existing product lines, additions to, 105
existing products, improvements and revisions of, 105–6
narrative and financial review, of nondeveloper contribution, 106
changes in environment, response to, 107
cost of sales, 106
market share, 107
profits, 107
strengths, 107
new product lines, 105
"new to the world" products, 105
repositioning, 106
New product lines, 105
New product plans, development of, 11–15
consumers and their needs, understanding, 11–12
market, understanding, 12–13
product attributes and specifications, 13
schedules, resources, financials, and documentation, 14–15
New products, opportunities for, 9, 10
"New to the world" products, 105
Newham Council, 22
Newtonian/non-Newtonian viscosities, 495–6
Nickel-based superalloy development, 217–18
Nickel aluminides, 223–4
Noise, 282
analysis, 395
control, 285–6
effects:
on health, 284–5
on performance, 283–4
guidelines to control, 285–6
Noise factors, 320
Noise-induced hearing loss, 284–5
Nondeveloper contribution, 106
changes in environment, response to, 107
cost of sales, 106
market share, 107
profits, 107
strengths, 107
Notified body, 446–7
Nuisance and distraction, 284

O

Objective tree method, 29, 33
for child car seat, 33
for safe automobile travel, 34
Objective tree method, *see* Function tree method
Official Journal of the European Union (OJEU), 445
Optimal lifting task, 252
Optimum checking rate, 80–1
Organization, perspectives external to, 126
global product and business development, 128–9
market investigation, 129–33
marketplace
defining, 126
understanding, 126–8
Organization, perspectives internal to:
consistency, 120–2
continuity, 122–3
dead-ended single-product programs, 120
flow, 124–5
future, 117
five-year product line catalog, 117
industry trends, 117
narrative of plan and reality, 117
narrative of vision, 117
product scope definition, 117
specific market segments, 117
theme/mission statement development, 117
introspection, 115–17
leverage, 123–4
product fit:
company's capabilities, understanding, 112–13
goals and agendas of company, 113
new product opportunity, selecting, 112
sacrifice, 113

Organization, perspectives internal to (*continued*)
product strategy:
company's strategy in pursuing market, 113
contingency planning, sales scenario with, 114
development team, views of, 113–14
market and company, connection between, 114
planning, framework for, 115
strategic planning, benefits of, 114–15
pursuit, 125
transition:
action plans, 119
external requirements, 119
implementation, 118
internal requirements, 119
investment/return, 118–19
recommendation, 119
requirements, 119
segment strategy, 117–18
Oshonsoft PIC Simulator, 406, 409
Ovako working posture analyzing system (OWAS), 252
Overheads, 300
Oxifluorination, 553

P

Packard, Hewlett, 78
Paconius, 21
Parabolic light, 289
Part reliability, 313
PBTP bumper skin, 584
Penalties, 381, 382
apportionment of costs during guarantee, 381
payment according to downtime, 381
Percentile, 243
Perception, 238
Performance specifications, 34
Permanent hearing loss, 284–5
Personal analogies, 37
Personal factors, 233
case study, 245
data collected, 245
method, 245
physical capabilities, 241
anthropometry, 242–5
muscular strength and endurance, 241–2
sensory and cognitive capabilities, 233
auditory sense, 237
cognitive capabilities, 237–40
information processing, summary of, 241
visual sense, 233–7
Phacomp, 216
Phase one matrix, *see* House of quality (HOQ) matrix
Phoenix Bridge Company, 23
Physical capabilities, 241
anthropometry, 242–5
muscular strength and endurance, 241–2
Physical factors, 249
behavior, 259–60
case study, 263–4
ergonomics program, 260–3
hand-intensive repetitive tasks, 257–9
manual handling screening checklist, 264–9
manual handling tasks, 250–6
musculoskeletal disorders (MSDs), 249–50
musculoskeletal ergonomics program gap analysis checklist, 269–71
PIC microcontroller, 401
architecture, 401
device, choosing, 402–3
families, 401–2
flowcode, 406
PIC16F84, features of, 403–4
PIC16F877, features of, 404–5
practical investigation, 406–13
programming, 405–6
software development cycle, 409
PIC16C/F84 instruction set, 410–11
PIC16C54, 402
PIC16C55, 402
PIC16C56, 402
PIC16F84, features of, 403–4
PIC16F877, features of, 404–5
PIC16F877A microcontroller, investigating, 406–13
PIC17C42, 402
Pilot lighting system, installation of, 289–90
Pipe, designing, 538
RP pipe, 539–41
thermoplastic pipe, 538–9
Planguage methods:
and design engineering process, 60–1
Plastic lumber, 575
Plastic plate, designing, 510–12
Plastic products, 584
Plastics design techniques, 483–5
behaviors, 494
morphology/molecular structure/mechanical property, 494–5
Newtonian/non-Newtonian, 495–6
rheology and mechanical analysis, 496
viscosities and melt flows, 495
construction materials, 485–8
commodity and engineering plastics, 493
cross-linked thermoplastics, 490–1
elastomers/rubbers, 493–4
reinforced plastics, 491–3
thermoplastics, 488–90
thermosets, 490
fabricating processes, 499–500
extrusions, 500
injection moldings, 500–2
thermoformings, 502–3
FALLO approach, 506
foams, 503
processing-to-performance interface, 496
drying operations, 498–9
glass transition temperatures, 496–7
melt temperatures, 497–8
processing and moisture, 498
reinforced plastics, 504–6
variables, 506
Plastics design technology, 507
automobile, 563–7
beam, 512–14
bearing, 526
building, 554–5
column, 517
contact stress, 526
cushioning, 553–4
cylinder, 530
design, 520–3
design limitation and constraint, 581–2
EI theory, 509–10
electrical/electronic, 567–70
Euler's formula, 517–18
furniture, 573–5
gasket, 527–8
gear, 523–6
geometric shape, 509
grommet, 527
hinge, 548
house of future, 555–6
house top, 556–7
lumber, 575–81
medical, 570
dental product, 571–2
health care, 572
surgical product, 571
missile and rocket, 567
modulus of elasticity, 509
monocoque structure, 509
packaging, 551–2
permeability, 552–3
pipe, 538–41
plate, 510–12
press fit, 548
PV factor, 526–7
recreation, 572–3
reinforced plastic (RP), 508–9

rib, 514–16
sandwich structure, 519–20
shape, 528–9
snap fit, 548–50
sphere, 530–1
spring, 541–8
tank, 531–8
tape, 550–1
torsion, 518–19
transportation, 557–63
water filter, 575
Platform management, concept of, 145–6
PMMAs, 571–2
Pneumatic tire, 36
Poisson's ratio, 541
Poka Yoke, 336
Polar charts, 584
Pole-zero analysis, 393–4
Polyelectrolytes plastics, 571
Polyethylene (PE), 490
Polymer selection:
car body, tubular frames of, 585–6
car door skin, 585
constant cross section, beam torsion of, 589–91
non-constant cross-section, beam torsion of, 591–3
property combinations, 584
ribs
buckling of, 588–9
on injection mouldings, 586–8
selection packages, 583–4
shape selection to optimise stiffness, 586
Polymers, 193, 201
Polytetrafluoroethylene (PTFE), 527
Portable electronic devices in aircraft, 416–17
Portfolio management, 11
Postformings, 500
Post-it note, 37
Potato harvesting machine:
function analysis and structure for, 38
morphological chart for, 40
Potential design ideas, 59
Power Line Communications (PLC), 424
Power Line Telecommunication (PLT), 424
compliance status, 426
cumulative effects, 426
floodgates, opening, 426
HF spectrum, use of, 425
interference capability, 425
and interferers, 425–6
politics, 424–5
Press fit, 548
Prevention costs, 300
Price per unit weight diagram:
for materials, 199
for products, 200
Prickly heat, *see* Heat rash
Primary creep, 224
Primary market research, 129–30
Priority determination, 70, 71
planguage information supports, 71
role of resource in, 71
subjective weighting approach for, 70–1
Priority strategy, 71
Probabilistic design techniques, 314
hole in plate analogy of, 315
versus deterministic design, 316
PRocess Information MAps (PRIMAs), 336
Process integrity, 298–9
Process of design engineering:
beginning, 61–2
diagrams/icons, 72–3
and planguage methods, 60–1
principles, 69–70
process description, 65
entry conditions, 65
exit conditions, 69
procedure, 66–9
Product character, 188–9, 190
Product concept synthesis:
business concept embodied on product idea, 139–41
customer's problem, solving, 141–2
idea evaluation within business framework, 142–4
Product cost, 4
Product design specification (PDS), 461–4
Product development:
disposition of rework in, 296
external means to, 160–7
brand label, joint venture, and acquisitions, model for, 160–3
comparative analysis, with industry averages, 164–7
valuation of company, 163–4
valuations with risk analysis, structuring, 164
origination and elimination of faults in, 296
Product development, successful, 3
attributes of, 4–5
key factors, 5
customer focus and market orientation, 5
execution of activities, 5–6
experience and core competencies, 7
good product, 6
homework, 5
innovativeness, 7–8
market attractiveness, 7
nature of benefits, 8
order of entry, 7–8
organizational structure and climate, 6
project selection decisions, 6
sharp and early product definition, 5
speed without compromising quality, 6–7
systematic new product process, availability of, 7
top management, role of, 6
uniqueness, 5
new business activity, progression of actions in, 3
strategy for, 8
company's growth expectations from new products, determination of, 8
existing opportunities, determining, 9, 10
new product options list, development of, 9
new product plans, development of, 11–15
portfolio management, 11
product inclusion in portfolio, setting criteria for, 9–11
product portfolio, creating, 11
strategic information, gathering, 8–9
Product failure:
mechanisms, 304–7
risk as basis for design, 308–10
and variability, link between, 307–8
Product inclusion in portfolio, setting criteria for, 9–11
Product integration into business:
business development game
playing, 145
winning, 145–7
flow chart, 148–51
new product, 148
profitability, 144–5
Product liability and safety legislation, 385
Consumer Protection Act 1987, 386
general situation, 385, 386
Common law, 385
Contract law, 385
Statute law, 385
Health And Safety at Work Act 1974, 386–7
insurance and product recall, 387–8
strict liability, 386
Product liability and service provision, 201–2
Product lines, additions to, 105

Product per forma, 14
Product plan:
  for blender, 14
  for rechargeable electric toothbrush, 14
Product portfolio, creating, 11
Product specifications:
  for rechargeable electric toothbrush, 13
Product tolerances, 297
Product use-patterns, matrix of, 175
Production energy, 176
Program evaluation and review technique (PERT), 14
Project management:
  functional safety capability, 374
  planning, feasibility and allocation, 371–2
  programme activities, 372–3
  responsibilities, 373–4
  setting objectives and specifications, 371
  standards and guidance documents, 374–5
Property combinations, for material selection, 584
Property detractors, 581
Proprioceptors, 233
PV factor, 526–7

Q

Q10 effect, 279
Quacking duck, 419
Quality, of product, 43
  design review case study, 53–4
  high quality in design stages, incorporating, 45
    design for six sigma (DFSS), 45–6
    design review, 49–53
    mistake proofing (Poka-Yoke), 46–7
    quality function deployment, 47–9
  quality control, need for, 43–4
  reactive versus proactive quality control, 44–5
  six sigma case study, 55–6
Quality and reliability, 295
  costs of quality, 300
    failure costs, cases studies in, 301–3
    quality-cost estimating methods, 303–4
  designing, 310–15
  product failure:
    and variability, link between, 307–8
    mechanisms, 304–7
    risk, as basis for design, 308–10
  statement of problem, 295–300
Quality assurance versus quality control, 44
Quality control, need for, 43–4
Quality-cost models, 303–4, 365
Quality-cost scaling, *see* Quality Loss Function
Quality dimensions, 44
Quality failure costs, 300–1
Quality function deployment, 47–9
Quality Function Deployment (QFD), 310
Quality Loss Function, 304
Quality loss versus product price, 44
Quebec Bridge:
  collapsed, 24
  under construction, 24
Quebec Bridge Company, 23

R

Radiation, 278
Radio-frequency interference (RFI), 567, 568, 569
Radome, 570
RAM measurement, levels of, 372
RAMS cycle, 373
Random stimulus method, 38
Random vibration, 286
Rapid entire body assessment (REBA), 252
Rapid solidification processing (RSP) alloys, 229–30
Rapid upper limb assessment (RULA) method, 258
Reaction injection moldings (RIM), 504–5, 521
Reactive versus proactive quality control, 44–5
Realistic reliability, maintainability and safety (RAMS), 371
Rechargeable electric toothbrush:
  customer needs statement for, 12
  product plan for, 14
  product specifications for, 13
Recycling, 584
Reflected/specular glare, *see* Indirect glare
Regular vibration, 286
Reinforced/foamed plastic, 515–16
Reinforced plastic (RP), 486, 491–2, 504–6, 508–9
  bulk molding compounds, 493
  compounds, 492
  ductilities, 492
  prepregs, 492–3
  sheet molding compounds, 493
  thermal expansions, 492
  tolerances/shrinkages, 492
  toughness, 492
Reinforced reaction rejection molding (RRIM), 520
Reinforced thermoplastics (RTPs), 491, 504
Reinforced thermoset plastics (RTSs), 491, 504
Reinforced thermoset resin (RTR), 539
Reliability and maintainability, 377, 378
  analysis, 379–80
  programme, 379
Reliability and quality, *see* Quality and reliability
Reliability prediction techniques, 313
Repetitive strain injury (RSI), 249
Repositioning, 106
Requirement specification, 60
Requirements problem, solving, 57
  alternative design ideas, need for, 59
    choosing best combined set from all alternatives, 59
    choosing best from alternatives, 59
    design optimization, 60
    use of alternative design, reducing risk by, 59
  design and design engineering, difference between, 57
  design engineering process, beginning, 61–2
  design ideas, 58
    design constraints, 59
    potential design ideas, 59
  design specification template, 74–5
  design specifications masquerading as requirements, case study, 72
  detailed requirement specification, 57–8
  diagrams/icons, 72–3
  dynamic priority evaluation, 71–2
  language core, 63
  planguage information supports priority determination, 71
  planguage methods and design engineering process, 60–1
  principles, 69–70
  priority determination, 70
    role of resource in, 71
    subjective weighting approach for, 70–1
  priority strategy, 71
  process description, 65–9
  requirement specification, design engineering and Evo:
    iterative processes, 57
  rules, 63–5
Requirements pyramid, 187–8

Research and development (R&D), relationship between, 109–12
  fundamental type, 110
  incremental research, 110
  radical type, 110
Resin matrix material system, 535
Rheology, 496
Rib, designing:
  plastic, 514–15
  reinforced/foamed plastic, 515–16
Risk as basis for design, 308–10
Robust design, 52, 311, 312
  parameter design, 52–3
  system design, 52
  tolerance design, 53
Ronan Point construction, 22, 23
Rotational moldings (RM), 505–6
Royal Navy, 26
RP leaf springs, 542, 546
RP spring model, 543

S

S Factor analysis, 53
S product planning matrix, 47–8
Safe lifting task, 251–2
Safety critical failures, potential cost of, 303
Safety critical systems, 202
Scale effect, 23–4
Science-push, 197
Scribe, 86
Seals, 476–7
Second EMCD, changes in, 440
Second-order transition, 497
Secondary market research, 130–1
  articles, 131
  financial reports, 131
  lost sales reports, 131
  news pieces, 131
  observation of trends, 131
  patent search, 131
  textbooks, 131
  trade literature search, 131
  trade shows, 130–1
Segmental vibration, 287–8
Segmentation research, 11
Selective attention, 238
Self certification and internal production control, 444–6
Senses and materials, *see* Aesthetic attributes
Sensitivity analysis, 395
Sensory and cognitive capabilities, 233
  auditory sense, 237
  cognitive capabilities, 237–40
  information processing, summary of, 241
  visual sense, 233–7
Sensory information, 237
Sensory memory (SM), 238–9
Shape-memory effect (SME), 227
Shape of products, 528
  filament winding (FW) shapes, 528
  netting analysis, 528–9
Sheet molding compounds (SMCs), 493
Shielding-effectiveness (SE), 568
Shivering, 280
Short-term and long-term trade-offs, benefits of, 168–9
Short-term memory (STM), 239
Sight, 191–2
Silicon spheroidal graphite (SG) iron, 216
Silver, Spencer, 37
Simplified design decision matrix, 40
"Six degrees of separation", 146–7
Six sigma case study, 55–6
Skin packaging, 551–2
Skylab program, 109
SLIM (Simpler Legislation for the Internal Market), 433
Small-signal distortion analysis, 394–5
Snap fit, 548–50
Soft-hard (SH) review, 51
Solenoid end assembly, case study, 357–63
Solenoid security cover, case study, 355
Sound, 282
Source document, 81
Specific gravity, 494
Specification quality control (SQC), 77
  case study, 93–4
  continuous process improvement, 99–100
  diagrams/icons, 94
  document quality, understanding, 94–9
  economics of, 91
  effectiveness and estimating remaining defects, determining, 91–2
  effectiveness of, 91
  entry conditions, 86
  example, 78–80
  exit conditions, 88
  extending from SQC into specification review, 92–3
  finding defects, 94
  improving process, 77
  increasing productivity, 78
  industrial usage, 78
  language core:
    checking rate, 80
    checklist, 81
    defect, 80
    Defect Detection Process (DDP), 80
    Defect Prevention Process (DPP), 80
    issue, 80
    kin documents, 81
    logical page, 80
    main specification, 81
    major defect, 80
    optimum checking rate, 80–1
    remaining major defects, 81
    rule, 81
    source document, 81
  principles, 90–1
  purposes for using, 100
  and rules, 92
  saving time, 78
  simplified SQC:
    form, 99
    process, 88–90
  standards, 81
    guidelines for assessing overall functioning of SQC, 81–4
    process roles and responsibilities, 84–6
  sub-processes, 86–8
    checking, 87
    edit audit, 87
    Editor Advice Log, 87
    entry, 86
    exit, 87–8
    kickoff, 87
    planning, 86
    process meeting, 87
    specification meeting, 87
    SQC statistics, 87
    SQC strategy, 86
Speech and communication, 283
Sphere, designing, 530–1
SPICE (Simulation Program with Integrated Circuit Emphasis), 392, 397–400
Springs, 477–8
  designing, 541
    cantilever spring, 547
    leaf spring, 542–6
    special spring, 546–7
    torsional beam spring, 547–8
Stainless steels:
  compositions and properties of, 214
Stampable reinforced thermoplastic sheet (STX), 576, 579, 580
Standards check point design review, 49
Static anthropometry, *see* Structural anthropometry
Static dimensions, *see* Structural anthropometry
Static effort, *see* Static work
Static work, 241, 242
Statistical maintainability test, 380–1
Statistical Process Control (SPC), 296

Statute law, 385
Statutory Instruments (SIs), 432
Steel Tank Institute, 532
Steels, commercial:
  alloy steels, 208–9
  dual-phase (DP) steels, 210–12
  high-strength low-alloy (HSLA) steels, 209–10
  maraging steels, 209
  mechanically alloyed (MA) steels, 212–12
  plain carbon steels, 207–8
  steels, designation of, 212
Steels, designation of, 212
Strategic development activities, 11
Strength design limit (SDL), 256
Strength upper limit (SUL), 256
Stress–strain (S–S) analysis, 540
Strict liability, 386
Strip hybrids, 559
Structural anthropometry, 243
Structural foam (SF), 521, 522
Structure, of design process, 19
  changing design process, 20–1
  defining problem and setting objectives, 29–33
  design paradigms, 21–7
  developing provisional designs, 35
    analogies and chance, 36–8
    analytic methods, 38–9
    brainstorming, 36
  engineering design process, steps of, 28–9, 30, 31
  evaluation and decision-making, 39–41
  function analysis, 33–5
  performance specifications, 34
  problem confronting designers, 27–8
  requirements, 27
    setting, 34
Style, 195
Superalloys:
  basic alloying features, 215–17
  dispersion-hardened superalloys, 218
  nickel-based superalloy development, 217–18
Superglue/Krazy Glue, *see* Cyanoacrylate
Superplastic aluminum alloys, 229
Surface engineering, 201
Surface roughness chart, 330
Survey and observation tools, 258
Surveys, 129
Sustainability, design for, 173
Sustained attention, 238
Swing solutions, 58
Symbolic analogies, 37
Synectics, 37
System effectiveness, 377
System reliability, 313
Systematic methods, 38
Szlapaka, P. L., 23

T

Tacoma Narrows Bridge, 26
Taguchi's method, 52
Tank, designing:
  corrosive resistant tank, 537–8
  fabricating RP tank, 531–2
  highway tank, 537
  hopper rail car tank, 534–7
  underground storage tank, 532–4
  very large tank, 537
Task lights, 289
Team leader, 84–5
Technical Construction File (TCF), 433, 445, 447
Telescopic lever assembly, case study, 355–7
Temperature, 278
  comfort and discomfort zones, 280–1
  effects of cold on performance, 279
  effects of heat:
    on health, 279–80
    on performance, 278–9
  hot environment, work tolerance in, 281–2
  recommendations to improve working conditions, 282
Temporary hearing loss, 284
Tesco, 203
Thermal analysis, 395
Thermoformings, 502–3
Thermoplastics (TPs), 486, 487, 488–90
Thermoset plastics (TSs), 486, 487, 490
Thermostats, 418–19
Tin can method, 569
Titanium alloys:
  basic alloying and heat-treatment features, 219–20, 221
  commercial titanium alloys, 220–2
  processing of, 222
Titanium aluminides, 224–5
Torsion, 518–19
Torsional beam spring, 547–8
Touch, 190–1
Transient analysis, 393
Tree diagram, 55, 56
Tribology, 474
Triggers, 262
TS polyurethane (PUR) plastics, 504–5
Tunnel vision, avoiding, 26–7
Twentieth-century engineering drawing, 21

U

UK Ministry of Defence Standard–40 reliability and maintainability, 375
UK motorcycle industry, 302
Undesired gas emissions, 176
Unidirectional composite (UDC), 559
Unwanted sound, 282
US Military Standard 470 Maintainability Programme Requirements, 375
User-interface design review, 49

V

Vacuum casting, 217
Vaporization, 177
Variability, 299
  and failure, link between, 307–8
  in manufacturing, 320–1
Variability risk table, 348, 349
  for fixing bolt assembly, 351
Variables, 506
Vasoconstriction, 280
Verein Deutscher Ingenieure (VDI), 29
Vibration, 286
  effects:
    on health, 287–8
    on performance, 287
  guidelines to reduce or control vibration, 288
Vigilance, *see* Sustained attention
Viscosity of plastics, 495
Visibility versus visual acuity, 235–6
Visual acuity, 235–6
Visual field, 234
Visual performance, effects of vibration on, 287
Visual sense, 233
  accommodation of eye, 234
  adaptation process, 234–5
  age, 236–7
  color vision, 235
  visual acuity, 235–6
  visual field, 234
Vulcanized rubber, 36

W

Waspaloy, 217
Wavelength of light, 273
White-heart malleable iron, 214
White irons, 214
Whole-body vibration, 287
Wireless Telegraphy Act 1949, 432
Wood, 193
Working memory, *see* Short-term memory (STM)
Writer, 86

Z

Z-Air System, 53, 54