7 Embodiment Design

Embodiment design is the part of the design process in which, starting from the principle solution or concept of a technical product, the design is developed in accordance with technical and economic criteria and in the light of further information, to the point where subsequent detail design can lead directly to production (see Section 4.2).

The draft guideline VDI 2223: *Systematic Embodiment of Technical Products* [7.295] builds on recommendations from the fourth German edition of this book along with other sources. In doing so, it presents a generally established systematic procedure for embodiment design.

7.1 Steps of Embodiment Design

Having elaborated the principle solution during the conceptual phase, the underlying ideas can now be firmed up. During the embodiment phase at the latest, designers must determine the overall layout design (general arrangement and spatial compatibility), the preliminary form designs (component shapes and materials) and the production processes, and provide solutions for any auxiliary functions. During all of this, technological and economic considerations are of paramount importance. The design is developed with the help of scale drawings, critically reviewed, and subjected to a technical and economic evaluation.

In many cases several embodiment designs are needed before a definitive design appropriate to the desired solution can emerge.

In other words, the *definitive layout* must be developed to the point where a clear check of function, durability, production, assembly, operation and costs can be carried out. Only when this has been done is it possible to prepare the final production documents.

Unlike conceptual design, embodiment design involves a large number of corrective steps in which analysis and synthesis constantly alternate and complement each other. This explains why the familiar methods underlying the search for solutions and evaluation must be complemented with methods facilitating the identification of errors (design faults) and optimisation. The collection of information on materials, production processes, repeat parts and standards involves considerable effort.

The embodiment process is complex in that:

- many actions must be performed simultaneously
- several steps must be repeated at a higher level of information
- additions and alterations in one area have repercussions on the existing design in other areas.

Because of this, it is not always possible to draw up a strict plan for the embodiment design phase. However, it is possible to suggest a general approach with main working steps, see Figure 7.1. Particular problems may demand deviations and subsidiary steps, which can rarely be predicted precisely. The approach has to be planned to match the problem at hand, realising that further modifications will have to be made. Basically, the process will proceed from the qualitative to the quantitative, from the abstract to the concrete, and from rough to detailed designs. It is important to make provision for checks and, if necessary, for corrections.

1. Starting with the principle solution, and using the requirements list, the first step is to identify those *requirements that have a crucial bearing* on the embodiment design:
   - size-determining requirements, such as output, throughput, size of connectors, etc.
   - arrangement-determining requirements, such as direction of flow, motion, position, etc.
   - material-determining requirements, such as resistance to corrosion, service life, specified materials, etc.

   Requirements such as those based on safety, ergonomics, production, assembly and recycling involve special embodiment considerations, which may affect the size, arrangement, and selection of materials (see Sections 7.2 to 7.5).
2. Next, the *spatial constraints* determining or restricting the embodiment design must be identified (for instance clearances, axle positions, installation requirements, etc.).
3. Once the embodiment-determining requirements and spatial constraints have been established, a rough layout, derived from the concept, is produced with the emphasis on the overall embodiment-determining *main function carriers*, that is, the assemblies and components fulfilling the main functions. The following subsidiary questions must be settled, with due regard paid to the principles of embodiment design (see Section 7.4):
   - Which main functions and function carriers determine the size, arrangement and component shapes of the overall layout (for instance, the blade profiles in turbomachines or the flow area of valves)?
   - Which main functions must be fulfilled by which function carriers jointly or separately (for instance, transmitting torque and allowing for radial movement by means of a flexible shaft or by means of a stiff shaft plus a special coupling)? This step is similar to division into realisable modules, as shown in Figure 1.9.

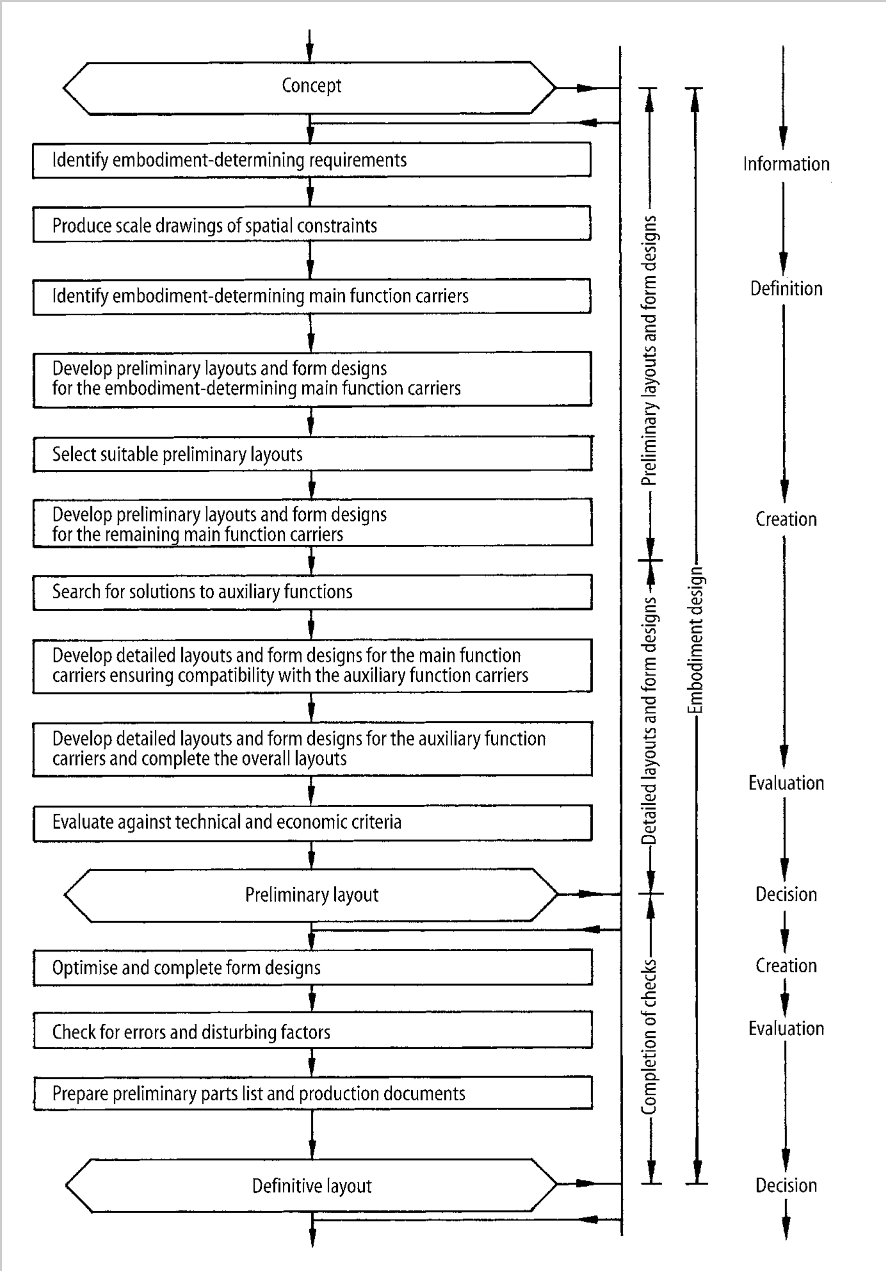


Figure 7.1. Steps of embodiment design

4. Preliminary scale layouts and form designs for the embodiment-determining main function carriers must be developed; that is, the general arrangement, component shapes and materials must be determined provisionally. To that end, it is advisable to work systematically through the items under the heading "layout" in the checklist shown in Figure 7.3. The result must meet the overall spatial constraints and then be completed so that all of the relevant main functions are fulfilled (for instance by specifying the minimum diameters of drive shafts, provisional gear ratios, minimum wall thicknesses, etc.). Known solutions or existing components (repeat parts, standard parts, etc.) must be shown in simplified form. It may be useful to start working on selected areas only, combining these into preliminary layouts later.
5. One or more suitable *preliminary layouts* must be selected in accordance with the procedure described in Section 3.3.1 (modified if necessary) by considering the relevant items in the checklist shown in Figure 7.3.
6. Preliminary layouts and form designs must now be developed for the remaining main function carriers that have not yet been considered because known solutions exist for them or they are not embodiment-determining until this stage.
7. Next, determine which essential *auxiliary functions* (such as support, retention, sealing and cooling) are needed and, where possible, *exploit known solutions* (such as repeat parts, standard parts, catalogue solutions). If this proves impossible, *search for special solutions*, using the procedures already described in Section 3.2 and Chapter 6.
8. *Detailed layouts and form designs for the main function carriers* must now be developed in accordance with the embodiment design rules and guidelines (see Sections 7.3 to 7.5), paying due attention to standards, regulations, detailed calculations and experimental findings, and also to the problem of compatibility with those auxiliary functions that have been realised. If necessary, divide into assemblies or areas that can be elaborated individually.
9. Proceed to develop the *detailed layouts and form designs for the auxiliary function carriers*, adding standard and bought-out parts. If necessary, refine the design of the main function carriers and combine all function carriers into overall layouts.
10. *Evaluate* the layouts against technical and economic criteria (see Section 3.2.2). If a particular project requires several concepts to be put in more concrete form prior to evaluation, then the embodiment process must not, of course, be pursued beyond what the evaluation of the variants demands. Depending on the circumstances, it is thus possible, in some cases, to take a decision just as soon as the main function carriers have reached the preliminary layout stage, while in other cases the decision will have to be deferred until after a great deal of detail design. In either event, all of the designs to be compared must be at the same level of embodiment, since no reliable evaluation is possible otherwise.

11. Fix the preliminary overall layout. The overall layout describes the complete construction structure (see Figure 2.13) of the system or product being designed.

12. Optimise and complete the form designs for the selected layout by *eliminating the weak spots* that have been identified during the course of the evaluation. If it should prove advantageous, repeat the previous steps and adopt suitable subsolutions from less favoured variants.

13. *Check* this layout design for *errors* (design faults) in function, spatial compatibility, etc. (see Figure 7.3), and for the effects of *disturbing factors*. Make what improvements may be needed. The achievement of the objectives with respect to cost (see Chapter 11) and quality (see Chapter 10) must be established at this point at the latest.

14. Conclude the embodiment design phase by preparing a preliminary *parts list* as well as a preliminary *production and assembly documents*.

15. Fix the *definitive layout* and pass on to the detail design phase.

The representation of the spatial constraints and the embodiment is now generally obtained by creating a full 3-D digital model. Irrespective of whether a 2-D or 3-D representation is used [7.213]:

- the function and type of the objects must be shown
- the positions of and the necessary space for the objects must be recognisable through characteristic dimensions, e.g. the overall dimensions, which can be used to check the overall spatial compatibility and assembly operations.

When 2-D CAD systems or drawing boards are still used simplifications, such as those proposed by Lüpertz [7.174], could be used.

In the embodiment phase, unlike the conceptual phase, it is not necessary to lay down special methods for every individual step, however the following recommendations might prove useful.

The *search for solutions* for auxiliary functions and other subsidiary problems should be based either on the procedure described in Chapter 3, but simplified as far as possible, or else directly on catalogues. Requirements, functions and solutions with appropriate classifying criteria have already been elaborated.

The *embodiment* (layout and form designs) of the function carriers should be based on the checklist (see Figure 7.3) and involves reference to the principles of mechanics and structures, and to materials technology. It calls for calculations ranging from the simplest through to complex differential equations and finite element analyses. For these calculations, the reader is referred to the literature listed in Section 7.5.1, and for even more complex calculations to the domain specific literature. In some cases it might be necessary to build prototypes or to undertake specific tests.

In the elaboration of embodiment designs, many details have to be clarified, confirmed and optimised. The more closely they are examined, the more obvious it becomes as to whether the right solution concept has been chosen. It may appear that this or that requirement cannot be met, or that certain characteristics of the chosen concept are unsuitable. If this is discovered during the embodiment phase, it is advisable to re-examine the procedure adopted in the conceptual phase, for no embodiment design, however perfect, can hope to correct a poor concept. This is equally true of the working principles applicable to the various subfunctions. However, even the most promising concept can cause difficulties in embodiment and detail design. This often happens because various features were originally treated as subordinate or as not in need of further clarification. Attempts to solve these subproblems compel designers to reiterate the appropriate steps while retaining the selected working structure and overall arrangement.

Experience with the proposed approach for embodiment design has confirmed its basic validity, but has also revealed the following important points [7.211]:

- If prior research has been undertaken or embodiment variants already exist, the step of producing preliminary embodiments can often be left out.
- Preliminary embodiments can always be left out when only detailed improvements are required.
- The solutions for auxiliary functions usually influence the preliminary embodiment of the main function carriers, so working on these solutions must not be left until too late in the process.
- A characteristic of successful designers is that they continuously check and monitor their actions to identify direct and indirect effects.

Many products are not developed from scratch, but are developments or improvements of existing ones that take into account new requirements, new knowledge and experiences. Experience has shown that it is useful to start by analysing the failures and disturbing factors for an existing solution (see Sections 10.2 and 10.3) and, based on that analysis, to develop a new requirements list (see Figure 7.2). The result of the clarified task will show whether a new working structure—a new principle solution—is required, or whether it is sufficient to modify the existing embodiment. It is possible to start at many different places within the overall approach. In some cases a new product can be produced by making improvements to the details. In other cases, tests of the existing or modified modules may be necessary. The required steps in the overall approach must be selected appropriately.

To sum up, embodiment design involves a flexible approach with many iterations and changes of focus. The individual steps have to be selected and adapted to the particular situation. The ability to organise one's own approach while paying due regard to the fundamental links between the steps and the recommendations we provide is important (see Section 2.2.1).

In embodiment design, the rules and principles elaborated in Sections 7.2 to 7.5 should be followed. Because of the fundamental importance of the identification of errors (design faults) in several of the steps, the reader is referred to Chapter 10 in particular.

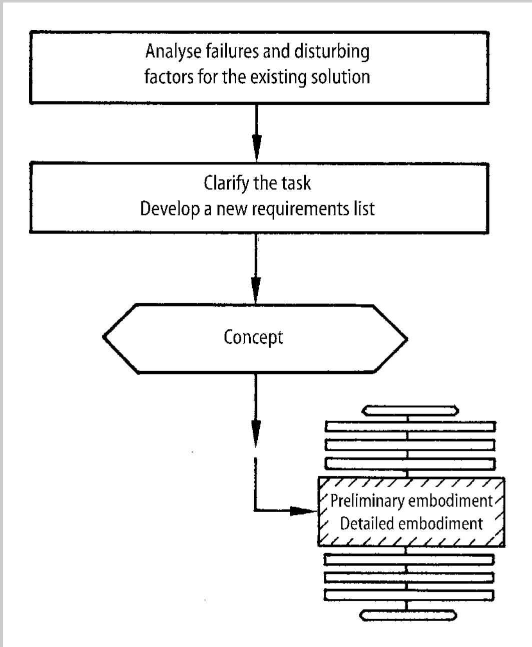


Figure 7.2. Embodiment design phase based on the development of an existing solution. Which of the steps shown in Figure 7.1 needs to be completed follows from an analysis of failures and disturbing factors

7.2 Checklist for Embodiment Design

Embodiment design is characterised by *repeated deliberation and verification* (see Section 7.1). Every embodiment design is an attempt to fulfil a given function with appropriate layout, component shapes and materials. The process starts with preliminary scale layouts based on a rough analysis of spatial requirements, and proceeds to consider safety, ergonomics, production, assembly, operation, maintenance, recycling, costs and schedules.

In dealing with these factors, designers will discover a large number of interrelationships, so that their approach must be progressive as well as iterative (verification and correction). Notwithstanding this double character, however, the approach must always be such as to allow the speedy identification of those problems that must be solved first.

The checklist shown in Figure 7.3 has been derived from the general objectives and constraints discussed in Section 2.1.7. Although the factors are interrelated, this checklist presents them in a useful procedural order and provides designers with a systematic check on each one. The checklist thus not only provides a strong mental impetus, but also ensures that nothing essential is forgotten.

All in all, continuous reference to the headings will help designers to develop and test their progress in a systematic and time-saving way. Each heading should be examined in turn, regardless of its interrelationship with the rest.

Headings	Examples
Function	Is the stipulated function fulfilled? What auxiliary functions are needed?
Working principle	Do the chosen working principles produce the desired effects and advantages? What disturbing factors may be expected?
Layout	Do the chosen overall layout, component shapes, materials and dimensions provide: adequate durability (strength) permissible deformation (stiffness) adequate stability freedom from resonance unimpeded expansion acceptable corrosion and wear with the stipulated service life and loads?
Safety	Have all the factors affecting then safety of the components, of the function, of the operation and of the environment been taken into account?
Ergonomics	Have the human–machine relationships been taken into account? Have unnecessary human stress or injurious factors been avoided? Has attention been paid to aesthetics?
Production	Has there been a technological and economic analysis of the production processes?
Quality control	Can the necessary checks be applied during and after production or at any other required time, and have they been specified?
Assembly	Can all the internal and external assembly processes be performed simply and in the correct order?
Transportt	Have the internal and external transport conditions and risks been examined and taken into account?
Operation	Have all the factors influencing the operation, such as noise, vibration, handling, etc. been considered?
Maintenance	Can maintenance, inspection and overhaul be easily performed and checked?
Recycling	Can the product be reused or recycled?
Costs	Have the stipulated cost limits been observed? Will additional operational or subsidiary costs arise?
Schedules	Can the delivery dates be met? Are there design modifications that might improve the delivery situation?

Figure 7.3. Checklist for embodiment design

The actual sequence is no indication of the relative importance of the various headings, but ensures a systematic approach. For instance, it would be futile to deal with assembly problems before ascertaining if the required performance or minimum durability is ensured. The checklist thus provides a consistent scrutiny of embodiment design and one that is easily memorised.

7.3 Basic Rules of Embodiment Design

The following basic rules apply to all embodiment designs. If they are ignored problems are introduced and breakdowns or accidents may occur. They underlie nearly all of the steps listed in Section 7.1. When used in conjunction with the checklist (see Figure 7.3) and with the design fault identification methods (see Chapter 10), they also provide essential assistance with selection and evaluation.

The *basic rules* of clarity, simplicity and safety are derived from the general objectives set out in Section 2.1.7, that is:

- fulfilment of the technical function
- economic feasibility
- individual and environmental safety.

The literature contains numerous rules of, and guidelines for, embodiment design [7.168, 7.180, 7.198, 7.205]. On closer analysis it appears that clarity, simplicity and safety are fundamental to all of them and are important prerequisites for a successful solution.

Clarity—that is, clarity of function or lack of ambiguity of a design—facilitates reliable prediction of the performance of the final product and in many cases saves time and costly analyses.

Simplicity generally guarantees economic feasibility. A smaller number of components and simple shapes are produced more quickly and easily.

Safety imposes a consistent approach to the problems of strength, reliability, accident prevention and protection of the environment.

In short, by observing these three basic rules, designers can increase their chances of success because they focus attention on, and help to combine, functional efficiency, economy and safety. Without this combination no satisfactory solution is likely to emerge.

7.3.1 Clarity

In what follows we shall be applying the basic rule of clarity to the various headings of the checklist in Figure 7.3.

Function

Within a given function structure, an unambiguous interrelationship between the various subfunctions and the appropriate inputs and outputs must be guaranteed.

Working Principle

The chosen working principle, in terms of the physical effects, must reveal a clear relationship between cause and effect, thus ensuring an appropriate and economical layout.

The chosen working structure, comprising several individual working principles, must guarantee an orderly flow of energy, material and signals. If it does not, undesirable and unpredictable effects such as excessive forces, deformations and wear may ensue.

By paying attention to the deformations associated with a given loading, and also to thermal expansion, designers must make the necessary allowances for possible expansion in a given direction.

The widely used bearing pairs, with a locating and a nonlocating bearing (see Figure 7.4a) have a clearly defined behaviour. The stepped bearing pair (see Figure 7.4b), on the other hand, should be specified only when the expected changes

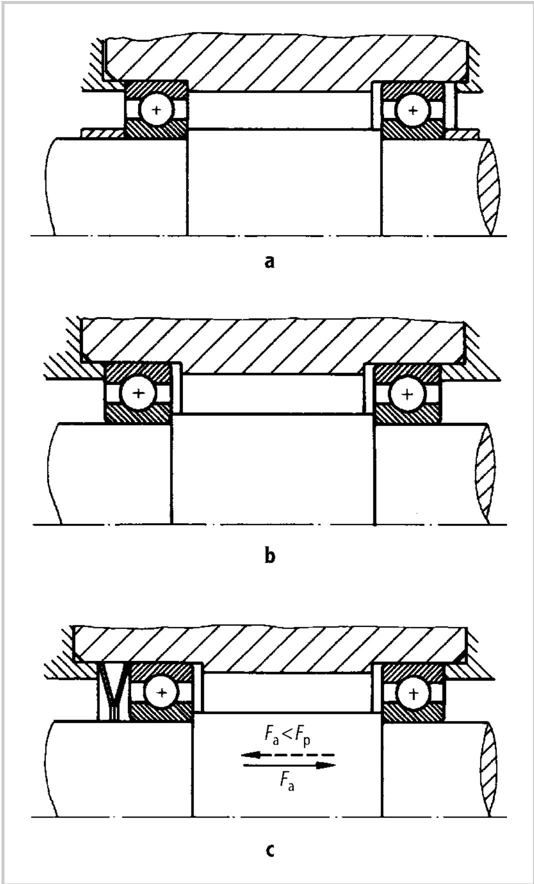


Figure 7.4. Basic bearing arrangements: **a** Locating and nonlocating arrangement: left-hand locating bearing takes up all the axial forces, right-hand sliding bearings permit unimpeded axial movement due to thermal expansion; accurate calculations are possible. **b** Stepped bearing arrangement: the axial loading of the bearings depends on the preload and thermal expansion and cannot be clearly determined; a modification is the "floating arrangement" in which the bearings are provided with axial clearance; in that case, thermal expansion is possible to a limited extent but there is no precise shaft location. **c** Spring-loaded bearing arrangement: here the disadvantages of the stepped bearing arrangement are largely eliminated, though the constantly applied axial load may reduce the bearing life; forces resulting from thermal expansion can be determined by spring force deflection diagrams; the shaft is located precisely provided the axial force F_a acts only towards the right or does not exceed the preloading F_p

in length are negligible or when the resulting play in the bearings is permissible. By contrast, a spring-loaded arrangement, in which the operating axial force F_a does not exceed the pre-load F_p, will permit a clear definition of the force transmission path (see Figure 7.4c).

Combined bearing arrangements often present problems. The combination shown in Figure 7.5a consists of a needle roller bearing which is intended to transmit the radial forces and a ball bearing which is meant to transmit the axial forces. However, this particular arrangement does not clearly define the transmission path for the radial forces, because the inner and outer races of both bearings

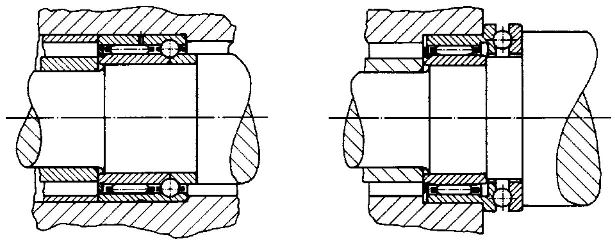


Figure 7.5. Combined rolling-element bearing. **a** Transmission path of radial forces not clear; **b** combined rolling bearing with the same elements as in **a**, but clear identification of the transmission paths of the radial and axial forces

are restrained radially. As a result, the service life cannot be predicted accurately. The arrangement shown in Figure 7.5b, on the other hand, satisfies the clarity rule with similar elements, provided the designer ensures during assembly that the right-hand race has enough radial play, thus making certain that the ball bearing transmits axial forces only.

Double fits conflict with the basic rule of clarity. These occur when a component is supported or guided by two surfaces at the same time, and these surfaces are either on different planes or on different cylindrical sections. In such cases, the surfaces have to be machined separately and will therefore have different dimensions caused by the tolerances. As a consequence, the force flow cannot be predicted clearly and assembly is made more difficult. Even though modern production machines have reduced the problems with tolerances, the lack of clarity will still affect function fulfilment and ease of assembly unless double fits are avoided. Double fits appear in various forms. Figure 7.6 shows examples and corrective measures.

Layout

The layout (general arrangement) and form design (shapes and materials) require a clear definition of the magnitude, type, frequency and duration of loads. If these data are not available, the implementation must be based on reasonable assumptions and the expected service life specified accordingly.

In any case, the embodiment must be such that the loads can be defined and calculated under all operating conditions. No impairment of the function or the durability of a component must be allowed to arise.

Similarly, following the checklist in Figure 7.3, behaviour with respect to stability, resonance, wear and corrosion must be clearly established.

Very often one comes across *double arrangements*, i.e. doubling up working principles for safety's sake, which conflict with the rule of clarity. Thus a shaft–hub connection designed as a interference fit will not have a better load-carrying capacity if it is also provided with a key, as in Figure 7.7. The extra element merely ensures correct positioning in the circumferential sense, but because of the reduction in the area at A, the resulting stress concentration at B and the presence

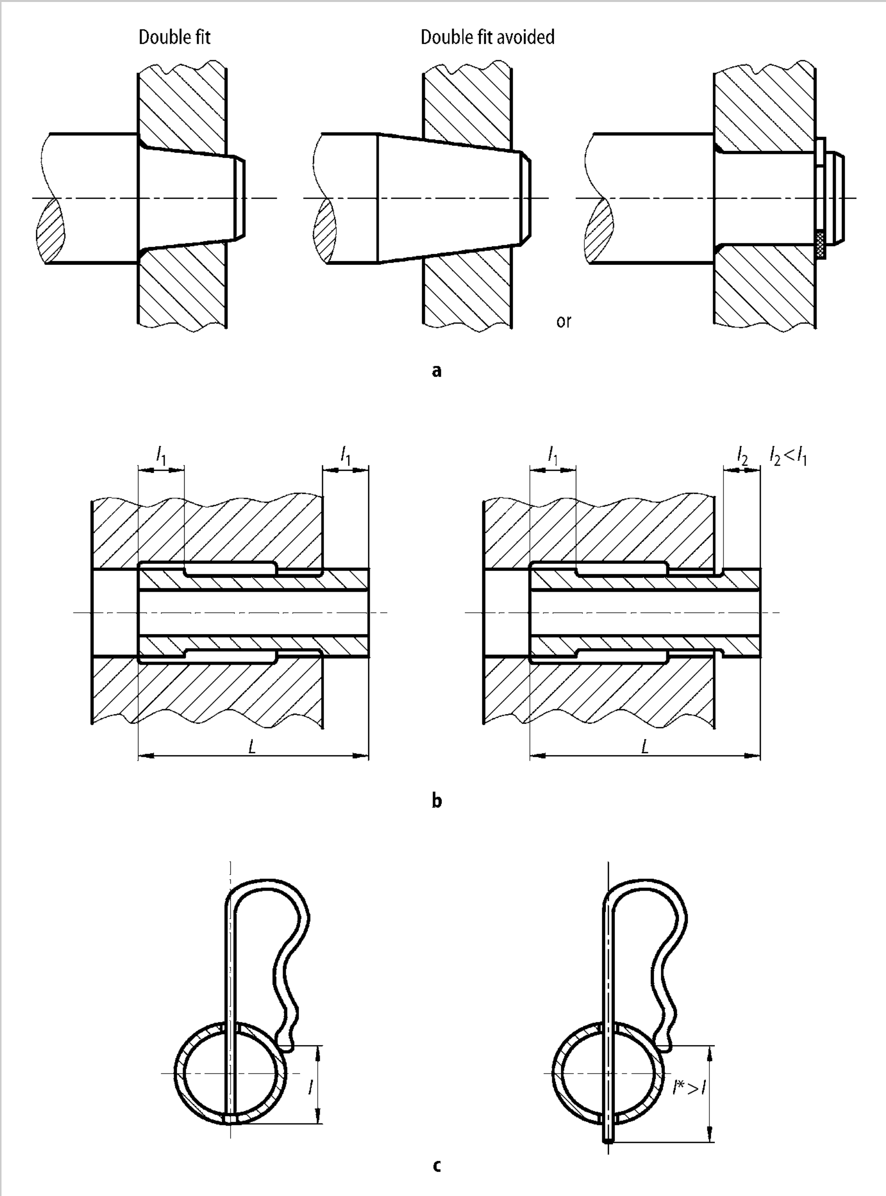


Figure 7.6. Avoiding double fits: **a** Tapered shaft–hub connection with interference (shrink) fit. The simultaneous axial location against the shaft collar and the taper seat creates a double fit: the radial pressure due to the interference fit cannot be determined. The right solution would be to use either a taper without a shaft collar or to use a cylindrical seat with a shaft collar. **b** Supported linear slide using a guiding sleeve in a housing. The simultaneous location of the housing at two points complicates the assembly process. A possible solution is shown in the figure on the right. **c** Spring clip of such a length that the lower end touches the tube at the same time as the pressure point touches the tube. The user will not be able to determine whether the clip is blocked by the tube or whether the spring force has to be overcome. The correct solution is shown in the figure on the right

of complicated and almost incalculable stresses at *C*, it decreases the strength in a drastic and fairly unpredictable manner.

Schmid [7.242] has shown that an axially preloaded taper joint for the transmission of torque requires a spiralling motion when the hub is assembled on the shaft in order to ensure a reliable interference fit, and the use of a key prevents this.

The employment of an interference fit to achieve the maximum torque capacity is only possible by leaving out the key. The solution shown in Figure 7.7 is only acceptable when the correct positioning of the hub relative to the shaft is the crux of the task, in which case a sliding fit is more appropriate.

Figure 7.8 shows a housing adapter for a centrifugal pump which can be used to provide various annulus profiles to fit different blade shapes so that new housings need not be constructed for each case. Unless the intermediate pressure in the gap between the adapter and the housing can be clearly regulated, or some other means of attachment is used, the adapter might travel upwards and damage the blades by rubbing against them.

This is particularly true when similar fits (H7-j6) are chosen for the two locating diameters which are approximately the same size. This is because, depending on production tolerances and working temperatures, gaps may appear, the relative sizes of which are unpredictable and which produce unknown intermediate pressures in the space between the adapter and the housing. The solution shown in Figure 7.8 (detail) ensures, by means of the specially designed connecting passage *A* (which must have a flow area roughly four to five times greater than the maximum gap area that might appear at the upper locating diameter), a clearly definable intermediate pressure, corresponding to the lower inlet pressure of the pump. As a result, the housing adapter is always pressed

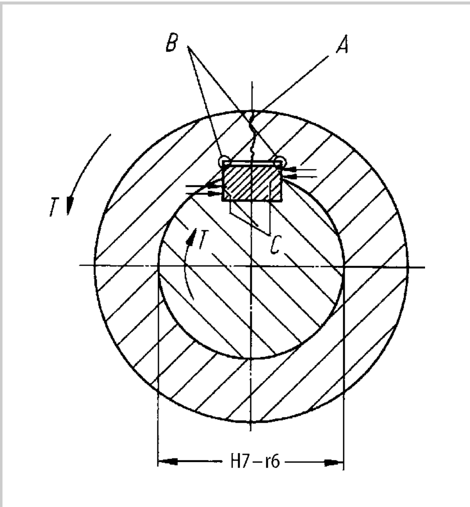


Figure 7.7. Combined shaft–hub connection achieved by means of shrink fit and key: an example of not applying the principle of clarity

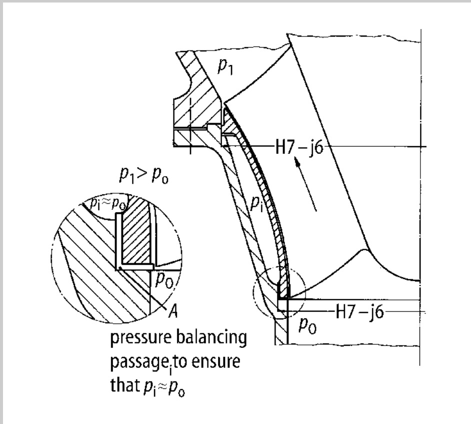


Figure 7.8. Housing adapter in a cooling-water pump

downwards when the pump is in operation, and attachments are only needed as locating aids for assembly and to prevent any tendency of the adapter to rotate.

Serious damage has been reported in gate valves whose operational or loading conditions were not clearly defined [7.130, 7.131]. When closed, gate valves separate, say, two steam pipes and at the same time close off the inside of the valve housing. The result is a self-contained pressure chamber, as shown in Figure 7.9. If condensate has collected in the lower part of the valve housing, and steam appears on the inlet side with the valve closed so that the valve is heated, then the enclosed condensate may evaporate and produce an unpredictable increase in pressure inside the valve housing. The result is either a ruptured housing or serious damage to the housing cover connection. If the latter is self-sealing, serious accidents may ensue since, in contrast to what happens with overloaded bolted flange connections, there is no preliminary leakage and hence no warning. The danger lies in the failure to specify clear operational and loading conditions. Possible remedies are as follows:

- Connect the inner chamber of the gate valve housing to an appropriate steam pipe, operational conditions permitting ($p_{\text{valve}} = p_{\text{pipe}}$)
- Protect the valve housing against excess pressure (p_{valve} restricted)
- Drain the valve housing, thus avoiding collection of condensate ($p_{\text{valve}} \approx p_{\text{external}}$)
- Design valves in such a way as to minimise the housing volume (collection of condensate kept low).

Similar phenomena in welded membrane seals are discussed in [7.206].

Safety

See basic rule in Section 7.3.3.

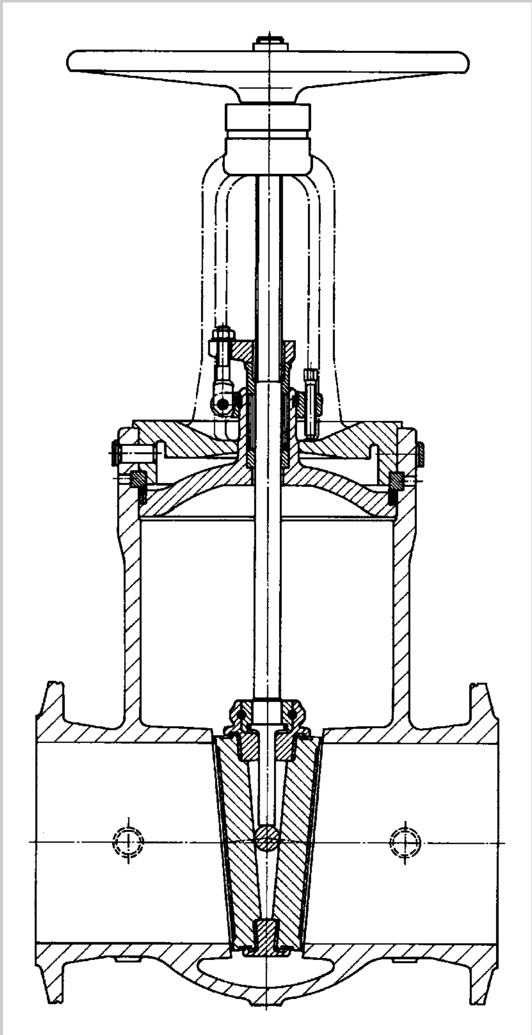


Figure 7.9. Gate valve with relatively large lower collecting area

Ergonomics

In human–machine relationships, correct operation must be ensured via the logical layout of equipment and controls.

Production and Quality Control

These must be facilitated by clear and comprehensive data in the form of product models as well as drawings, parts lists and instructions; and adherence to the prescribed production and quality control procedures.

Assembly and Transport

Much the same is true of assembly and transport. A clear assembly sequence preventing mistakes should be incorporated into the design (see Section 7.5.8).

Operation and Maintenance

Clear installation instructions and the appropriate embodiment design must ensure that:

- the performance is easily checked
- inspection and maintenance involves the smallest possible variety of tools and equipment
- the scope and schedules of inspection and maintenance are defined
- inspection and maintenance can be checked after they have been carried out (see Section 7.5.10).

Recycling

Designers should provide (see Section 7.5.11):

- clear separation of materials that are incompatible with regard to recycling
- clear sequences of assembly and disassembly.

7.3.2 Simplicity

For technical applications, the word “simple” means “not complex”, “easily understood” and “easily done”.

A solution seems simpler if it can be effected with fewer components, because, for example, the probability of lower production costs, less wear and lower maintenance is then greater. However, this is only true if the arrangement and shapes of the components are kept simple. Hence designers should always aim at the minimum number of components with the simplest shapes [7.168, 7.198, 7.206].

As a rule, however, a compromise has to be made. The fulfilment of a function always demands a certain minimum number of components. Cost efficiency often necessitates a decision between numerous components with simple shapes but with greater overall production effort, and, for example, a single cheaper cast component with the greater uncertainty it may entail in delivery. Simplicity must always be assessed from a holistic perspective—what constitutes “simpler” in individual cases depends on the problem and the constraints.

In what follows we shall be applying the basic rule of simplicity to the various headings of the checklist shown in Figure 7.3.

Function

In principle, only a minimum number and a clear and consistent combination of subfunctions should be pursued when considering the function structure.

Working Principle

In selecting working principles, only those involving a small number of processes and components, that have obvious validity and involve low costs should be taken into consideration.

In the development of the one-handed mixing tap (see Section 6.6.1), several solution principles were proposed. One group (see Figure 6.36) involved the use of only one component to realise two independent adjustments in directions tangential to the valve seat face (types of motion: translation and rotation). The other group (see Figure 6.33), though involving only movements in one direction (normal or tangential to the seat face), required an additional coupling mechanism to convert the two single adjustments into one direction of movement. Quite apart from the fact that, in the second group, the preset temperature is often lost when the tap is shut off, all solutions represented in Figure 6.33 involve a greater design effort than those in the first group. Hence, designers should always begin with a group like that depicted in Figure 6.36.

Layout

Here the simplicity rule requires:

- geometrical shapes which can be analysed simply for strength and stiffness
- symmetrical shapes which provide clearer identification of deformations during production and under mechanical or thermal loads.

In many cases, designers can reduce the work of calculation and experimentation significantly if they try, by means of a simple design, to facilitate the application of basic mathematical principles.

Safety

See under Section 7.3.3.

Ergonomics

The human–machine relationship should also be simple (see Section 7.5.5) and can be significantly improved by means of:

- obvious operating procedures
- clear physical layout
- easily comprehensible signals.

Production and Quality Control

Production and quality control can be simplified, and at the same time made faster and more accurate, if:

- geometrical shapes permit the use of well-established, time-saving methods
- production operations are minimised and have short setting-up and waiting times
- shapes are chosen to facilitate the inspection process.

Leyer, when discussing changes in production methods [7.166], uses the example of a sliding control valve approximately 100 mm long to demonstrate how the replacement of a complicated casting by a brazed product made of geometrically simple turned parts helped to overcome difficulties and paved the way for more economical production. Even though modern casting techniques now allow more intricate shapes to be produced relatively easily, further simplifications might still be expedient (see Figure 7.10). Step 3 helps to simplify the geometrical shape of the central, tubular part. Step 4 (fewer parts) can be taken when the surface areas at right angles to the valve axis need not be retained.

A further example is provided by the one-handed mixing tap discussed earlier. The design of the lever arrangement shown in Figure 7.11 is expensive to make, difficult to clean (slits, open recesses) and not aesthetically pleasing. The one shown in Figure 7.12 is much simpler and also more suitable for longer production runs. The lever, whose end can slide and rotate in a circumferential groove, requires a smaller number of parts and avoids wear in areas that are difficult to readjust. All in all, therefore, this solution is by far the better because it is more economic, easier to clean and looks nicer.

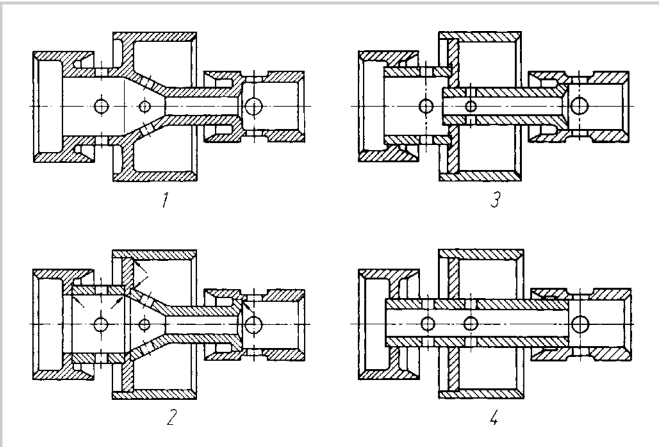


Figure 7.10. Simplification of a sliding control valve: 1 Casting is difficult and expensive; 2 Improvement by splitting into simple, brazed parts; 3 Simplification of central tubular part; 4 Further simplification possibility (1 and 2 after [7.166])

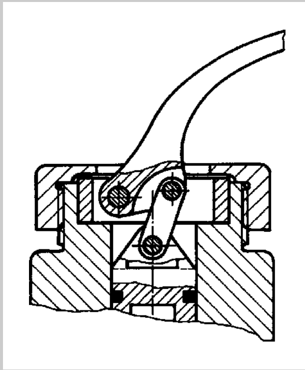


Figure 7.11. Proposed lever arrangement for a one-handed mixing tap with translational and rotational movements

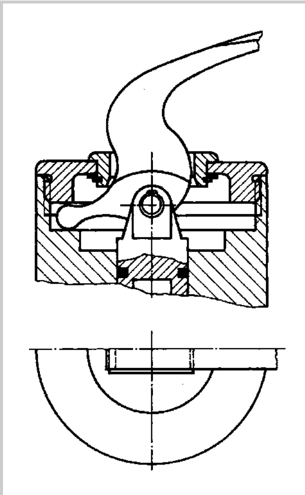


Figure 7.12. Simpler solution with improved embodiment (based on Schulte)

Assembly and Transport

Assembly is simplified—that is, facilitated, speeded-up and rendered more reliable—if:

- the components to be assembled can be identified easily
- the assembly instructions can be followed easily and quickly
- no adjustment has to be repeated
- reassembly of previously assembled components is avoided (see Section 7.5.9).

During assembly, the adjustment ring of a small steam turbine has to be moved vertically and horizontally with the turbine shaft already assembled, in order

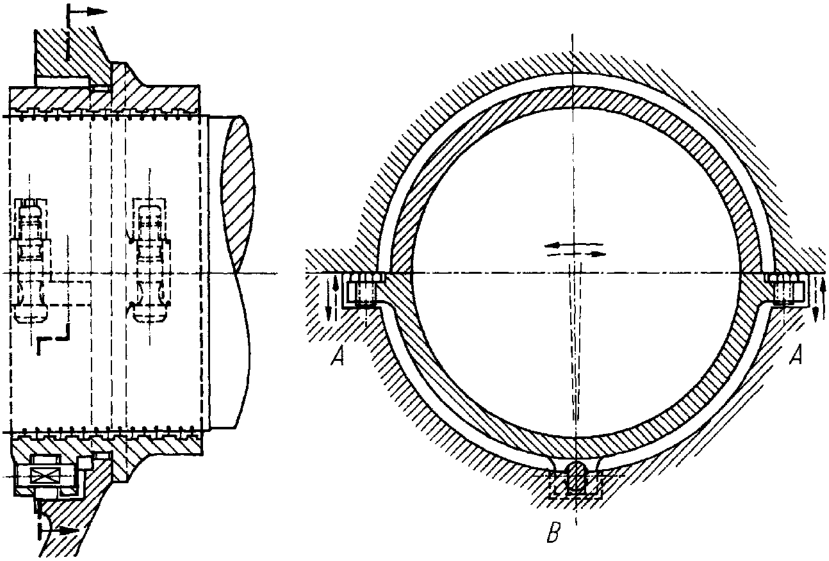


Figure 7.13. Adjustable sealing ring of an industrial steam turbine; adjustments at *A* in the same sense produce vertical movement, adjustments at *A* in the opposite sense produce a rotation about *B* that approximates to a horizontal movement

to ensure uniform clearance around the labyrinth seal. Doing this without having to remove the shaft several times for adjustment poses a problem that can be solved by the design shown in Figure 7.13. The adjustment can be made at the joint by rotating the adjustment screws *A* in the same sense, producing vertical movement only, and by rotation in the opposite sense, producing a tilting movement about pivot *B* that approximates to horizontal movement. The pivot itself must, however, allow for vertical movement during the adjustment and also for radial heat expansion when the turbine is operating. This is achieved with a few easily produced elements with simple shapes. A suitable arrangement of the surfaces, moreover, obviates the need to secure the pivot pin with additional locking elements: it is located in such a way that it can not fall out.

Operation and Maintenance

With respect to operation and maintenance, the simplicity rule means:

- operation must be possible without special or complicated instructions
- the sequence of operations must be clear and simple, and any deviations or faults easily identified
- maintenance must not be awkward, laborious and time-consuming.

Recycling

Simplicity for recycling can be realised by:

- use of recyclable materials
- simple assembly and disassembly processes
- simplicity of the parts themselves (see Section 7.5.11).

7.3.3 Safety

1. Nature and Scope of Safety Measures

Safety considerations affect both the reliable fulfilment of technical functions and also the protection of humans and the environment. Designers have recourse to a safety methodology that, following the German industry standard DIN 31 000 [7.57], includes the following three levels:

- direct safety
- indirect safety
- warnings.

In general, designers should try to guarantee safety by using *direct safety*, that is, by choosing a solution that precludes danger from the outset. Only when this proves impossible should they have recourse to *indirect safety*, in other words, constructing special protective systems [7.58 to 7.60]. *Warnings*, which merely point out dangers and indicate danger areas, can be used to support direct and indirect safety measures by, for example, pointing out special features, obstructions and disturbances. Only as a last resort should warnings be used on their own, and never as an easily implemented safety measure.

In the solution of technical problems, designers are faced with several constraints, not all of which they can hope to overcome simultaneously. They must nevertheless strive to provide a solution that comes nearest to satisfying all the requirements. The strength of an unavoidable safety requirement may, under certain circumstances, put the realisation of the whole project in doubt. A high demand for safety can greatly complicate a design and, by reducing clarity, may even lower the inherent safety of the product. Moreover, safety provisions may also render a product uneconomic and lead to its abandonment.

Such cases, however, are exceptional, because safety and economy generally go hand-in-hand in the long term. This is particularly true of expensive and complex plant and machinery. Only smooth, accident-free and safe operation can ensure long-term economic success. Protection against accidents or damage, moreover, goes hand-in-hand with reliability [7.75, 7.312]. Reliability makes it possible to operate a machine to full capacity, even though poor reliability may not necessarily lead to accidents or damage. All in all, it is therefore advisable to achieve safety by treating direct and indirect safety measures as an integral part of system design.

There are many different ways of applying safety measures in mechanical engineering. Therefore, we consider it necessary to provide some definitions before discussing the measures in detail. The withdrawn German industry standard DIN 31 004 (1979) defined safety as "being free from danger", a "danger" being a threat for which the type, size and action is known. A dangerous situation is one that can cause damage to persons or things. This DIN standard was replaced in November 1982 by DIN 31 004 Part 1 [7.61]. The basic terms are defined as follows:

Safety	is a state in which the risk is smaller than the risk limit.
Risk limit	is the largest but still acceptable system-specific risk relating to a particular technical process or situation.
Risk	is described by the frequency (probability) and the expected extent of the damage (scope).

Whereas the initial DIN standard defined protection as the limitation of danger in order to prevent damage, the 1982 standard uses the following definition:

Protection	is the reduction of risk by suitable means in order to reduce the frequency of occurrence and/or the extent of damage.

The DIN EN 292 standard [7.57] now uses these terms in a more general way. This development of the standard demonstrates that there is no absolute safety in the sense of complete freedom from danger. In common with many aspects of life, the use of technical systems always involves a certain risk. Safety measures aim to reduce risks to an acceptable level. However, what is acceptable (the risk limit) can only be quantified in a few cases. Now and in the future this limit will be determined by technical knowledge and social standards, and in no small measure by the experience and responsibilities of design engineers.

In the context of safety, it is very important to ensure reliability:

Reliability	is the ability of a technical system to satisfy its operational requirements within the specified limits and for the required life (definition based on [7.75, 7.76]).

It is clear that the reliability of individual components of a machine or the machine itself, as well as the reliability of any protective systems and devices, are important requirements for safety. Without state-of-the-art quality that ensures reliability, protective measures are of doubtful value.

One measure of reliability is the operational availability of a technical system.

Availability	is the percentage of time the system is available for operation compared to the maximum possible time or compared to a particular target time.

Safety concerns the following areas (see Figure 7.14):

Operational safety	is the limitation of danger (reducing risk) during the operation of technical systems in order to prevent damage to the systems themselves and their immediate environment, such as the workplace, neighbouring systems, etc.

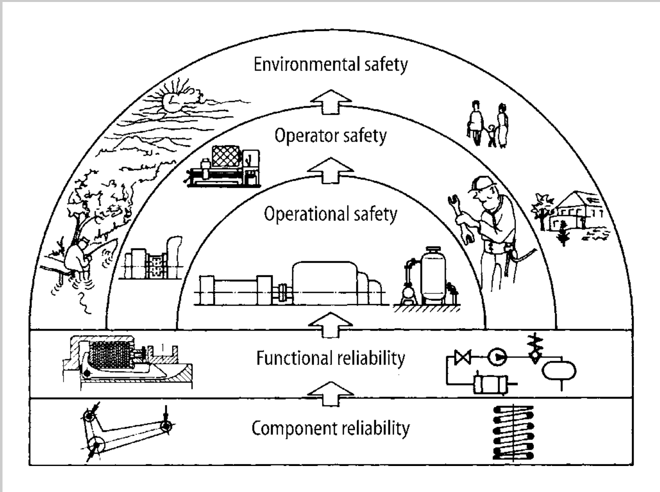


Figure 7.14. Relationship between component and functional reliability on the one hand and operational, operator and environmental safety on the other

Operator safety	is the limitation of danger to persons using technical systems either at their workplace or outside, for example for sport or leisure.
Environmental safety	is the limitation of damage to the environment in which technical systems are used.
Protective measure	is the use of protective systems or devices to limit existing dangers and reduce risks to acceptable levels where these cannot be achieved through direct safety measures.

The reliability of assemblies and of their interaction—that is, the functional reliability of a machine or a protective system—is crucial for operational, operator and environmental safety [7.179]. For designers, all these areas of safety are closely connected when developing a concept and its embodiment. A safety methodology should therefore give equal weight to each of the areas [7.210].

2. Direct Safety

Direct safety measures achieve safety through systems or components actively involved in the performance of a particular task. To ensure and evaluate the safe functioning and durability of components, designers can adopt one of several safety principles [7.210]. There are three basic principles, namely:

- safe-life principle
- fail-safe principle
- redundancy principle.

The *safe-life principle* demands that all components and their connections be constructed in such a way as to allow them to operate without breakdown or malfunction throughout their anticipated lives. This is ensured by:

- clear specification of the operating conditions and environmental factors, such as the anticipated loads, service life, operating conditions, etc.
- adequately safe embodiment based on proven principles and calculations
- numerous and thorough inspections during production and assembly
- analysis of components or systems to determine their durability when they are overloaded (load levels and/or running time) or subjected to adverse environmental influences
- determination of the limits of safe operation, with due regard being paid to possible breakdowns.

It is characteristic of this principle that it bases safety exclusively on accurate qualitative and quantitative knowledge of all of the influences at work or on the determination of the limits of failure-free operation. The application of this principle calls for a great deal of experience, or for costly and time-consuming preliminary investigations, and for continuous monitoring of the state of components. If a failure should nevertheless occur, and if a safe-life is essential, then as a rule there will be a serious accident, for instance the fracture of an aeroplane wing or the collapse of a bridge.

The *fail-safe principle* allows for the failure of a system function or for a component fracture during the service life by ensuring that grave consequences do not ensue. To that end:

- a function or capacity, however small, must be preserved to prevent dangerous conditions
- a restricted function must be fulfilled by the failing component or by some other component until such time as the plant or machine can be removed from operation without danger
- the failure or breakdown must be identifiable
- the effect of the failing component on the overall safety of the system must be assessable.

In essence, the impairment of a main function must be signalled. The signal can take various forms (increasing vibrations, loss of sealing, loss of power, slowing down), each without causing immediate danger. In addition, special monitoring systems may be provided to indicate the incipient failure to the operator. Their layout should be governed by the general principles of protective systems. The fail-safe principle presupposes knowledge of the progress of a failure and provides a means for taking over or maintaining the impaired function.

By way of example, let us consider a spherical rubber element in an elastic coupling (see Figure 7.15). The first visible crack appears on the outer layer, but the function is not yet impaired (State 1). Only when the number of revolutions under load is increased does the stiffness begin to decrease with a consequent change in the behaviour of the coupling, which manifests itself, for instance, by a lowering of the critical speed (State 2). With further operation, the crack grows larger and causes the stiffness to decrease still further (State 3), but even if the

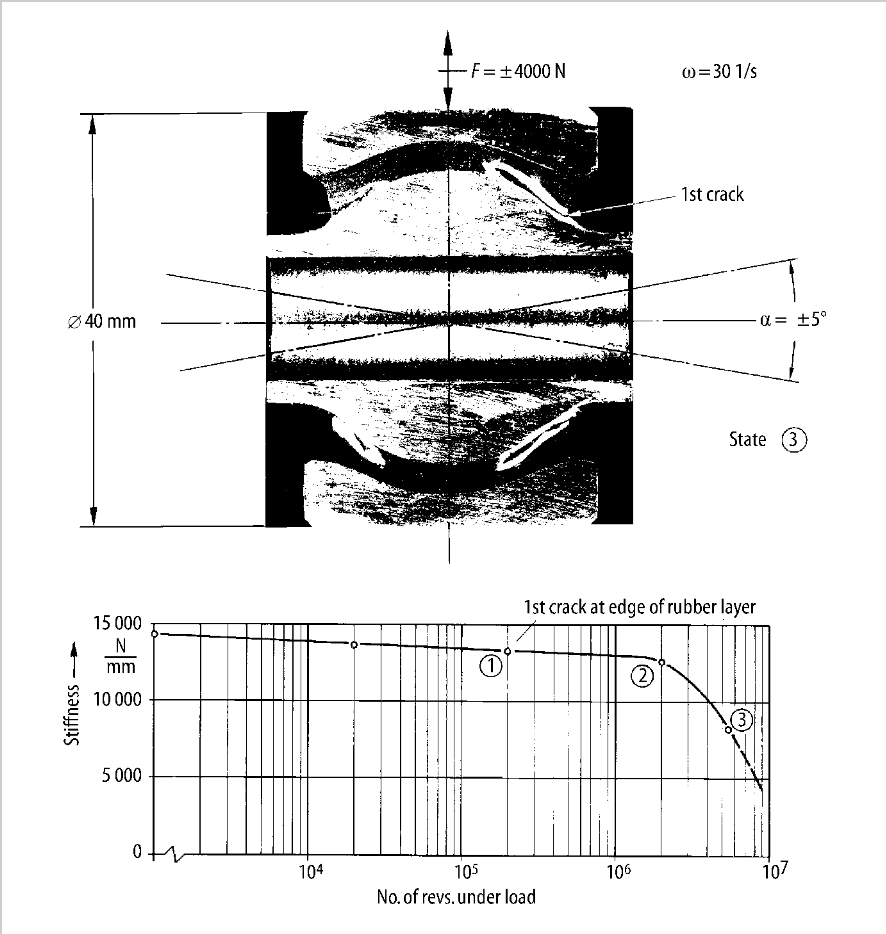


Figure 7.15. Fail-safe behaviour of an elastic coupling: crack-state and stiffness against number of revolutions

crack went right through, there would not be a complete failure of the coupling. Therefore, no sudden effect with serious consequences need be feared.

Another example is the behaviour of flange bolts made of a tough material which, on overloading, exceed their yield strength and deform plastically, resulting in a reduction of preload and, hence, a reduction of the clamping force. Their impaired function is indicated by the resulting loss in flange sealing but does not give rise to sudden failure.

Figure 7.16 illustrates two safe methods of fastening components. The means of attachment should be designed such that, even if the bolts begin to fail, the mountings remain in place, no broken parts can migrate, and the equipment continues to function to some extent [7.206].

The *redundancy principle* provides another means of increasing both the safety and the reliability of systems.

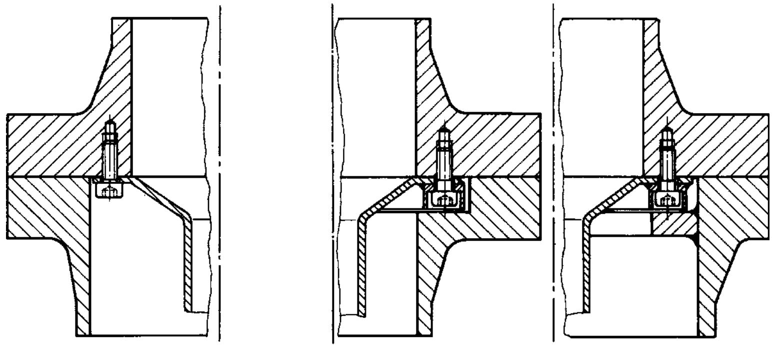


Figure 7.16. Fastening of components: the covering of the bolted connection maintains function and prevents broken parts migrating in the event of bolt failure

In common usage, redundancy means superfluity or excess. In information theory, redundancy refers to that fraction of a message that may be eliminated without loss of essential information. Redundancy is often used deliberately to allow for transmission losses, and hence to safeguard the system. The fact that this safety principle is common in electronics and information technology is useful when integrating these technologies with mechanical engineering systems.

Redundant safety arrangements lead to an increase in safety, provided that the breakdown of a particular element of the system is not dangerous in itself, and that other elements, arranged in parallel or in series, can take over its function fully or at least in part.

The provision of several engines in aircraft, of multistrand cable for a high-voltage transmission line, and of parallel supply lines or generators, all ensure that, should a particular element break down, the function is not completely impaired. In that case, we speak of *active redundancy*, because all the components are actively involved. Partial breakdowns lead to a corresponding reduction in energy or performance.

If reserve elements (for instance alternative boiler feed pumps)—usually of the same type and size—are provided and put into operation during breakdowns, then we speak of *passive redundancy*.

If a multiple arrangement is to be equal in function but different in working principle, then we have *principle redundancy*.

Depending on the situation, safety-enhancing elements can be arranged in parallel, for instance emergency oil pumps, or in series, for instance filter installations. In many cases, layouts in parallel or series will not suffice and crossover links will have to be introduced to guarantee transmission, despite the breakdown of several elements (see Figure 7.17).

In a number of monitoring systems, signals are collected in parallel and compared with one another. *Selective redundancy* (two out of three) and *comparative redundancy* arrangements are shown in Figure 7.17.

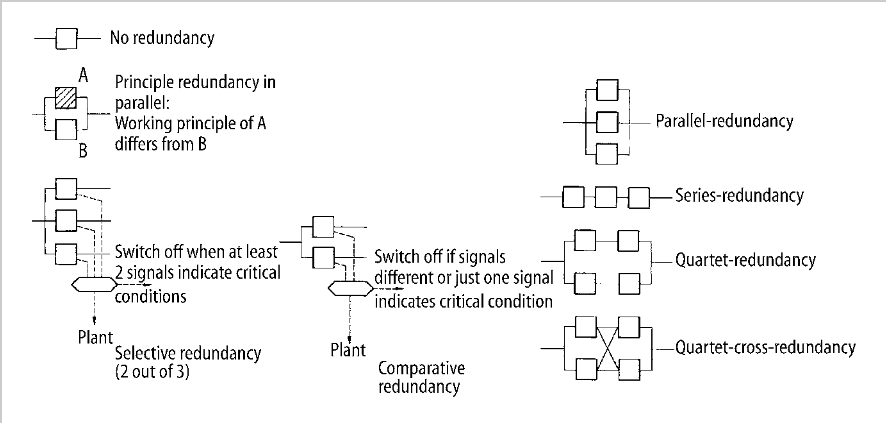


Figure 7.17. Redundant arrangements

Redundancy layouts cannot, however, replace the safe-life or fail-safe principles. Two cable cars operating in parallel will, admittedly, increase the reliability of passenger transport, but this will contribute nothing to the safety of the individual cars. The redundant layout of aircraft engines will not increase safety if any of the engines might explode and hence to endanger the system. In short, an increase in safety can only be guaranteed if the redundant element satisfies the safe-life or the fail-safe principle.

Adherence to all the principles we have mentioned—that is, the attainment of safety in general—is greatly facilitated by the principle of the division of tasks (see Section 7.4.2) and by the two basic rules of clarity and simplicity, as we shall now try to show with the help of an example.

The principle of the division of tasks and the clarity rule have been applied with great consistency to the construction of a helicopter rotor head (see Figure 7.18), and helped the designers to come up with a particularly safe construction based on the safe-life principle. Each of the four rotor blades exerts a radial force on the rotor head due to the centrifugal inertia force, and a bending moment due to the aerodynamic loading. The rotor blades must also be able to swivel so that their angles of incidence can be changed. A high safety level is achieved by the following measures:

- A completely symmetrical layout so that the external bending moments and the radial forces at the rotor head cancel out.
- The radial forces are transmitted exclusively by the torsionally flexible member *Z* to the main central component where they cancel each other out.
- The bending moment is only transmitted through part *B* and is taken up by the roller bearings in the rotor head.

As a result, every component can be optimally designed in accordance with its task. Complicated joints and shapes are avoided and the necessary high level of safety is attained.

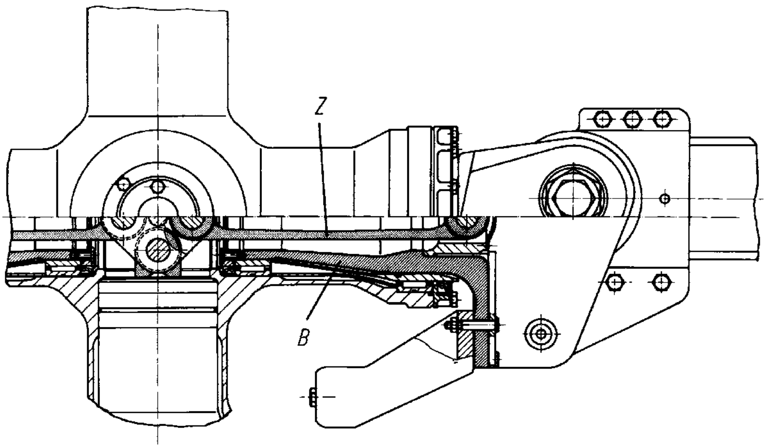


Figure 7.18. Rotor blade attachment of a helicopter based on the principle of the division of tasks (Messerschmitt–Bölkar system)

3. Indirect Safety

Indirect safety measures involve the use of special protective systems and protective devices. They are applied whenever direct safety measures prove inadequate. A detailed discussion of indirect safety measures for technical systems can be found in [7.215]. In what follows, the most important elements of these measures are described.

Protective systems react when danger occurs. To that end, their function structure includes a signal transformation with an input that captures the danger and an output that removes it.

The working structure of a protective system is based on a function structure with the following main functions: capture–process–act. Examples are the multiple redundant monitoring of temperatures in a nuclear reactor; the monitoring of robots in inaccessible workplaces; the sealing of areas when they are subject to X-rays; and the automatic checking of the locking of centrifuge covers prior to operation. The required actions can involve removing, limiting or separating.

Protective devices fulfil protective functions without transforming signals.

Examples are a pressure safety valve (see Figure 7.22); a shaft coupling that slips with torque overload; a pin that shears to limit excessive forces; and safety belts in cars. Their main action is removing or limiting. They can form part of a protective system.

Protective barriers fulfil protective functions without acting.

These barriers are passive, and not able to act on their own. They do not transform signals and therefore do not require a function structure that involves this transformation. They protect by separating; that is, by keeping persons and equipment at a distance from danger using physical barriers, covers, fences, etc. They are described in DIN 31 001, Parts 1 and 2 [7.58, 7.59]. Locking devices, according to Part 5 of this standard [7.60], are regarded as protective systems.

Basic Requirements

Indirect safety measures have to fulfil the following basic requirements:

- operate reliably
- function when danger occurs
- resist tampering.

Operate Reliably

Reliable operation means that: the working principle and the embodiment allow unambiguous operation; the layout follows the established rules; production and assembly are quality-controlled; and the protective systems and devices are rigorously tested. The safety modules and their functional links should be based on direct safety principles and demonstrate safe-life or fail-safe behaviour.

Function When Danger Occurs

This requirement means that:

- the protective function has to be available from the start of the dangerous situation and must last throughout the period of danger
- the protective function should not cease or the protective device should not be removed before the dangerous situation has completely ended.

Figure 7.19 shows example layouts for safety fence contacts for a machine guard. Closed contacts signal that the safety fence is in position. Layout *a* has severe deficiencies because the contact movement relies upon the spring force alone and is not bi-stable (see Section 7.4.4). If the spring breaks or the contacts stick together, the contact will not be broken, that is, the machine can be started with the safety

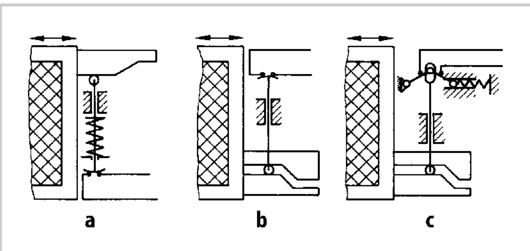


Figure 7.19. Layouts for safety fence contacts for a machine guard. **a** Protection not guaranteed because contact movement relies on a spring force alone. **b** Protection guaranteed because activation relies on form fit. **c** Bi-stable behaviour added to form fit activation in **b**

fence open. Layout *b* will always function when danger occurs. Sticking contacts will be opened because the effect relies on form rather than spring force, and if parts break they will not fall onto the contacts. Layout *c* also makes use of form for activation, but adds spring force and bi-stable behaviour. Further examples can be found in [7.215].

Resist Tampering

Resistance to tampering means that the protection cannot be reduced or removed by unintended or intended actions. If we consider the safety fence contact in Figure 7.19, it should be designed such that actions that prevent correct operation are not possible. The best way to achieve this is to use a cover that cannot be opened without tools or without stopping the machine.

The requirements of protective systems and devices are listed in the following paragraphs followed by those of protective barriers.

Protective Systems and Devices

Protective systems and devices render endangered plant or machinery safe automatically, with the aim of preventing danger to persons and machinery. In principle, the following approaches are available:

- When danger occurs, prevent the consequences by disabling the plant or machinery or preventing any plant or machinery in a dangerous state from being put into operation.
- When there is a continuous danger, avoid its effects by introducing protective measures.

The basic requirements "operate reliably", "function when danger occurs", and "resist tampering" are supported by fulfilling the following requirements.

Warning

When a protective system notes changes in the working conditions, a warning must be provided that indicates the change and the cause of the warning. Examples are "oil level too low", "temperature too high", and "safety fence open". Recommended acoustic and optical signals are given in DIN 33 404 [7.69], colours for warning lights and push buttons in DIN IEC 73/VDE 0199 [7.77], and special safety symbols in DIN 4844 [7.40–7.42].

Two-Step Action

If the dangerous situation emerges so *slowly* that operator action can reduce the danger, then a warning should be given before a protective action is initiated.

Between the two steps, there should be a sufficiently large and clearly defined change in the danger variable. For example, if pressure is the danger variable being monitored, a warning could be given at 1.05 p_{normal} and shutdown initiated at 1.1 p_{normal}.

If the dangerous situation emerges too *quickly*, the protective system should react immediately and signal its response clearly. The terms “slowly” and “quickly” must be interpreted in the context of the cycle time of the technical process and the reaction time required [7.243].

Self-Monitoring

A protective system must be self-monitoring; that is, it must be triggered not only when the system breaks down, but also by faults in its own system. This requirement is best satisfied by the *stored energy principle*, because, when this is applied, the energy needed to activate the safety device is stored within the system and any disturbance of or fault in the protective system will release that energy and switch off the plant or machinery. This principle can be used not only in electronic protective systems but also in mechanical, hydraulic and pneumatic systems.

The stored energy principle has been used in the valve shown in Figure 7.20. When the valve opens, the spring is compressed by the operating oil pressure. When the oil pressure reduces, the spring extends and the valve closes. Failure of the spring will not inhibit the closure of the valve because of the particular configuration used. The flow direction selected and the suspended configuration support the requirement of always functioning when danger arises.

A further example of the use of the stored energy principle in a hydraulic system is shown in Figure 7.21. In this protective system, pump 1 with a pressure-regulating valve 2 ensures a constant pre-pressure p_{P}. The protective system with the pressure p_{S} is connected to the pre-pressure system by means of an orifice 3. Under normal conditions, all outlets are closed, so that the quick-action stop valve 4 is held open

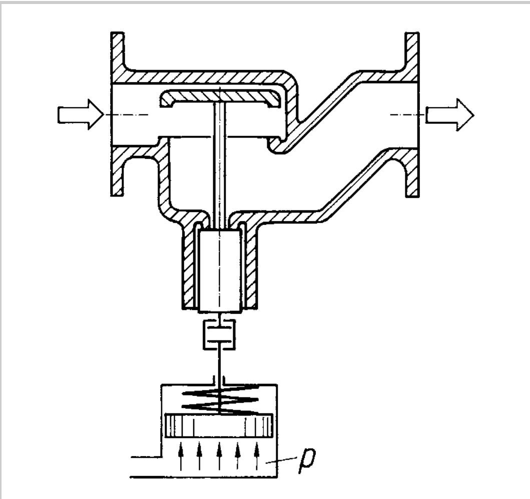


Figure 7.20. Layout of a quick-action valve. In the event of a drop in oil pressure p, the spring force, the flow pressure on the valve face and the weight of the valve act together to guarantee the rapid closure of the valve

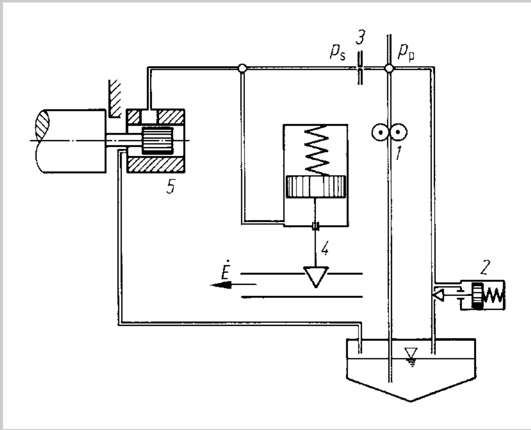


Figure 7.21. Hydraulic protection system employed to prevent incorrect axial shaft positions based on the stored energy principle

by the pressure p_s, allowing energy to be supplied to the machine. In the case of a faulty axial shaft position, the piston valve 5 at the end of the shaft opens, the pressure p_s drops, and further energy supplies are cut off by the quick-action stop valve 4. The same effect is produced by damage to the pre-pressure or protective system, for example by pipe fracture, lack of oil or pump failure. The system is self-monitoring.

A system operating on the *active energy principle*, where energy is only generated in the case of danger, cannot detect a failure in its own system. Therefore, this approach should only be used to provide the warning signals of a protective system when a monitoring system is also available and the system is checked regularly. The possibility that a protective system based on the stored energy principle can cause interruptions that are not caused by a dangerous situation but instead by the protective system itself should be met by increasing the reliability of the system elements, and not through application, for example, of the active energy principle.

Redundancy

The failure of a protective system or device should be seen as a real possibility. Because a single protective system may break down, its mere doubling or replication ensures greater safety: it is unlikely that all the systems will fail at once. A solution that is often applied in protective systems is redundancy based on two from three selection. Three sensors are used to detect the same danger signal (see Figure 7.17). Only when at least two sensors signal the critical value is the protective action—such as machine shutdown—initiated. Thus the failure of a single sensor does not reduce the protective cover, and its failure will not trigger an unnecessary protective action [7.179].

This is however only true provided that the replicated protective systems do not all fail due to a common fault. Safety is considerably increased if the double or multiple systems work independently of one another and are, moreover, based on different working principles (principle redundancy). In this case, common faults—for instance those due to corrosion—will not have catastrophic consequences: the simultaneous breakdown of all such systems is highly improbable.

Figure 7.22 illustrates protective devices employed to prevent excessive pressure in pressure vessels. Mere doubling would not protect against common failures such as corrosion or inappropriate materials. The use of different working principles, however, reduces the possibility of simultaneous failure.

When redundant configurations are linked in parallel or series, the values at which they are triggered should be carefully staggered within an appropriate range. In this manner, primary and secondary protection can be established. In the example in Figure 7.22, the configuration should be chosen such that the safety valve is activated at a lower excess pressure than the shear plate.

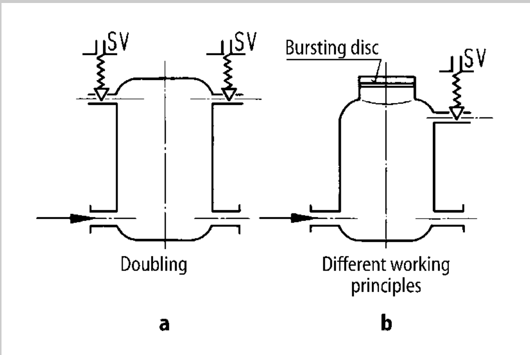


Figure 7.22. Protective devices employed to protect against excessive pressure build-up in pressure vessels: **a** two safety valves (not safe against common faults); **b** safety valve and shear plate (principle redundancy)

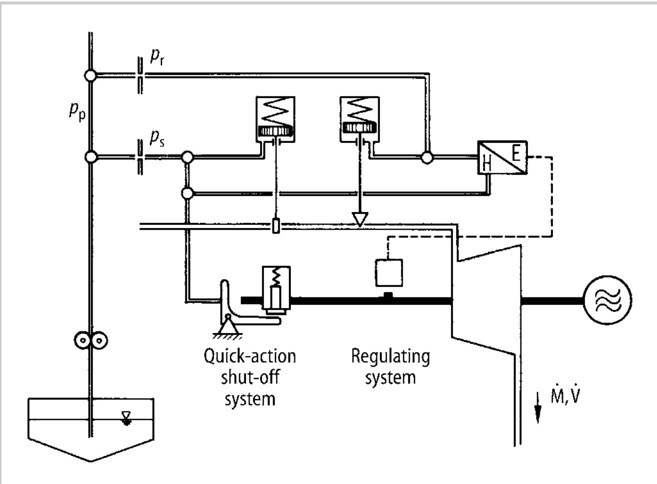


Figure 7.23. Stored energy protective system against overspeeding based on principle redundancy

In many cases the primary protection system can receive its signals from an existing control system, if it has the characteristics of a protective system. This requirement is met in the control of steam turbines shown in Figure 7.23 [7.272]. In the case of overspeeding, the energy supply is cut off by two systems that differ in principle. Increases in speed first invoke the regulating system, whose speed measurement and regulating valve are independent of, and different in principle to, the quick-action shut-off system.

Speed is measured by three identical but independent magnetic sensors. They take their measurements from a gear wheel on the turbine shaft (see Figure 7.24). Their primary purpose is to control the speed of the machine through electronics and hydraulics. In addition, each signal is compared with a reference signal in order to prevent excess speed. This comparison is based on the two from three

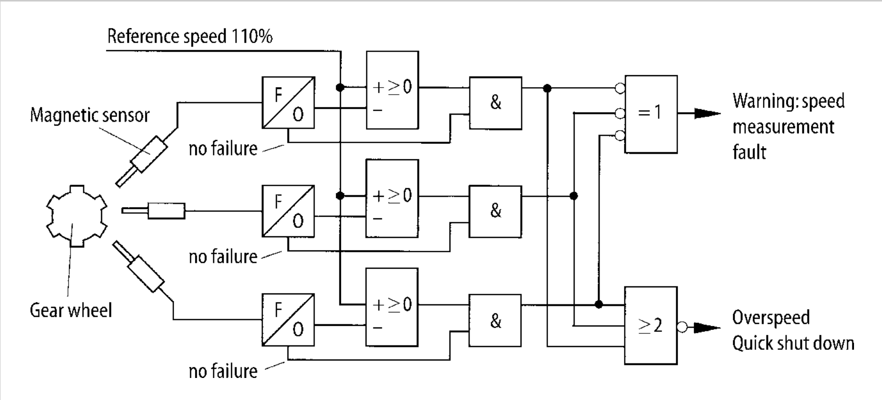


Figure 7.24. Electronic speed control and speed monitoring using a redundant layout based on the two from three principle (simplified representation). Safety is based on the stored energy principle, which is also applied to the quick-action shut-off system

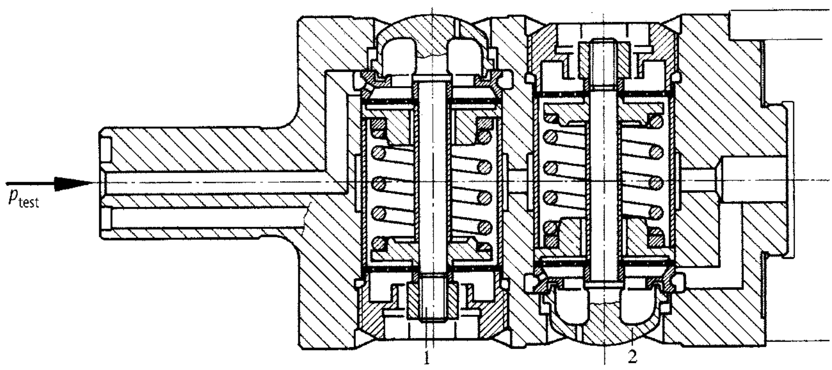


Figure 7.25. Stored energy protective system against overspeeding based on two triggering values

principle. Each measurement circuit is monitored separately, and any failures are signalled. If two fail, the quick-action shut-off system is activated immediately.

The measurement and the activation of the quick-action system, however, are based on a mechanical principle. Figure 7.25 shows quick-action pins that, in the case of excess speed, move out rapidly against their retaining springs and strike a trigger. This in turn activates the quick-action shut-off system hydraulically. The turbine is provided with two such bi-stable devices that trigger at 110% and 112% excess speed respectively (see Section 7.4.4).

A common hydraulic supply to the control and quick-action shut-off system based on the stored energy principle is acceptable because both are based on a common self-monitoring principle.

Bi-Stability

Protective systems and devices must be designed with a clearly defined triggering value. When this value is attained, the protective reaction must be initiated immediately and unambiguously. This can be achieved by using the bi-stable principle (see Section 7.4.4). Below the triggering value, the system is in a stable state. When the triggering value is attained, an unstable condition is created deliberately. This avoids intermediate states and transfers the system rapidly into its second stable state. This bi-stable characteristic must be realised without intermediate states occurring when the triggering value is reached in order to achieve clarity in the behaviour of the protective system or device.

Preventing System Restarts

After a protective system or device has been activated, that system should not automatically return a machine to normal operation, even if the danger recedes. The activation of a protective system is always triggered by an unusual situation. After shutdown, the situation should be checked and evaluated, and the subsequent restart should follow a clearly structured procedure. For example, the safety regulations covering protective systems and devices [7.256], as well as other machines used in production [7.334], prescribe procedures for restarting.

Testability

A protective system or device should allow its functioning to be tested without having to create a situation with real danger. However, it might be necessary to simulate a dangerous situation in order to trigger the protective system. During a simulation, the effects used must be similar to the real danger and all possible danger conditions checked.

In our speed control system example, this means a planned increase in speed up to the excess speed, at which point the protective system triggers. If this is not possible or it is not desirable, it is possible to simulate the centrifugal inertia force by using oil pressure to trigger the system. The machine does not have to be shut down for this simulation. Figure 7.25 shows the oil channel. The oil simulates an increase in the centrifugal inertia force on the quick-action shut-off pins so that they are triggered and their action tested without attaining an excess speed.

With redundant protective systems, it is possible to isolate individual systems from the machine to test them. Any other redundant protective systems can remain active and continue to monitor safety during the test. Care must be taken to ensure that the protective system automatically returns into its fully operational state after test procedures that only check part of the system.

From the previous paragraphs, the following points emerge:

- protection must be retained during testing
- testing must not introduce new dangers
- after testing, the parts tested should return automatically to their fully operational state.

Often a *start-up check* is useful, or even prescribed. This check permits the operation of a machine only after its functions have been tested by activating the protective system. Safety regulations, for example, often prescribe this type of start-up check for power tools with safety devices [7.256].

Protective systems and devices must be *tested regularly*, that is:

- before the first operation
- at regular predetermined intervals
- after every service, repair or modification.

The procedures should be described in operating manuals and the results documented.

Relaxing the Requirements

At this point, one may question whether it is necessary to meet the testability requirement as well as that of self-monitoring. However, even protective systems based on the stored energy principle include elements whose full functionality can only be assessed through testing. Examples include the operation of the quick-action pins in Figure 7.25, and sticking contacts in an electric switch.

Relaxation of the safety system requirements is only permissible when the probability of failure is so small and the consequences of any failure are so limited that the overall risk is acceptable. This will only be the case with redundancy requirements when system tests are easy and carried out regularly. This occurs when these tests are part of normal operation, for example when start-up checks are implemented. This often applies to protective systems associated with safety at work.

If human life is endangered or large-scale damage may occur, leaving out redundancy is neither justified nor economic. Which redundancy is applied, for example two from three selection, replication of the same principle, or principle redundancy, depends on the specific context and the level of risk.

Protective Barriers

The purpose of a protective barrier is to isolate people and objects from the source of danger, and to protect them from a variety of dangerous effects. DIN 31 001 Part 1 [7.58] and Part 2 [7.59] deal mainly with protection against physical contact with dangerous static and moving parts, and against objects and particles that break away. Elaborate illustrations and examples are given in [7.215].

The desired solution principles (see Figure 7.26) prevent contact by providing:

- full enclosure
- cover for a particular side
- fence, used to maintain a safe distance.

Safety distances play an essential role when it is possible to reach through or around fences or barriers. These distances are determined by body dimensions and ranges of reach. DIN 31 001 Part 1 [7.58] gives clear safety distances, depending on body dimensions and posture.

With respect to contact protection and protection against objects and particles that break away, DIN 31 001 Part 2 [7.59] only permits the use of those materials that can fulfil their protective function on the basis of their durability, shape stability, temperature resistance, corrosion resistance, resistance to aggressive substances, and their permeability to those aggressive substances.

4. Designing for Safety

The checklist in Figure 7.3 can prove a great help. Safety criteria must be scrutinised with respect to all the headings listed [7.303].

Function and Working Principle

It is important to establish whether or not the function is fulfilled safely and reliably by the chosen solution. Likely faults and disturbing factors must be taken

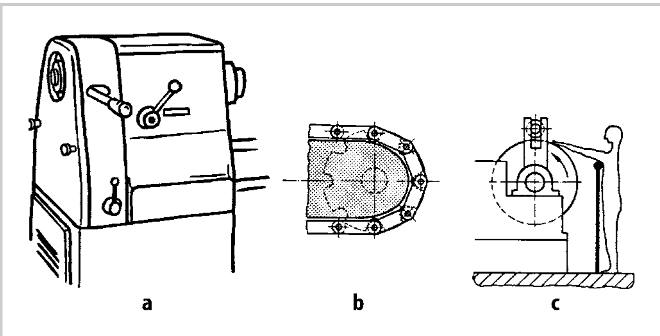


Figure 7.26. Examples of protective barriers: **a** full enclosure; **b** cover for a particular side; **c** fence used to maintain a safe distance

into account as well. The extent to which allowances must be made for exceptional, purely hypothetical, circumstances that could affect the function is not always clear, however.

The correct estimation of the scope and likelihood of a risk should be based on the successive negation of each of the functions to be fulfilled and on an analysis of the likely consequences (see Section 10.2). Sabotage need not necessarily be considered in this context, because measures to prevent human errors are likely to cover most possible circumstances.

What we have to consider and prevent first and foremost are failures due to possible disturbances of the structure, operation and environment of a machine, as well as those caused by operator error. Harmful effects that are not due to technological factors cannot be eliminated by the technical system itself, but the system must be able to survive them and, if possible, limit them.

A further question is whether the direct safety measures we have been discussing are adequate, or whether safety should be increased by additional protective systems and devices. Finally, we might also ask whether the whole project should be abandoned if it proves to be impossible to make adequate safety provisions in a particular case. The answer depends on the degree of safety that has been attained, on the probability of unpreventable damage or accident, and on the magnitude of the possible consequences. Objective standards are often lacking, particularly in the case of new applications. It has been argued that technical risks must be no greater than the risks humans must expect from natural causes [7.138]. However, this is always a matter for discretion. The final decision should, in any case, reflect a responsible attitude towards the human race.

Layout

External loads produce stresses in components. Through analysis we determine their magnitude and frequency (steady and/or alternating loads). The various types of stress produced can be determined by calculation or experiment. The calculated stresses in a component are then, using an appropriate failure hypothesis, converted into an *equivalent stress* σ_E, which should correctly represent the combined direct and shear stresses. The maximum equivalent stress should not exceed the *allowable stress* σ_A. When the two are equal, the material utilisation is 1.0. In general, the ratio of the equivalent stress divided by the allowable stress is smaller than 1.0, because the choice of dimensions is also influenced by standards and other embodiment considerations.

Materials technology provides designers with *material stress limits* σ_L or particular conditions (tension, compression, bending, shear and torsion), beyond which the material will fail or permanently deform. These values are usually obtained from test specimens and not from the components themselves. The strength of a component is also affected by uneven loading, and by its size, surface finish and shape. Only when these are taken into consideration can adequate durability be guaranteed. Thus the component stress limit is usually lower than the material stress limit.

The ratio of the material stress limit (or of the component stress limit) to the allowable stress is the *Safety Factor*, $(SF) = \sigma_L/\sigma_A$. This value must be greater than 1.0. Safety factors are provided in reference manuals for specific situations and types of materials, and the allowable stress σ_A in a component can easily be calculated using these.

The value of a safety factor depends on uncertainties in the determination of the material stress limits; on uncertainties in the load assumptions; on the calculation methods; on the production processes; on the (uncertain) influences of shape, size and environment; and also on the probability and importance of possible failures.

The determination of safety factors still lacks generally valid criteria. An investigation by the authors has shown that published recommended safety factors cannot be classified by type of product, branch of engineering or other criteria such as toughness of material, size of component, probability of failure, etc. Tradition, figures based on one-off and often inadequately explained failures, hunches and experiences are often the basis for numerical data from which no generally valid statements can be derived.

The figures that are given in the literature must therefore be treated with circumspection. Their application usually calls for a knowledge of the individual circumstances and of the special practices or regulations of the branch of engineering in question. In general, however, safety factors smaller than 1.5 should only be used when more precise calculation procedures have been used, experimental data are available, a sufficiently ductile material is used, or there is experience with the specific application. For brittle materials subject to stresses that lead to brittle fracture, the safety factor will be nearer to 2.0.

Toughness—that is, the ability to undergo plastic deformation before failure and thus relieve stress concentrations caused by unevenly distributed loads—is one of the most important safety features any material can have. The usual overspeed spinning tests of rotors with the correspondingly high stresses they set-up, and also the required overpressure tests of pressure vessels—provided that they are built of tough materials—are good examples of the direct safety method aimed at reducing stress concentrations in finished components.

Because toughness is a crucial safety-enhancing property of materials, it is not enough simply to aim at greater yield strength. Since, in general, the toughness of materials decreases with increasing yield strength, it is essential to ensure a minimum toughness, otherwise the benefits of plastic deformation are no longer guaranteed. Also dangerous are those cases in which the material turns brittle with time or for other reasons (for instance, due to radiation, corrosion, heat, or surface coatings). This is particularly true of synthetic materials.

If the safety of a component is calculated merely by the difference between the computed stress and the maximum permissible stress, a vital point is missed.

Of the utmost importance is the loading condition and the effect on the properties of the material due to ageing, heat, radiation, weathering, operating conditions and production processes, for instance welding and heat treatment. Residual stresses must not be underestimated either: brittle (fast) fractures without plastic deformation can occur suddenly and without warning. The avoidance of a build-up of additive stresses, of brittle materials, and of production processes that encourage brittle fractures, is therefore an essential requirement of direct safety.

If plastic deformation is monitored at a critical point, or can be used to impede the function in such a way that the danger can be noticed before humans or machines are endangered, it becomes fail-safe [7.206].

Elastic deformations must not be allowed to disturb the smooth functioning of a machine, for instance through loss of clearance. If this happens, the force transmission paths or the expansions can no longer be determined with certainty and overloading or fracture may ensue. This is just as true of stationary as it is of moving parts (see Section 7.4.1).

By *stability* we refer not only to the basic stability of a machine but also to its stable operation. Disturbances should be counteracted by stabilising effects, that is, by automatic return to the initial or normal position. Designers must ensure neutral equilibrium or that potentially unstable states do not lead to a build-up of disturbances that might get out of control (see Section 7.4.4).

Resonances produce increased stresses that cannot be accurately determined. They must be avoided unless the amplitudes can be sufficiently damped. This applies not only to the stability problem, but also to such associated phenomena as noise and vibration, which impair the efficiency and health of operators.

Thermal expansions must be taken into account under all operating conditions, in particular during unsteady processes, if overloading and impairment of the function are to be avoided (see Section 7.5.2).

Inefficient *seals* are a common cause of breakdown or trouble. Careful choice of seals, provision for pressure relief at critical sealing points and careful attention to fluid dynamics help to overcome these problems.

Wear and the resulting particles can also impede operational safety, and must therefore be kept within tolerable limits. In particular, designers should ensure that such particles do not damage or interfere with other components. They should be removed as near as possible to their point of origin (see Section 7.5.13).

Uniform *corrosion* reduces the designed thickness of components. Local corrosion, particularly of components subject to dynamic loading, may appreciably increase stress concentrations and lead to fast fractures with little deformation. There is no such thing as permanent stability under corrosion—the load capacity of components decreases with time. Apart from fretting corrosion and fatigue corrosion, stress corrosion can also be very serious for certain materials subject to tensile stresses in the presence of corrosive media. Finally, corrosion products can impede the functioning of machines, for instance by jamming valve spindles, control mechanisms, etc. (see Section 7.5.4).

Ergonomics

The application of ergonomic principles to industrial safety involves the careful scrutiny of sources and locations of danger as well as of human–machine relationships. Possible human errors and fatigue must also be included. Machines and products therefore have to be designed ergonomically (see Section 7.5.5).

Table 7.1. Harmful effects associated with various types of energy

Protect humans and environment against harmful effects Headings	 Examples
Mechanical	Relative movement of human and machine, mechanical vibrations, dust
Acoustic	Noise
Hydraulic	Jets of liquid
Pneumatic	Jets of gas, pressure waves
Electrical	Passage of current through body, electrostatic discharges
Optical	Dazzle, ultra-violet radiation, arcs
Thermal	Hot and cold parts, radiation, inflammation
Chemical	Acids, alkalis, poisons, gases, vapours
Radioactive	Nuclear radiation, X-rays

Table 7.2. Minimum industrial safety requirements in mechanical devices

In mechanical devices, protruding or moving parts should be avoided in areas where human contacts may occur
Protective equipment is required for the following, regardless of the operational speed: • for gear, belt, chain and rope drives • for all rotating parts longer than 50 mm, even if they are completely smooth • for all couplings • in cases of danger from flying parts • for potential traps (slides coming up against stops, components pushing or rotating against each other) • descending components (weights, counter-weights) • for slots, for example at material inputs. The gaps between parts must not exceed 8 mm; in the case of rollers, the geometrical relationship must be examined and, if necessary, special guards must be installed
Electrical installation must always be planned in collaboration with electrical experts. In the case of *acoustic*, *chemical* and *radioactive* dangers, expert advice must be sought for the requisite protection

A great many books and papers have been devoted to this subject [7.26, 7.65, 7.189, 7.255, 7.303]. In addition, DIN 31 000 [7.57] specifies the basic requirements of design for safety, and Parts 1, 2 and 10 of DIN 31 001 [7.58, 7.59] deal with protective equipment. Regulations by various professional bodies, factory inspectorates, etc., must be scrupulously observed in all branches of engineering, and so must a great deal of special legislation [7.115] (see also [7.334]). In this book it is impossible to examine every aspect of industrial safety.

Tables 7.1 and 7.2 provide a introductory guide to the sources of danger and the minimum requirements for industrial safety.

Production and Quality Control

Components must be designed in such a way that their qualities are maintained during production (see Chapter 10). To that end, special quality controls must be instituted, if necessary by special regulations. Through appropriate design measures, designers must help to avoid the emergence of dangerous weak spots in the course of production processes (see Sections 7.3.1, 7.3.2 and 7.5.8).

Assembly and Transport

The loads to which a product will be subjected during assembly and transport must be taken into consideration during the embodiment design phase. Welds carried out during assembly must be tested and, where necessary, heat treated. All major assembly processes should, whenever possible, be concluded by functional checks.

For safe transportation, firm bases, support points and handling points should always be provided and marked clearly. The weights of parts heavier than l00 kg should be marked where they can be seen easily. If frequent dismantling is called for, the appropriate lifting points must be incorporated.

Operation

Operation and handling must be safe [7.57, 7.58]. The failure of any automatic device must be indicated at once so that the requisite actions can be taken.

Maintenance

Maintenance and repair work must only be undertaken when the machine is shut down. Particular care is needed to ensure that assembly or adjusting tools are not left behind in the machine. Safety switches must ensure that the machinery is not started unintentionally. Centrally placed, easily accessible and simple service and adjustment points should be provided. During inspection or repair, safe access should be possible through the provision of handrails, steps, nonslip surfaces, etc.

Costs and Schedules

Cost and schedule requirements must not affect safety. Cost limits and delivery dates are ensured by careful planning, and by implementing the correct concepts and measures, not by cutting corners. The consequences of accidents and failures are generally much greater and graver than the effort needed to prevent them.

7.4 Principles of Embodiment Design

The general principles of embodiment design have been discussed at some length in the literature. Kesselring [7.148] set out principles of minimum production costs, minimum space requirements, minimum weight, minimum losses, and optimum handling (see Section 1.2.2). Leyer discussed the principle of lightweight construction [7.167] and the principle of uniform wall thickness [7.168]. It is obviously neither possible nor desirable to have all of these principles implemented in every technical solution—one of them might be crucial, the rest merely desirable. Which principle should be prioritised in a given case can only be deduced from the task and the company's facilities. By proceeding systematically, elaborating a requirements list, abstracting to identify the crux of the problem, and also by following the checklist given in Figure 5.3, designers transform these principles into concrete proposals that enable them to determine production costs, space requirements, weights, etc. These have to be consistent with the requirements list.

The systematic approach also highlights the question of how, with a given problem and a fixed solution principle, a function can be best fulfilled and by which type of function carrier. Embodiment design principles facilitate this part of the design process. In particular, they help with Steps 3 and 4, but also with Steps 7 to 9 as listed in Section 7.1.

Initially embodiment problems focus predominantly on issues of channelling, combining and storing. For the relatively common task of *transmitting* (channelling) forces or moments, it seems advisable to establish special "principles of force transmission". *Changing* the type or *varying* the magnitude of a force are primarily fulfilled by the appropriate physical effects, but designers must also apply the "principle of minimum losses" [7.148] for energy conservation or economic reasons, which they do by adopting a small number of highly efficient steps. This principle also applies to the efficient conversion of one type of energy into another, whenever this should be required. *Storing* energy involves the accumulation of potential and kinetic energy, be it directly or indirectly through the collection of material. The storage of energy, however, raises the question of the stability of the system, and the consequent application of the "principles of stability and bi-stability".

Often, several functions have to be fulfilled by one or several function carriers. Here the "principle of the division of tasks" may be useful to designers. Its application involves careful analysis of the functions and their assignment to function carriers. This analysis of functions is also helpful for the application of the "principle of self-help" when supplementary effects must be identified and exploited.

When applying embodiment design principles, designers may find that they run counter to certain requirements. Thus, the principle of uniform strength may conflict with the demand for minimum costs; the principle of self-help may conflict with fail-safe behaviour (see Section 7.3.3); and the principle of uniform wall thickness chosen for the purpose of simplifying the production process [7.168] may conflict with the demand for lightweight construction or uniform strength.

These principles represent many strategies that are only applicable under certain conditions. In using them, designers must strike a balance between competing demands. To that end, the present authors have developed what they consider to be important embodiment design principles, which will now be presented. Most are based on energy flow considerations and, by analogy, they apply equally well to the flow of material and of signals.

7.4.1 Principles of Force Transmission

1. Flowlines of Force and the Principle of Uniform Strength

The problems solved in mechanical engineering generally involve forces and/or motions and their connection, change, variation or channelling, and involve the conversion of energy, material and signals. The generally applicable function "channel forces" includes the application of loads to, the transfer of forces between, and the transmission of forces through components and devices. Guidelines are provided in [7.168, 7.278]. In general, designers should try to avoid all sudden changes of direction in the flowlines of force—that is, in the force transmission path—caused by sharp deflections and abrupt changes of cross-section. The idea of "flowlines of force" aids the visualisation of the force transmission paths (load paths) through components and devices, and is analogous to flowlines in fluid mechanics. Leyer [7.167, 7.168] has dealt with the transmission of forces at some length, so we can dispense with a detailed discussion of the problem. Designers are advised to consult these important texts. Leyer, moreover, emphasises the complex interaction between the functional, embodiment and production aspects. The concept of force transmission can be summarised as described below.

Force transmission must be understood in a broad sense; that is, it must include the application, transfer and transmission of bending and twisting moments. First, it is important to remember that *external loads* applied to a component produce axial and transverse forces as well as bending and twisting moments at every section. These set up *stresses* (direct and shear) that produce *elastic* or *plastic deformations* (longitudinal, lateral (Poisson), and shear strains, along with bending and twisting).

The section dimensions transmitting the forces are obtained by "mental dissection" of the components at the point under consideration. The sum of the stresses over these sections produces internal forces and moments which must be in equilibrium with the external loads.

The stresses, determined at the relevant section, are then compared with the material properties of tensile strength, yield strength, fatigue strength, creep strength, etc., with due regard being paid to stress concentrations, surface finish and size effects.

The *principle of uniform strength* [7.278] aims, with the help of appropriate materials and shapes, to achieve uniform strength throughout a mechanical device over its anticipated operational life. Like the principle of lightweight construction [7.167], it should be applied whenever economic circumstances allow.

This important consideration often misleads designers into neglecting the deformations (strains) associated with the stresses. It is, however, these very deformations that often throw light on the behaviour of components and tell us what we need to know about their integrity (see Section 7.4.1).

2. Principle of Direct and Short Force Transmission Path

In agreement with Leyer [7.168, 7.208] we consider the following principle to be of great importance:

- If a force or moment is to be transmitted from one place to another with the *minimum possible deformation*, then the shortest and most direct force transmission path is the best.

This principle, which leads to the minimum number of loaded areas, ensures:

- minimum use of material (volume, weight)
- minimum deformation.

This is particularly true if it is possible to solve a problem using tensile or compressive stresses alone, because these stresses, unlike bending and torsional stresses, produce smaller deformations. When a component is in compression, however, special attention must be paid to the danger of buckling.

If, on the other hand, we require a flexible component capable of *considerable elastic deformation*, then a design using bending or torsional stresses is generally the more economical.

The principle is illustrated in Figure 7.27—the mounting of a machine frame on a concrete foundation—where different requirements demand supports with different stiffnesses. This, in turn, has repercussions on the operational behaviour of the machine: different natural and resonant frequencies, modified response to additional loads, etc. The more rigid solutions are obtained with minimum material and space requirements by means of a short support under compression; the most flexible solution by means of a spring, which transmits the force in torsion. If we look at other design solutions, we find many examples of the same principle: for example, in the torsion bar springs of motor cars, or in flexible pipes that rely on bending or torsional deformations.

The choice of means thus depends primarily on the nature of the task; that is, on whether the force transmission path must be designed for durability with

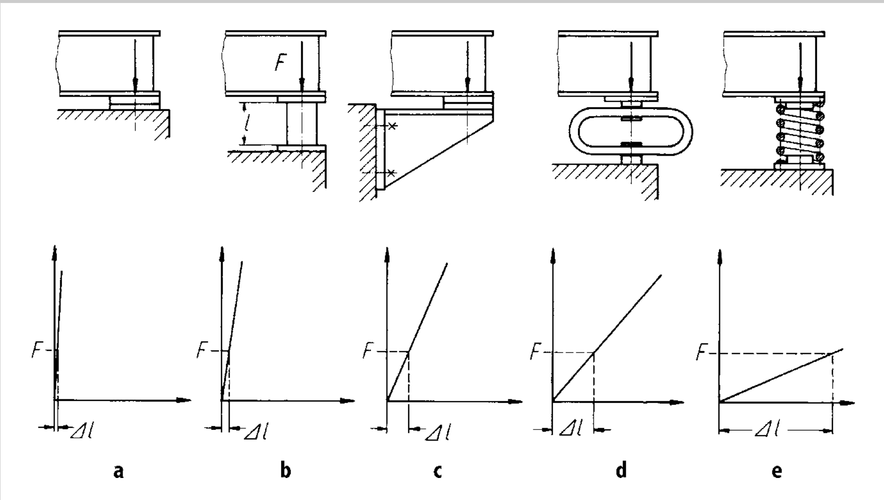


Figure 7.27. Supporting a machine frame on a concrete foundation: **a** very rigid support due to short force transmission path and low stress on the baseplates; **b** longer force transmission path, but still a rigid support with tubes or box sections under compression; **c** less rigid support with pronounced bending deformation (a stiffer construction would involve the greater use of materials); **d** more flexible support under bending stresses; **e** very flexible support using a spring, which transmits the load in torsion. This can be used to alter the resonance characteristics

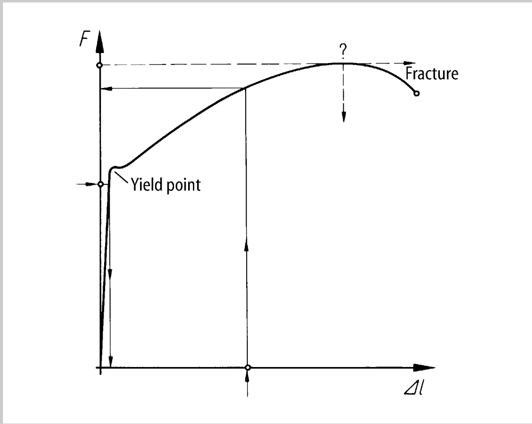


Figure 7.28. Force deformation diagram of tough materials. *Arrows* indicate the cause–effect relationships

maximum stiffness, or whether certain force–deformation relationships must be satisfied first and durability can be treated as a subsidiary problem.

If the *yield point is exceeded*, then the following have to be taken into consideration (see Figure 7.28):

- When a component is *loaded by a force*, it is invariably subjected to deformation. If the yield point is exceeded, then the linear-elastic relationship between the force and the deformation no longer holds. Relatively small changes in the force near the peak of the force–deformation curve may produce unstable conditions leading to fracture, because the load-bearing cross-section may be reduced more rapidly than the strength is increased due to strain hardening. Examples are tie rods, centrifugal inertia forces on a disc, and weights on a rope. The necessary safety precautions must always be taken.
- When a component is *deformed*, then a reaction force is set up. So long as the impressed deformation does not change, the force and the stress remain unchanged as well. If the peak is not reached, the component remains stable so that the yield point can be exceeded without danger. Beyond the yield point, a large change in deformation will lead to only a small change in the force. Admittedly, any preload must not be augmented with further operational loads in the same sense, otherwise the conditions described above will prevail. Further requirements are the use of tough materials and the avoidance of a build-up of multiaxial stresses in the same sense. Examples are highly distorted shrink-fits, preloaded bolts and clamps.

3. Principle of Matched Deformations

Designs matched to the flowlines of force avoid sharp deflections of the transmission path and sudden changes in cross-section, thus preventing the uneven distribution of stresses with high stress concentrations. A visualisation of the flowlines of force, though very graphic, does not always reveal the decisive factors involved. Here, too, the key is the deformation of the affected components.

The principle of matched deformations states that related components must be designed in such a way that, under load, they will deform *in the same sense* and, if possible, *by the same amount*.

As an example, let us take soldered or glued connections in which the solder or adhesive layer has a different modulus of elasticity from that of the material to be joined. Figure 7.29a illustrates the resulting deformation [7.181]. The deformations and the thickness of the solder or adhesive layers have been greatly exaggerated. The load F, which is transmitted across the junction of parts 1 and 2, produces distinct deformations in the overlapping parts, the adhesive layer being subjected to particularly marked deformation near the edges due to differences in the relative deformation of parts 1 and 2. While part 1 bears the full load F at the upper edge of the adhesive layer and is therefore stretched, part 2 does not yet bear a load. The relative shift in the adhesive layer sets up a local shear stress that exceeds the calculated mean value.

A particularly unsatisfactory result is shown in Figure 7.29b where, as a result of opposite and unmatched deformations of parts 1 and 2, the deformation in the adhesive layer is considerably increased. This example makes it clear why provision should be made for deformations to take place in the same sense and, if possible, to be equal in magnitude. Magyar [7.177] has made a mathematical study of the relationships between load and shear stress: the result is shown qualitatively in Figure 7.30.

The same phenomenon also occurs between nuts and bolts in bolted joints [7.328]. The nut (see Figure 7.31a) is in compression and the bolt is in tension, that is, they are deformed in the opposite sense. In the modified nut (see

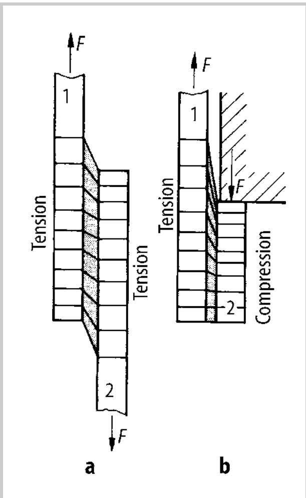


Figure 7.29. Overlapping adhesive or solder joint with strongly exaggerated deformation from [7.181]: **a** Parts 1 and 2 deformed in the same sense; **b** Parts 1 and 2 deformed in the opposite sense

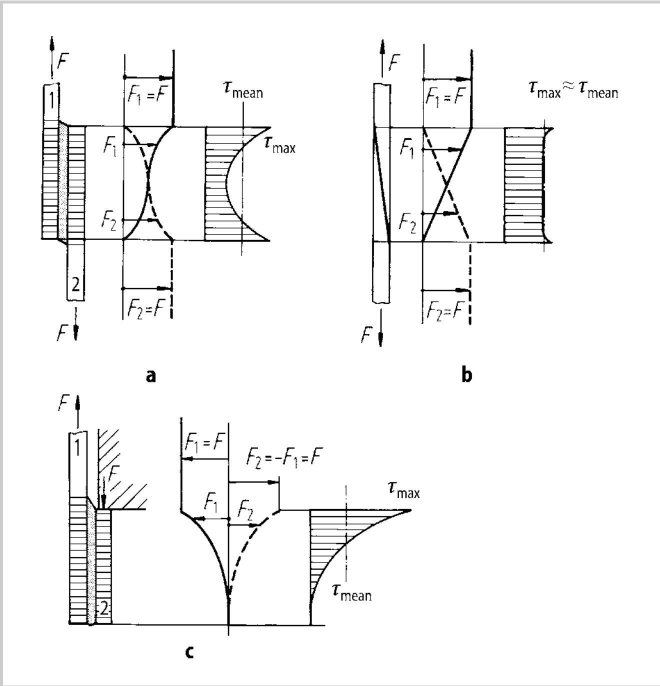


Figure 7.30. Distribution of forces and shear stresses in overlapping joints with layer of adhesive or solder, after [7.177]: **a** overlapped on one side (bending stress neglected); **b** spliced with linearly decreasing thickness; **c** pronounced "deflection of the flowlines of force" with deformations in the opposite sense (bending stress neglected)

Figure 7.31b) a deformation in the same sense is set up in the leading threads, which gives rise to a smaller relative deformation and hence a more even distribution of the load borne by individual threads. Wiegand [7.328] has been able to demonstrate this effect by showing that such nuts have a longer service life. Paland [7.214] has shown more recently that standard nuts are not as unsatisfactory as Maduschka [7.175] has suggested, because the moment $F \cdot h$ produces additional outward deformations of the nut at the contact surface and thus relieves the leading threads of their load. The load-relieving deformation of the nut due to this moment and also to the bending of the threads can be increased considerably by using material with a lower modulus of elasticity. If, on the other hand, the load-relieving deformations are resisted by a very stiff nut or a very small lever arm h, then the type of load distribution described by Maduschka would ensue.

As a further example, let us take a shaft–hub connection formed by a shrink fit. In essence, this too involves the deformation of two components [7.125]. In transmitting the torque, the shaft experiences a torsional deformation that decreases as the torque is transferred to the hub. The hub, for its part, is deformed in accordance with the transmitted torque.

Figure 7.32a shows that the maximum relative deformation occurs at A. In the case of alternating torques, this may lead to fretting corrosion; moreover, the right-

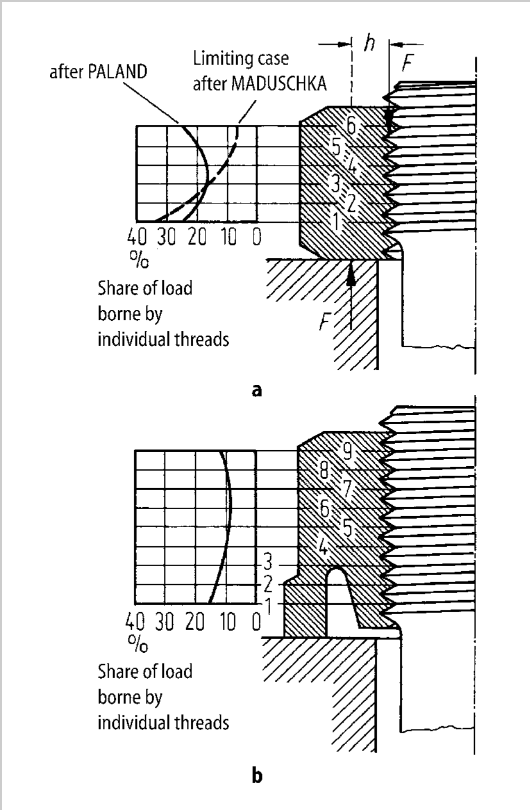


Figure 7.31. Nut shapes and load distribution, after [7.328]: **a** standard nut: limiting case after Maduschka [7.175] and case after Paland [7.214] allowing for deformation due to moment $F \cdot h$; **b** modified nut with matched deformations in the tension part

hand end, to all intents and purposes, contributes nothing to the transfer of the torque.

The solution shown in Figure 7.32b is much better because the resulting deformations are in the same sense. The best solution appears when the torsional stiffness of the hub is matched to that of the shaft. The transfer of torque then takes place along the whole length of the connection, ensuring uniform distribution of force flowlines and thus avoiding stress concentrations.

Even if the shrink fit were replaced with a keyed connection, the layout depicted in Figure 7.32a would, because the torsional deformations are in the opposite sense, set up very high contact stresses in the neighbourhood of A. The layout depicted in Figure 7.32b will, on the other hand, ensure an even contact stress distribution because the deformations are in the same sense [7.188].

The principle of matched deformations can also be applied to bearings, as in Figure 7.33. The embodiment of the bearings should ensure matched deformations between bearing and shaft, or provide for adjustment possibilities.

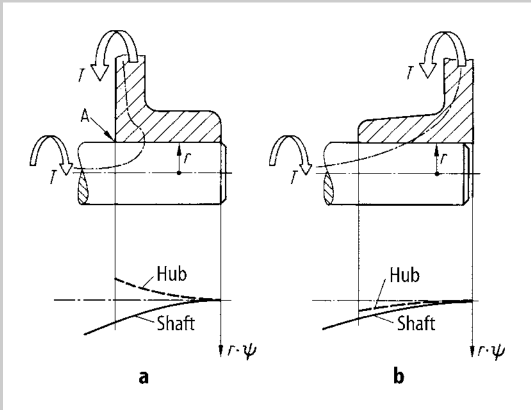


Figure 7.32. **a** Shaft–hub connection with strong "force flowline deflection". Torsional deformations of shaft and hub in opposite sense (ψ = angle of twist). **b** Shaft–hub connection with gradual "force flowline deflection". Torsional deformations of shaft and hub in the same sense

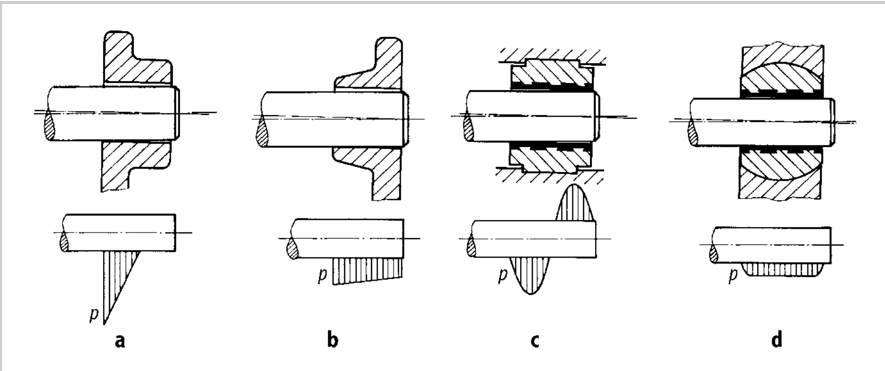


Figure 7.33. Force transmission in bearings: **a** edge compressing because of insufficient adaptation of the bearing to the deformed shaft; **b** more even bearing pressure because of matched deformations; **c** lacking adjustment to shaft deformation; **d** more even bearing pressure because of adaptability of bearing bush

The principle of matched deformations must be taken into account, not only in the transfer of forces from one component to another, but also in the division or combination of forces or moments. A well-known problem is the simultaneous propulsion of wheels that have to be placed at a considerable distance from one another, for instance in crane drive assemblies. In the layout shown in Figure 7.34a, the left side has a relatively high torsional stiffness due to the short force transmission path, and the right side a relatively low torsional stiffness because of its greater path length. When the torque is first applied, the left wheel will be set in motion, while the right wheel remains stationary until the right hand part of the shaft has twisted sufficiently to transmit the torque. The drive assembly has a tendency to run skew.

It is essential to provide the same torsional stiffness to both parts of the shaft so as to ensure an appropriate division of the initial torque. This can be achieved

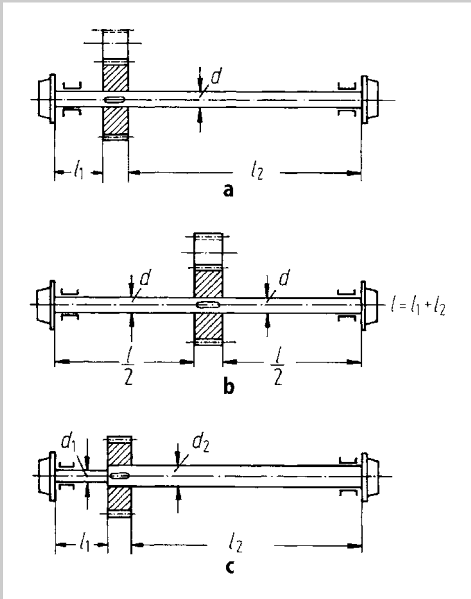


Figure 7.34. Application of the principle of matched—here equal—deformations in crane drives: **a** unequal torsional deformation of lengths l_1 and l_2; **b** symmetrical layout ensures equal torsional deformation; **c** asymmetrical layout with equal torsional deformation due to adaptation of torsional stiffnesses

in two distinct ways if the input torque is taken in one position only: either by symmetrical layout (see Figure 7.34b); or by adaptation of the torsional stiffness of the appropriate parts of the shaft (see Figure 7.34c).

4. Principle of Balanced Forces

Those forces and moments that serve the function directly, such as the driving torque, the tangential tooth force, and the load torque in a gearbox, can, in accordance with the definition of a main function, be described as *functionally determined main forces*.

In addition, there are many forces or moments that do not serve the function directly but that cannot be ignored, for instance:

- the axial force produced by a helical gear
- the force resulting from a pressure difference, for instance across the blades of a turbine or across a control valve
- tensile forces for producing a friction connection
- inertia forces due to linear acceleration or rotation of components
- fluid flow forces, inasmuch as they are not the main forces.

Such forces and moments accompanying the main ones are called *associated forces*, and may either produce a useful auxiliary effect or else appear merely as an unwanted effect that has be taken into account.

Associated forces place additional loads on the components and require an appropriate layout, or must be taken up by further surfaces and elements, such as stiffening members, collars, bearings, etc. As a result, weights are increased and further frictional losses may be incurred. For that reason, the associated forces must, whenever possible, be balanced out at their place of origin, thus obviating the need for a heavier construction or for reinforced bearing and transfer elements.

As has been shown in [7.204], this balance of forces is essentially ensured by two types of solution:

- balancing elements
- symmetrical layout.

Figure 7.35 shows how the associated forces can be balanced in a turbine, helical gears and a cone clutch, with the help of the principle of direct and short force

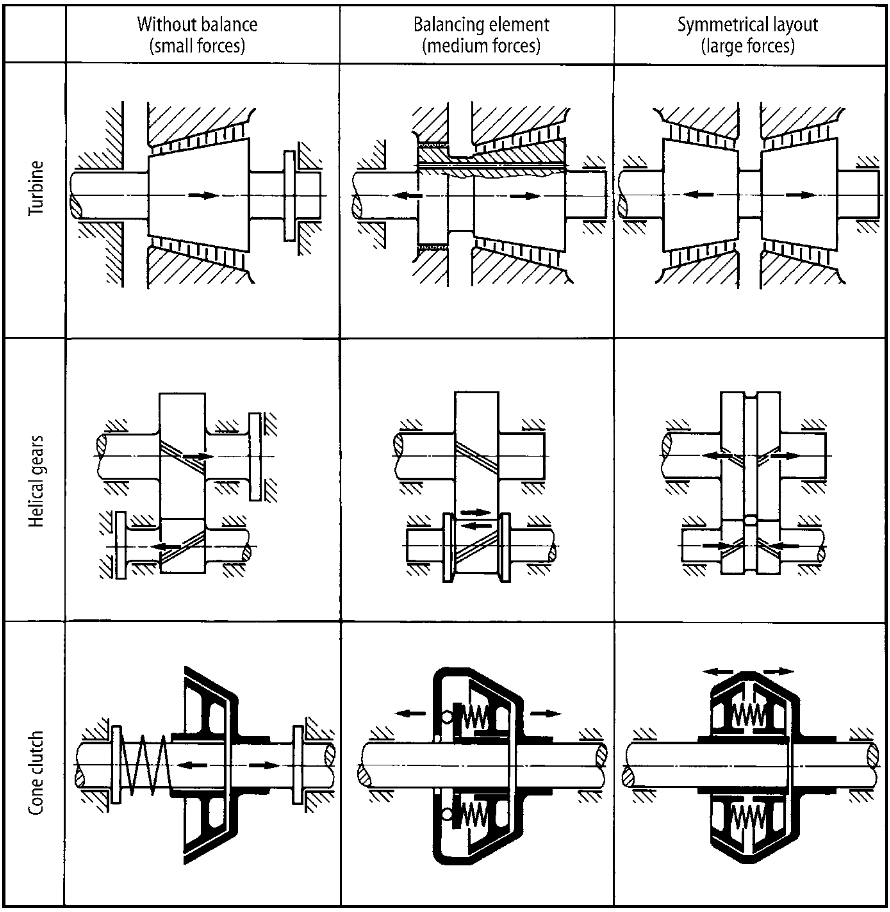


Figure 7.35. Fundamental solutions for balancing associated forces, illustrated via a turbine, helical gears and cone clutch

transmission path. As a result, no bearing position is loaded additionally and the designs are highly economical.

When it comes to the balancing of inertia forces, we find that a rotationally symmetrical layout is inherently balanced. The same solution principle is applied for reciprocating masses, as we know from automobile engineering. If the number of cylinders is too small to ensure a perfect balance, either special balancing elements, weights or shafts [7.228] are introduced, or cylinders are arranged symmetrically, as for instance in opposed cylinder engines.

As a general rule (which, however, can be ignored if there are overriding reasons for doing so), balancing elements should be chosen for relatively small or medium forces, and a symmetrical layout for relatively large forces.

5. Summary of Force Transmission Principles

Earlier we discussed the value of using the descriptive idea of flowlines of force when considering the transmission of forces during the embodiment of assemblies and components. The flowlines should fulfil the following criteria:

- the flowlines must always be closed
- the flowlines should, in general, be as short as possible, which can best be achieved by direct force transmission
- sharp deflections of the flowlines and changes in their "density" resulting from sudden changes in cross-section must be avoided.

In the case of complex force transmission situations, the definition or visualisation of a flowline envelope can be useful. This is the working zone outside of which the forces have no effect. The smaller the envelope, the shorter the force transmission paths. Figure 7.36 shows different concepts of a rotary bending test rig with the respective flowlines envelopes indicated.

The following principles complement the concept of flowlines:

- The *principle of uniform strength* which ensures, through the careful selection of materials and shapes, that each component is of uniform strength and contributes equally to the overall strength of a device throughout its service life.
- The *principle of direct and short force transmission path*, which ensures minimum volume, weight and deformation, and which should be applied particularly if a rigid component is needed.
- The *principle of matched deformations*, which ensures the matching of deformations of related components, so that stress concentrations are avoided and the function can be reliably fulfilled.
- The *principle of balanced forces*, which ensures, with the help of balancing elements or a symmetrical layout, that the associated forces accompanying the main ones are reacted as close as possible to their place of origin, so that material quantities and losses can be kept to a minimum.

In many situations, these principles cannot be applied to their full extent and often have to be applied in combination.

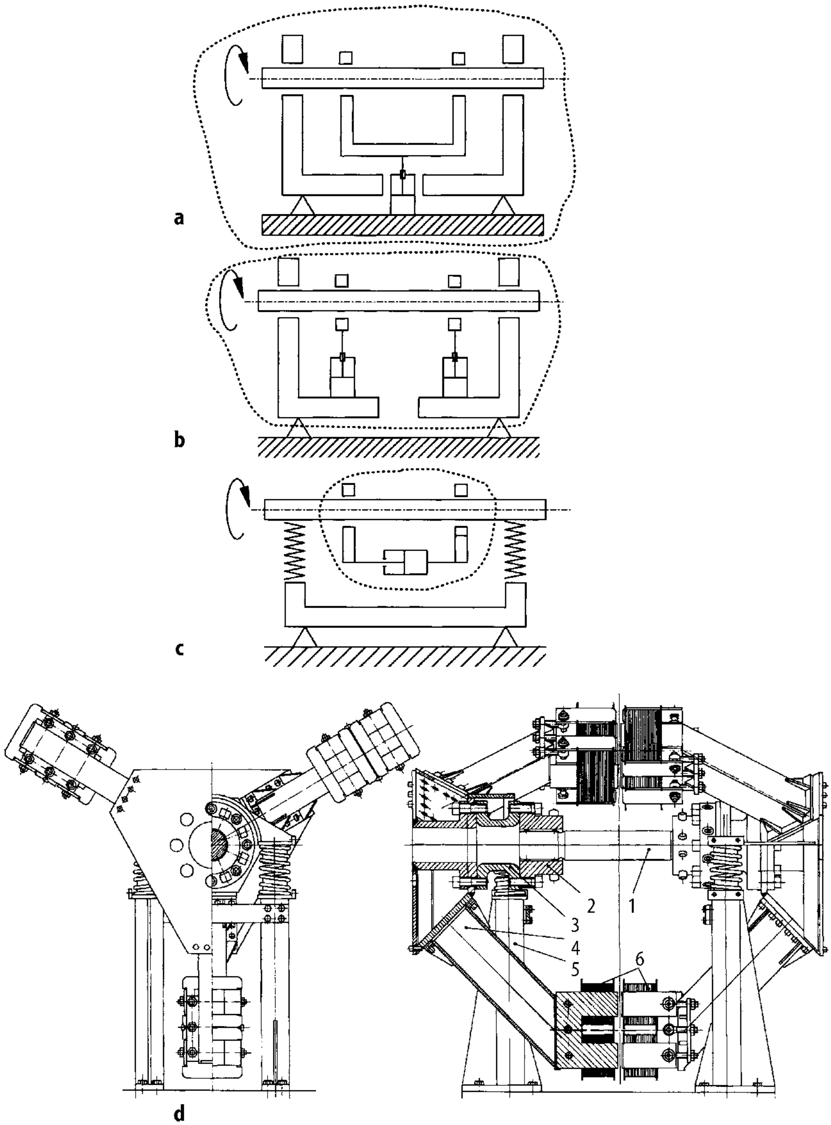


Figure 7.36. Force flow envelope (working zone of the forces) for a rotary bending test rig [7.330]. **a** Working zone includes the foundations; **b** working zone includes the supports; **c** working zone excludes the supports; **d** the test rig actually built using principle **c**, but with magnetic force excitation: 1 test shaft, 2 mounting flange, 3 connector, 4 support arm, 5 foundation supports, 6 magnet pair

7.4.2 Principle of the Division of Tasks

1. Assignment of Subfunctions

Even during the setting up and variation of the function structure, it is important to determine to what extent several functions can be replaced by a single one, or whether one function can be subdivided into several subfunctions (see Section 6.3).

These questions reappear in the embodiment phase, when the problem is to fulfil the requisite functions with the choice and assignment of suitable function carriers. We ask:

- Which subfunctions can be fulfilled with one function carrier only?
- Which subfunctions must be fulfilled with the help of several, distinct function carriers?

So far as the number of components and the space and weight requirements are concerned, a single function carrier fulfilling several functions would, of course, be the best. In terms of the production and assembly processes, however, this may prove disadvantageous, if only because of the complicated shape of the resulting component. Nevertheless, for economic reasons, an attempt should always be made to fulfil several functions with a single function carrier.

Numerous assemblies and components can fulfil several functions simultaneously or successively, as in the following examples:

- A shaft on which a gearwheel has been mounted transfers the torque and the rotating motion simultaneously, and, at the same time, takes up the bending moments and shear forces resulting from the normal tooth force. It also locates the gears axially and, in the case of helical gears, carries the axial force components from the teeth. In conjunction with the body of the gearwheel, it provides sufficient stiffness to ensure correct mating of the teeth.
- A pipe flange connection makes the connection and separation of the pipes possible, ensures the sealing of the joint, and transmits all forces and moments in the pipe resulting from residual tension, from thermal expansion and from unbalanced pipe loads.
- A turbine casing provides the appropriate inlet and outlet flow areas for the fluid, provides a mounting for the stationary blades, transmits the reaction forces to the foundation, and ensures a tight seal.
- A wall of a pressure tank in a chemical plant must combine a retaining with a sealing function and stave off corrosion, while not interfering with the chemical process.
- A deep groove ball bearing, apart from its centering task, transmits both radial and axial forces and occupies a relatively small volume.

The combination of several functions in a single function carrier may often prove economically advantageous, but may have certain drawbacks. These do not usually appear unless:

- the capacity of the function carrier has to be increased to the limit with respect of one or several functions
- the behaviour of the function carrier must be kept absolutely constant in one important respect.

As a rule, it is impossible to optimise the carrier of several combined functions. Instead, designers have recourse to the *principle of the division of tasks* [7.207], by which a special function carrier is assigned to every function. Moreover, in borderline cases, it may even be useful to distribute a single function over several function carriers.

The principle of the division of tasks:

- allows much better exploitation of the component concerned
- provides for greater load capacity
- ensures unambiguous behaviour, and hence fosters the basic rule of clarity (see Section 7.3.1).

This is because the separation of tasks facilitates optimum design in respect to every subfunction and facilitates more accurate calculations. In general, however, the constructional effort becomes correspondingly greater.

To determine whether the principle of the division of tasks can be usefully applied, the *functions must be analysed* with a view to determining if the simultaneous fulfilment of several functions in one carrier introduces constraints or mutual interferences. If it does, then it is best to settle for individual function carriers.

2. Division of Tasks for Distinct Functions

Examples from various fields illustrate the advantage of the division of tasks for distinct functions.

In large gearboxes, as found for instance between a turbine and a generator, it is advisable, because of thermal expansion of the foundations and bearings and also because of the torsional oscillations, to use a radially and torsionally flexible shaft whilst maintaining the shortest possible axial length on the output side [7.203]. However, because of the forces between the gear teeth, the transmission shaft must be as rigid as possible. Here the principle of the division of tasks leads to the following arrangement: the gearwheel is fitted to a stiff hollow outer shaft with the shortest possible distance between the bearings, while the radially and torsionally flexible component takes the form of an inner torsion shaft (see Figure 7.37).

Modern pressure-fed boilers are built with a membrane wall, as shown in Figure 7.38. The furnace must be gas-tight. Moreover, optimum heat transfer to the water demands thin walls with large surface areas. Beyond that, thermal expansion and pressure differences between the furnace and its environment must also be taken into consideration, and so must the weight of the walls. This complex problem is solved with the help of the principle of the division of tasks. The tubular walls with their welded lips constitute the sealed furnace. The forces resulting

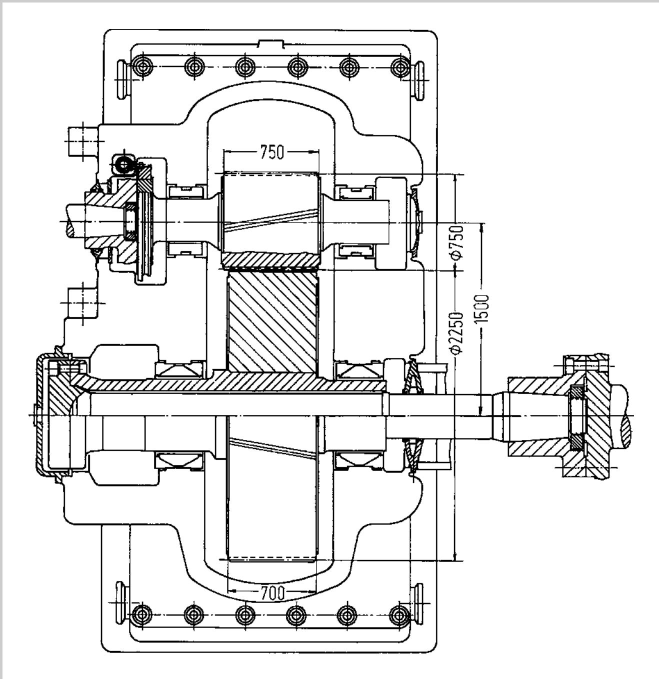


Figure 7.37. Large gearbox with an output torsion shaft; the bearing forces are transmitted over a stiff hollow shaft; the inner torsion shaft is radially and torsionally flexible, after [7.203] (Siemens-Maag)

from the pressure differences are transferred to special supports outside the heated area, which also carry the weight of the—usually suspended—walls. Articulated arms between the tubular wall and the supports allow for unimpeded thermal expansion. Thus every part can be designed in accordance with its special task.

The clamp connection in a superheated steam pipe shown in Figure 7.39 has also been designed based on the principle of the division of tasks. The sealing and load-carrying functions are assigned to different function carriers: the sealing function is performed by the welded membrane seal, which is axially loaded by the tension in the clamp. Tensile forces or bending moments should not be carried by the seal, whose function and durability would thereby be destroyed, so the load-carrying function is performed by the clamp which, in turn, is designed on the principle of the division of tasks. The clamp is made up of segments, which transmit forces and bending moments by means of a close-tolerance fit, and shrink rings hold the clamp segments together via friction in a simple and effective manner. Every part can he optimally designed for its particular task and is easily analysed.

The casings of turbines must ensure a tight seal under all operational and thermal conditions if they are to conduct the working fluid with minimum loss and turbulence. They must also provide an annular area and a support for the stationary blades. During temperature changes, sectioned casings with an axial flange have a particular tendency to distort and to lose sealing power due to

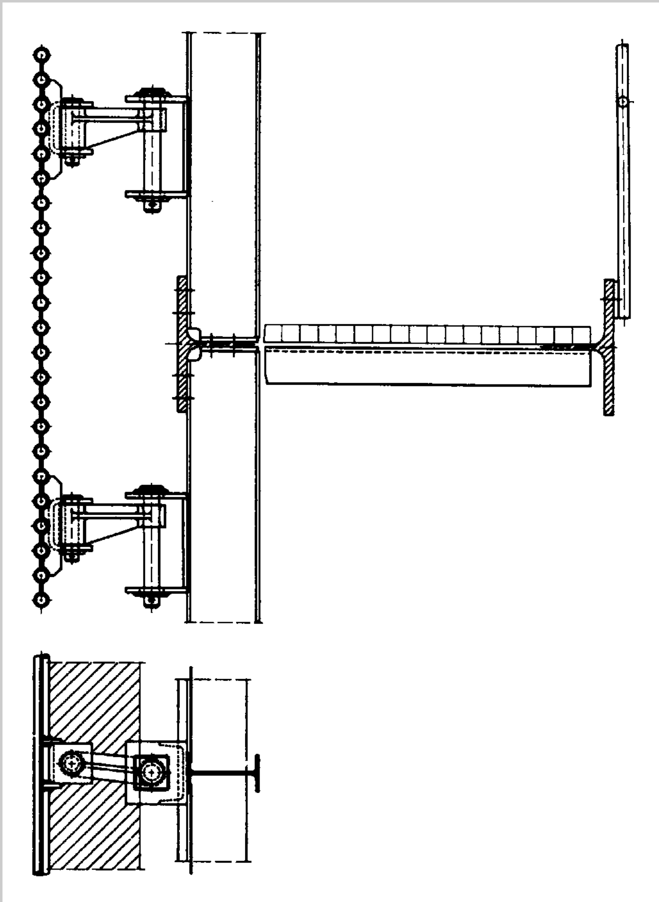


Figure 7.38. Section of boiler with membrane walls and separate supports (Babcock)

marked changes in shape at the inlet and outlet [7.224]. This effect can be offset by a separate blade carriers, that is, by a division of tasks. The annular area and stationary blade attachment can be designed regardless of the larger casing with its inlet and outlet sections. The outer casing can then be designed exclusively for durability and sealing (see Figure 7.40).

A further example is provided by the synthesis of ammonia, which involves feeding nitrogen and hydrogen into a container under high pressures and temperatures. If the hydrogen were allowed to come into direct contact with a ferritic steel container, it would penetrate into and decarbonise the latter, producing decomposition at the grain boundaries with the formation of methane [7.117]. The solution is again based on the division of tasks. The sealing function is provided by an inner casing of austenitic steel which is resistant to hydrogen, while support and strength are provided by a surrounding pressure chamber constructed from high-tensile ferritic steel, which is not resistant to hydrogen.

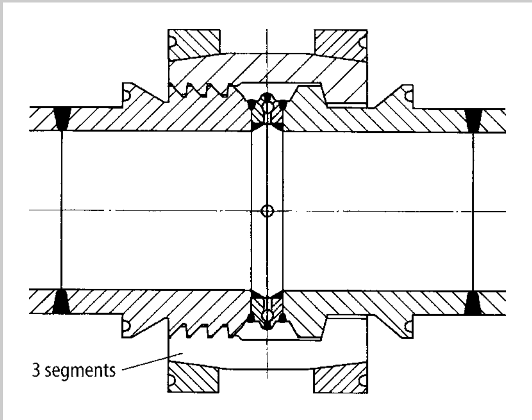


Figure 7.39. Clamp connection in a superheated steam pipe (Zikesch)

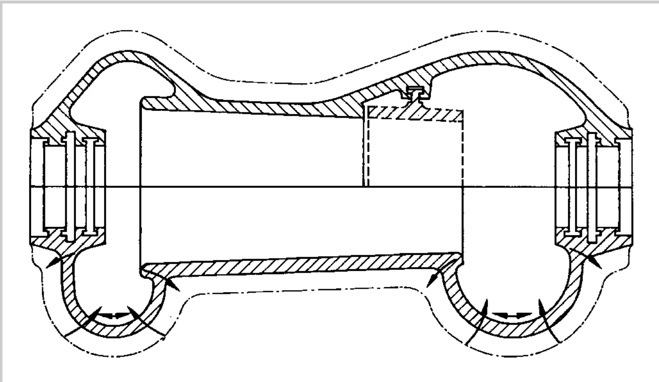


Figure 7.40. Axially divided turbine housing, after [7.224]: lower half conventional; upper half with separate blade carrier

In the electrical circuit-breaker illustrated in Figure 7.41, two or even three contact systems are provided. The breaker contacts 1 take the arcing current during the closing or opening of the switch, and the main contacts 3 carry the current under normal conditions. The breaker contacts 1 are subject to burning—that is, to wear and tear—and must be designed accordingly, while the main contacts must be designed to carry the full working current.

The division of tasks is also illustrated in Figure 7.42: the Ringfeder connectors carry the torque, while the corresponding cylindrical surfaces ensure the central location and seating of the pulley, which the Ringfeder connector cannot provide by itself when high accuracy is required.

A further example is provided by the design of rolling element bearings in which the service life of the locating bearing is increased by the clear separation of the transmission paths of radial and axial forces (see Figure 7.43). The outer race of the deep-groove ball bearing is not supported radially, and hence transmits axial forces only, while the roller bearing transmits radial forces only.

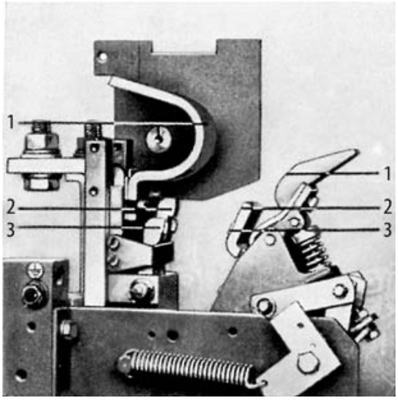


Figure 7.41. Arrangement of contacts in a circuit breaker (AEG): 1 breaker contacts; 2 intermediate contacts; 3 main contacts

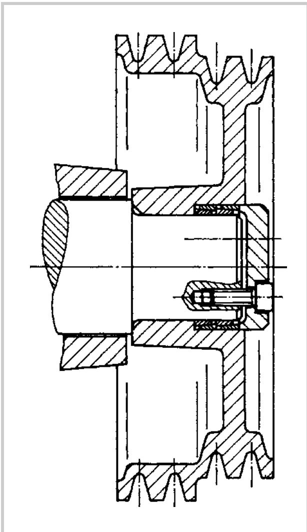


Figure 7.42. Ringfeder connector plus centralising surfaces

The principle of the division of tasks has been applied consistently to the construction of composite flat belts. They are made up, on the one hand, of a synthetic material capable of carrying high tensile loads and, on the other hand, of a chrome leather layer on the contact surface which provides a high coefficient of friction for the transfer of the load.

Yet another example is provided by the rotor blade attachment in a helicopter (see Figure 7.18).

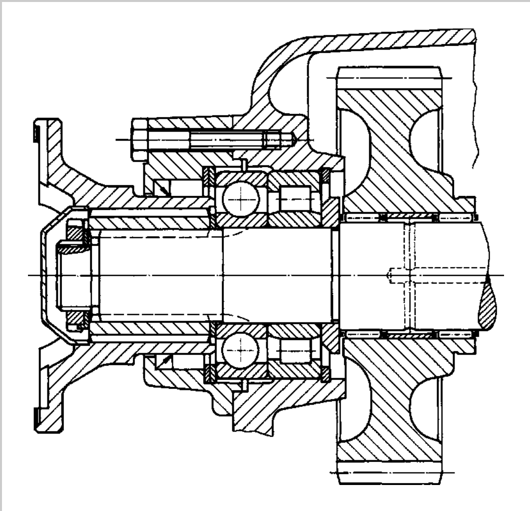


Figure 7.43. Locating bearing with separate transmission paths for radial and axial forces

3. Division of Tasks for Identical Functions

If increases in load or size reach a limit, a single function can be assigned to several, identical function carriers. In other words, the *load can be divided* and then recombined later. There are numerous examples of this.

The load capacity of a V-belt cannot be increased at will by increases in its cross-section (number of load-carrying strands per belt) because, for a given pulley diameter, an increase in the belt height *h* (see Figure 7.44) leads to an increase in the bending stress. As a result of the ensuing deformation, the rubber (which has hysteresis properties and is also a poor conductor of heat) becomes overheated and this reduces its life. A disproportionally wide belt, on the other hand, loses the stiffness needed to take up the normal forces acting on the wedge-shaped surfaces of the pulley. An increase in load-carrying capacity can, however, be obtained by dividing the overall load into part loads, each appropriate to the load limit and normal life of the individual belts (multiple arrangement of parallel V-belts).

The coefficient of thermal expansion of superheated steam pipes made of austenitic steel is approximately 50% higher than that of pipes made of the usual ferritic steel. Such pipes, moreover, are particularly stiff. At constant inner pressures and fixed material property limits, the ratio of outer to inner pipe diameter remains constant if the inner diameter is changed. However, while the throughput at constant flow velocities varies as the square of the inner diameter, the bending and torsional stiffnesses vary as its fourth power. The substitution of z pipe lines for a single large pipe would admittedly lead to increased pressure and heat losses for the same flow area, but would reduce the stiffness resisting thermal expansion by $1/z$. With four or eight pipelines, the individual reaction forces would then be no more than 1/4 or 1/8 of that present in a single pipe [7.29, 7.279]. In addition, the reduction in wall thickness leads to a reduction in thermal stresses.

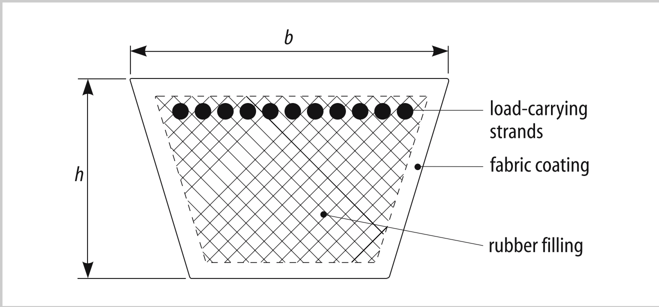


Figure 7.44. Cross-section of V-belt

Gearboxes, and epicyclic gearboxes in particular, make use of the principle of the division of tasks (or rather of forces) in the form of multiple meshing, which will increase the transmission capacity of the gearbox provided that the thermal effects can be kept within reasonable limits. In the symmetrical layout of epicyclic gearboxes based on the principle of balanced forces (see Section 7.4.1), even the bending moment in the shaft is eliminated because the forces produced by the gears cancel out. However, the torsional deformation is increased because of the greater load capacity (see Figure 7.45). In large gearboxes, this principle is applied to advantage in the form of multiple drives equipped with spur gears, which have external teeth only and hence are more easily produced. As Ehrlenspiel [7.96] has shown, it is possible to increase the load capacity with the number of force transmission paths, though not in direct proportion, because each step introduces a different flank geometry with a slightly greater flank loading. Basic arrangements are depicted in Figure 7.46.

One problem with the principle of the division of tasks is the uniform participation of all of the elements in the fulfilment of the function, that is, the provision of a *uniform distribution of forces or loads*. In general, this can only be achieved if:

- the participating elements adjust themselves automatically to balance out the forces
- appropriate flexibility is specially provided in the force transmission paths.

In the case of multiple V-belt drives, the tangential forces produce slight extensions of the belts which help to offset any dimensional errors in the lengths of the belts or in the pulleys, or any lack of parallelism in the shaft, and thus ensure equal load sharing.

In the case of the multiple pipeline discussed above, the individual pipe loss coefficients, the relationships between inflow and outflow, and also the geometry of the pipe layouts must be kept similar, or else the individual loss coefficients must be small and not greatly affected by the flow speeds.

In the case of multiple gears, either a strictly symmetrical arrangement must ensure equal stiffnesses and temperature distributions throughout the gearbox, or special flexible or adjusting elements [7.97] must ensure the equal participation of all of the components.

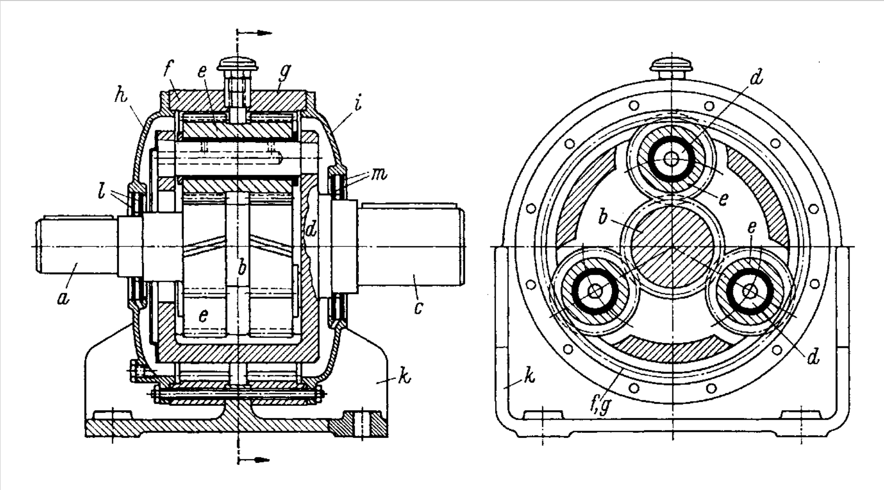


Figure 7.45. Epicyclic gearbox with balanced forces, after [7.97]

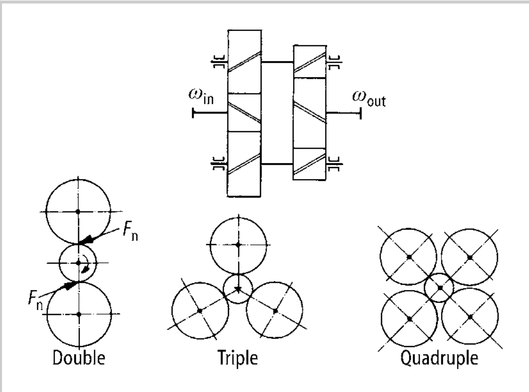


Figure 7.46. Basic arrangements of multiple gears, after [7.203]

Figure 7.47 illustrates a flexible arrangement. Further balancing components, such as elastic and articulated joints, are described in [7.97].

All in all, the principle of the division of tasks provides for increases in the maximum load capacity or for wider applications. By spreading tasks over several function carriers, we also gain a clearer picture of the relationship between forces and their effects, and, what is more, we can increase the output, provided only that a balanced division of forces is maintained by adjustable or self-regulating elements.

In supporting structures (such as bearing supports) where force transmission is divided, a more balanced load distribution can be achieved by adjusting the stiffness. During the stiffness analysis, the location and direction of the external forces have to be considered carefully, because they influence the deformation

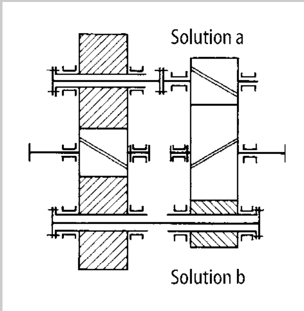


Figure 7.47. Balanced forces in multiple gears by means of flexible torsion shafts, after [7.203]

behaviour. This analysis can be facilitated by the use of Finite Element (FE) methods (see the principle of matched deformation in Section 7.4.1).

In general, the application of the principle of the division of tasks increases the number of components, which must be offset by greater overall economy or safety.

7.4.3 Principle of Self-Help

1. Concepts and Definitions

In the last section we discussed the principle of the division of tasks and showed how it could help to increase load capacity and provide a clearer definition of the behaviour of the components. To that end, we analysed the various subfunctions and assigned them to function carriers chosen such that they neither influence nor interfere with one another.

The same analysis can also be used in conjunction with the *principle of self-help* to achieve, through the appropriate choice of system elements and their arrangement, a mutual supportive interaction that improves the fulfilment of the function.

Under *normal conditions* (normal loading), self-help provides for greater effect by arranging the forces to work in the same direction as each other, or for relief by arranging the forces to offset each other. In *emergency situations* (overloading), self-help provides for greater protection and safety. In a self-helping design, the *overall effect* is made up of an initial effect and a supplementary effect.

The *initial effect* sets off the physical process required by the solution, but is insufficient on its own.

The *supplementary effect* is obtained from the functionally determined main forces (gearbox torque, sealing force, etc.) and/or from the associated forces (axial force produced by helical gears, centrifugal inertia force, force due to thermal expansion, etc.), provided, of course, that the two sets of forces are clearly correlated. A supplementary effect may also be obtained by appropriate changes to the type and distribution of the force transmission paths in order to increase load capacity.

The idea of formulating the self-help principle was first suggested by the Bredtschneider–Uhde self-sealing cover, which is particularly suitable for pressure vessels [7.237]. Figure 7.48 shows how it works. A relatively small force provided by the central bolt *2* suffices to press the cover *1* against the metal seal *5*. The initial effect of this force ensures that the parts make the proper contact. With increasing operational pressure, a supplementary effect is produced, which ensures that the sealing force between cover and tank is increased appropriately. The internal pressure thus provides the required sealing force automatically.

Inspired by this self-sealing solution, the principle of self-help was formulated in [7.206, 7.209] and further analysed and elaborated by Kühnpast [7.161].

It may be useful to specify the quantitative contribution of the supplementary effect S to the overall effect O in producing the degree of self-help:

$$\chi = S/O = 0 \dots 1$$

The gain from self-help solutions can be expressed in terms of one or several technical characteristics: efficiency, service life, use of materials, technical limit, etc. The self-help gain is defined as:

$$\gamma = \frac{\text{technical characteristic with self-help}}{\text{technical characteristic without self-help}}$$

Whenever the application of the self-help principle calls for a greater effort on the part of designers, then it must bring clear technical or economic advantages.

Identical design approaches may turn out to be *self-helping* or *self-damaging*, depending on the layout. Take the case of an inspection cover (see Figure 7.49). So long as the pressure inside the tank is greater than the pressure outside, the layout shown on the left is self-helping, because the pressure on the cover (supplementary effect) increases the sealing effect (overall effect) of the initial tension-screw force (initial effect).

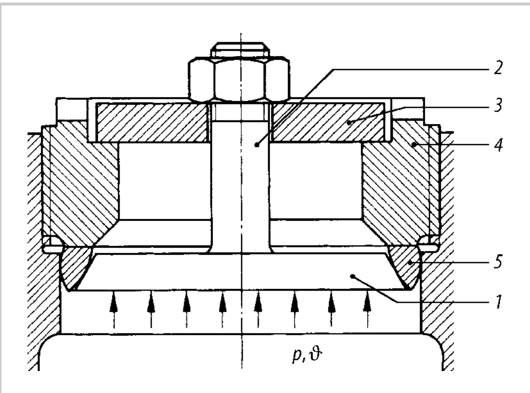


Figure 7.48. Self-sealing cover: *1* cover; *2* central bolt; *3* cross member; *4* element with sawtooth thread, *5* metal sealing ring; p = internal pressure, ϑ = temperature

The layout shown on the right, by contrast, is self-damaging because the pressure on the cover decreases the sealing effect O of the initial tension-screw force I. If, however, the tank were kept at below-atmospheric pressure, the left layout would be self-damaging, the right layout self-helping (see also Figure 7.50).

This example shows that the degree of self-help depends on the resultant effect: in the present case the effect on the sealing force resulting from the elastic forces, and not on the simple addition of the force exerted by the screw and the force acting on the cover. Figure 7.50 can also be considered to be a force–deformation diagram

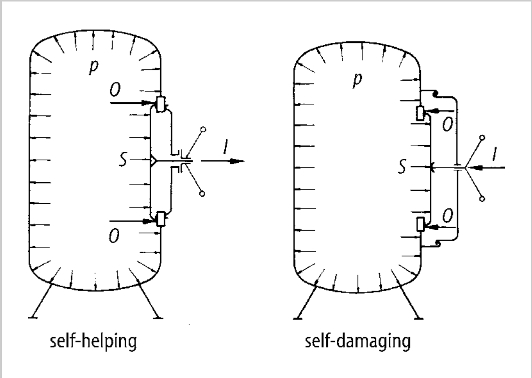


Figure 7.49. Layout of an inspection cover. I = initial effect; O = overall effect; p = internal pressure

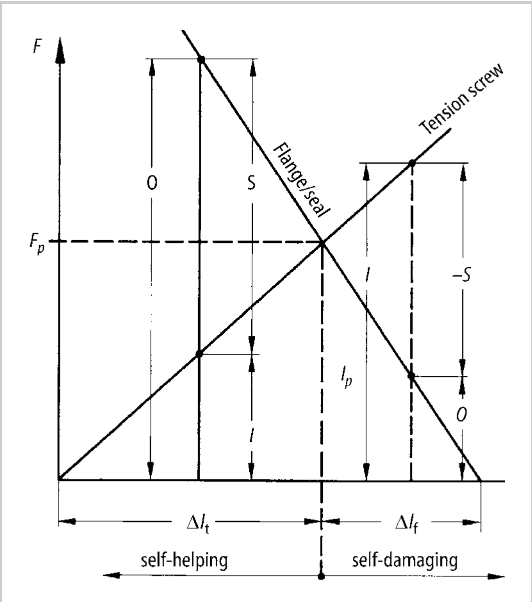


Figure 7.50. Force diagram for Figure 7.49: F = forces; F_p = preload; Δl = change in length; subscript t = tension screw; subscript f = flange/seal

Table 7.3. Summary of self-help solutions

	Normal load		Overload
Type of self-help	Self-reinforcing	Self-balancing	Self-protecting
Supplementary effect due to	Main and associated forces	Associated forces	Altered force transmission path
Important features	Main or associated forces act in the same sense as other main forces	Associated forces act in the opposite sense to main forces	Force transmission path altered by elastic deformation; limitation of function permissible

of a bolted connection with a preload and a working load. The conventional bolted flange connection may be called self-damaging inasmuch as, under operational conditions, the overall effect—that is, the flange sealing—becomes smaller than the preload. Also, the loading of the bolts is increased at the same time. If possible, therefore, only self-reinforcing arrangements that increase the overall effect while reducing the loading of the bolts should be chosen (Figs. 7.53a–d illustrate such arrangements).

For practical purposes, it is useful to classify self-helping solutions in accordance with Table 7.3.

2. Self-Reinforcing Solutions

In self-reinforcing solutions, the supplementary effect is obtained directly from a main or associated force and it adds to the initial effect to produce a greater overall effect.

This group of self-helping solutions is the most common. Under part-load conditions, it ensures greater service life, less wear, higher efficiency, etc., because the components are only loaded to an extent needed to fulfil the function at any particular moment.

As a first example, let us consider a continuously adjustable friction drive (see Figure 7.51).

The preload spring *a* presses the freely movable cup wheel *c* on the drive shaft *b* against the cone wheel *d*, thus providing the initial effect. Once a torque is applied, the roller follower *e* attached to shaft *b* is pressed against the cam *f* formed on the cup wheel *c*, where it produces a normal force F_n that can be resolved into a tangential force F_t and an axial force F_a, which, for its part, increases the contact force F_c applied to the cone wheel in a fixed proportion to the applied torque T:

$$F_a = T/(r \cdot \tan\alpha)$$

The force F_a represents the supplementary effect gained from the torque. The overall effect is obtained from the spring preload force F_p plus the axial force F_a,

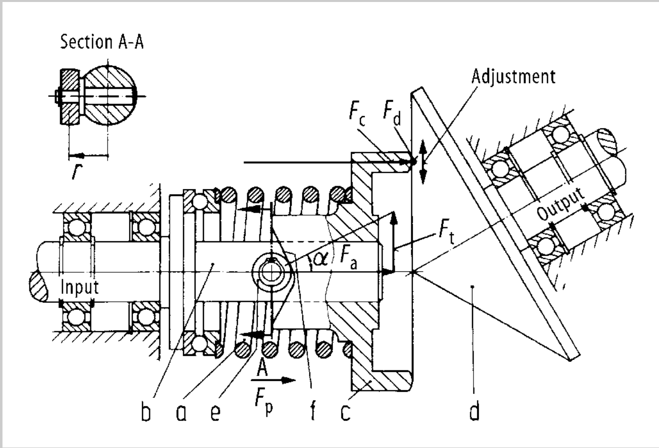


Figure 7.51. Continuously adjustable friction drive: *a* preload spring; *b* drive shaft; *c* cup wheel; *d* cone wheel; *e* roller follower; *f* cam formed on the cup wheel; *r* radius on which F_t and F_a act

which varies as the torque T (see Figure 7.52). The tangential driving force F_d on the cone, which determines the transmittable torque, is therefore:

$$F_d = (F_p + F_a) \cdot \mu$$

and the degree of self-help is:

$$\chi = S/O = F_a/(F_p + F_a)$$

It is obvious that the contact pressure between the wheels, which helps to determine the wear and the service life of the drive, must not exceed what is strictly necessary. A conventional solution (no self-reinforcement) would have demanded an axial force produced exclusively by the spring preload corresponding to the maximum torque, and would therefore have necessitated maximum pressure being applied to the contact area under all load conditions. As a result, the bearings would also have had to carry a considerably greater load, which would have led to a reduced service life or would have demanded a much heavier construction.

A rough calculation shows that if the actual loading is, say, 75% of the nominal maximum load, then the bearing load would be reduced by about 20% which, because of the exponential relationship of service life to load, can lead to the life of the bearings being doubled. In that case, with $n = 3$ the self-help gain with respect to the service life becomes:

$$\gamma_L = \frac{\text{Life with self-help}}{\text{Life without self-help}} = \left(\frac{C/0.8P}{C/P}\right)^n = 1.25^3 = 2$$

A typical example is provided by the SESPA drive [7.157].

Figure 7.53 shows various self-reinforcing layouts of contact surfaces loaded by bolts, in which the frictional forces are increased by the operational forces while the bolts themselves are off-loaded.

The application of the principle of self-help in the design of self-reinforcing brakes has been described by Kühnpast [7.161] and Roth [7.233]. Depending on the application, even self-damaging—and in this case self-weakening—solutions can prove interesting, inasmuch as they reduce the effect of variations of the coefficient of friction on the braking moment [7.107, 7.233].

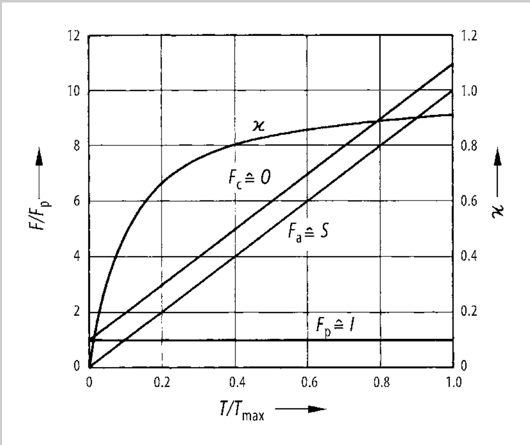


Figure 7.52. Degree of self-help (χ) and initial (I), supplementary (S) and overall (O) effect against the relative torque T/T_{max} for the friction drive (Figure 7.51)

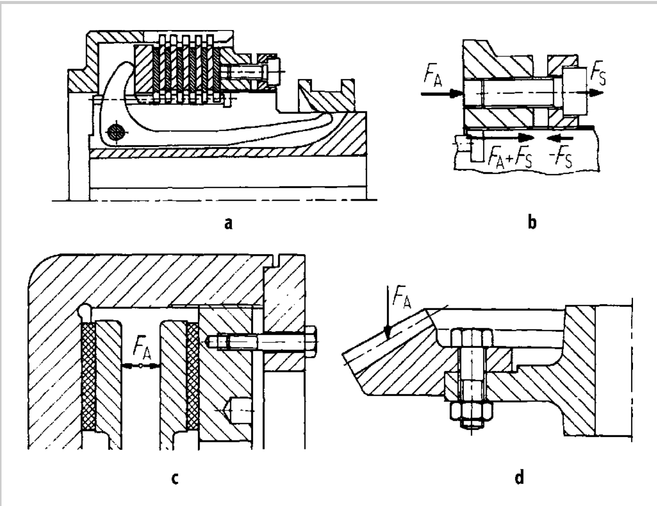


Figure 7.53. Self-helping bolted connections: **a** multiple disc clutch with adjustment ring; **b** force acting on the adjustment ring; **c** adjustable disc of two-disc friction clutch; **d** crown wheel attachment, symmetrical take-up of forces

Self-reinforcing seals (see Figure 7.54) provide us with further examples. In them, the operating pressure against which the seal has to be applied is used to produce the supplementary effect.

Finally, we must mention one case in which the supplementary effect is produced by an associated force. In hydrostatic axial bearings, the centrifugal inertia effect leads to an increase in oil pressure which, at high revolutions, will help to improve the load-carrying capacity, provided the heat can be removed (see Figure 7.55). The supplementary effect leads to an improvement in the load-carrying capacity due to the increased oil pressure resulting from the centrifugal effect alone; the overall effect is due to the load-carrying capacity of the combined static and dynamic pressures. According to Kühnpast [7.161], it should be possible at, say, 166 rev/s and $\chi = 0.38$, to obtain a gain in self-help of $\gamma = 1.6$ compared to static conditions.

The supplementary effect of another associated force, namely that caused by the effect of temperature on the shrink-fitted rings of a turbine, is discussed in [7.206].

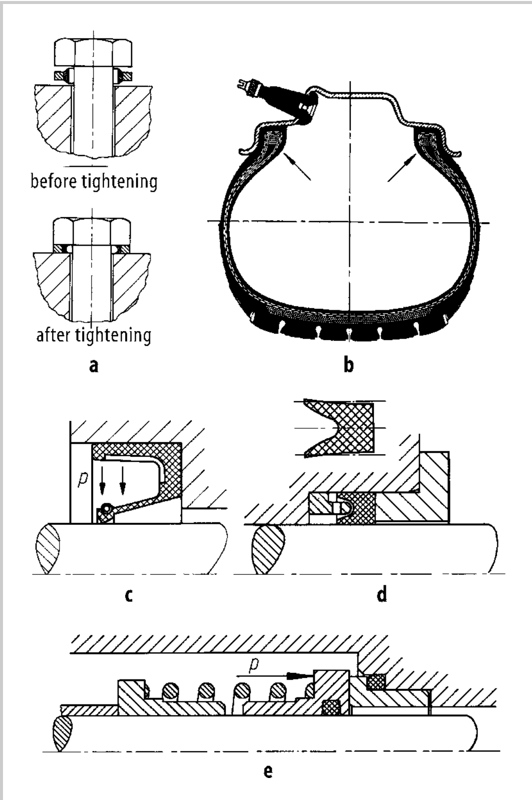


Figure 7.54. Self-reinforcing seals: **a** self-sealing washer; **b** tubeless tyre; **c** radial shaft seal; **d** sleeve seal; **e** sliding ring seal

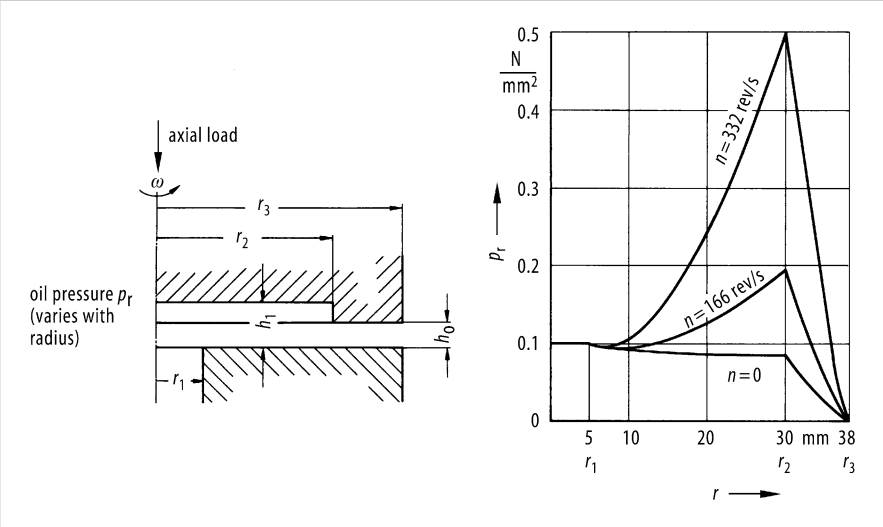


Figure 7.55. Self-help effect in hydrostatic axial bearings, after [7.161]

3. Self-Balancing Solutions

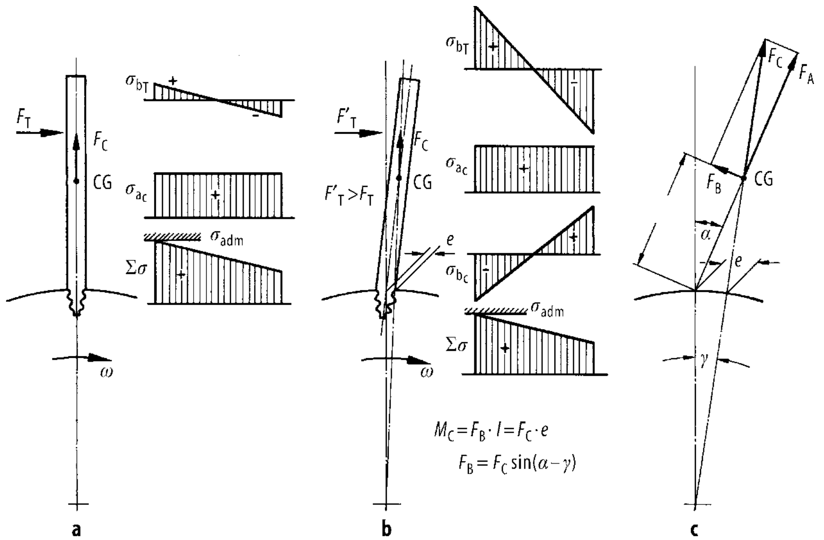
In self-balancing solutions, the supplementary effect is obtained from an associated force, and offsets the initial effect to produce an improved overall effect.

A simple example is provided by turbo-machines. A blade attached to a rotor is subject to a bending stress due to the tangential force acting upon it and also to an axial tensile stress due to the centrifugal inertia force. The two are additive and, because a certain stress must not be exceeded, the transferable tangential force is reduced (see Figure 7.56). If, however, the blade is attached at an angle, a supplementary effect is produced: an additional bending stress due to the centrifugal inertia force acting on the offset centre of gravity of the blade opposes the original bending stress and thus allows the application of a larger tangential force, that is, a greater overall effect. How far this balancing process can be carried depends on the aerodynamic and mechanical conditions.

A self-balancing effect can also be produced by allowing thermally induced forces (stresses) to oppose other forces (stresses), for instance, those resulting from excess or other mechanical loads (see Figure 7.57).

All of the examples we have given are intended to encourage the design of technical systems where:

- forces and moments with their resulting loads cancel out as far as possible, or
- additional forces or moments are produced in a clearly defined way so that it is possible to balance them out.



CG = Centre of gravity;
F_T = Tangential force; F_C = Centrifugal inertia force; σ_{bT} = Bending stress due to F_T; σ_{a_C} = Axial stress due to F_C;
σ_{b_C} = Bensing stress due to F_C; F_A = Axial component to F_C; F_B = Bending component of F_C

Figure 7.56. Self-balancing solution for turbine blades: **a** conventional solution; **b** leaning of the blade produces a balancing supplementary effect due to the additional bending stresses produced by the centrifugal inertia force (σ_{b_C}), which oppose the bending stresses caused by the tangential force (σ_{b_T}); **c** diagram of forces

4. Self-Protecting Solutions

In general, in the event of an overload, we do not want components to be destroyed, unless of course they have been deliberately designed as weak links. In particular, we try to protect components that are frequently subject to slight overloads. If special safety arrangements, for instance to limit the load, are not essential, then a self-protecting solution may prove advantageous. It will sometimes be simplicity itself.

Self-protecting solutions derive their supplementary effect from an additional different force transmission path that, in case of excess loading, is generally created after a given elastic deformation has taken place. As a result, the distribution of the flowlines of force is altered, which changes the nature of the loading and thereby increases the load-carrying capacity. Admittedly, in that case, the functional properties associated with normal conditions may become altered, limited or suspended.

The springs shown in Figure 7.58 have such self-protecting properties. In the case of excess loading, the spring elements, which are normally subject to torsional or bending stresses, will transmit the additional force directly by compressive stresses transmitted from coil to coil. The same effect may also be produced if the springs are shock-loaded (see Figure 7.58b).

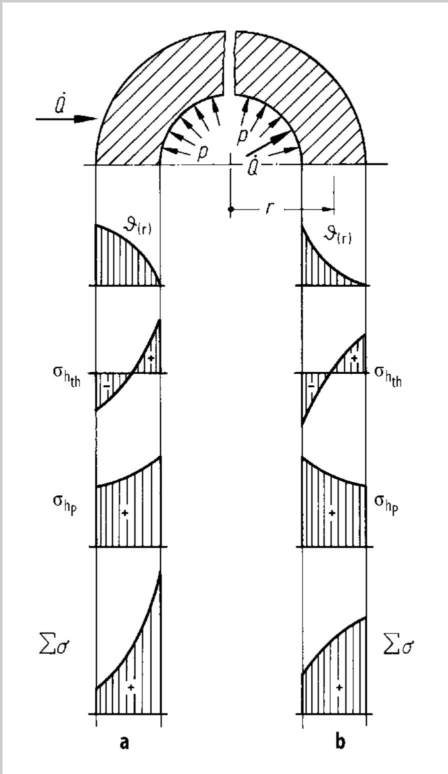


Figure 7.57. Hoop stresses in a thick-walled cylinder due to the internal pressure σ_{h_p}, and temperature differences at nearly steady heat flow $\sigma_{h_{th}}$: **a** nonbalancing solution, thermal stress is added to the maximum mechanical stress on the inner surface; **b** self-balancing solution, thermal stress opposes maximum mechanical stress on the inner surface

Figure 7.59 shows the layout of elastic couplings in which restriction of the spring movements provides additional and different force transmission paths with consequent loss of flexibility but with increased load-carrying capacity. The original springs are removed from the force transmission path. In Figure 7.59a, the load-carrying capacity of the bar springs is altered inasmuch as, besides the normal bending, a powerful shear force between the two halves of the coupling appears with overloads.

Figure 7.59b shows a coupling that, strictly speaking, may be considered a borderline case between a division of tasks and a self-protecting solution. The buffers will only take up forces in the case of overloading. In this case, the nature of the loading on the spring elements remains unchanged, although the force transmission path is altered after a given elastic deformation has taken place.

Kühnpast [7.161] also mentions cases in which there is an uneven stress distribution over a cross-section, and where plastic deformation can then be used for purposes of self-protection. In such cases, however, sufficiently tough materials and adequate dimensional stability are needed, and additive multiaxial stress situations are to be avoided.

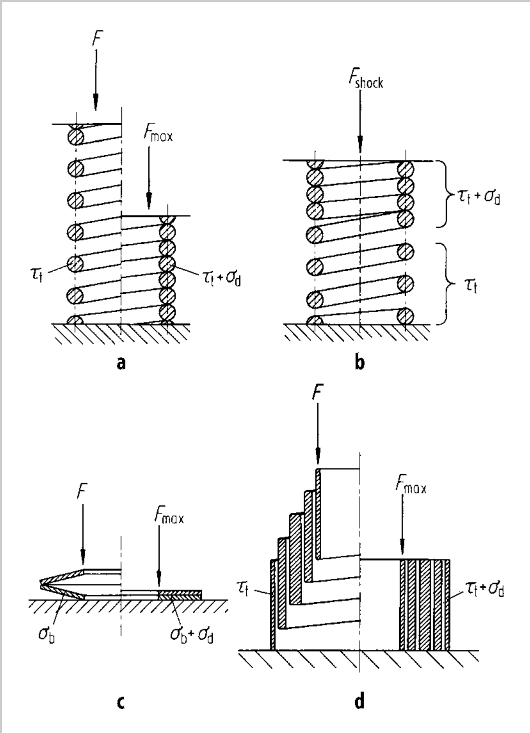


Figure 7.58. Self-protecting solution in springs: **a–d** force transmission path changed, the normal function is suspended or limited in the case of excess loading

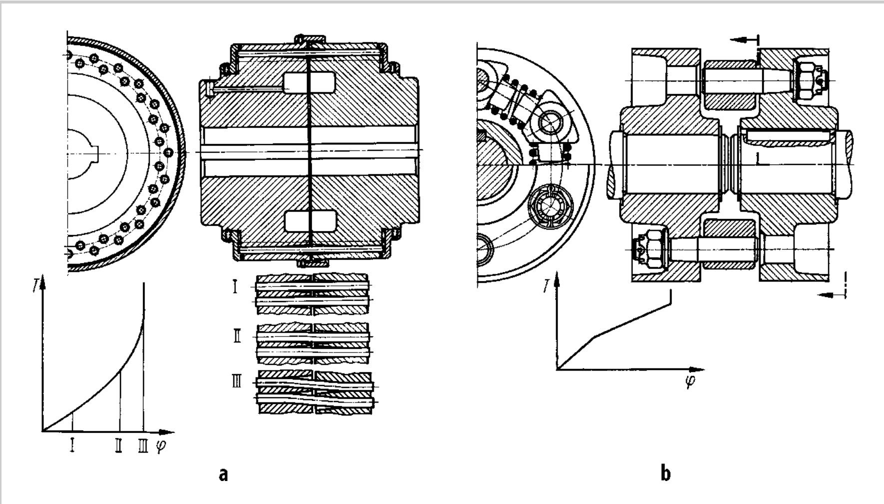


Figure 7.59. Self-protecting solution in couplings; change of force transmission paths with loss of elastic properties in case of overloading: **a** bar–spring coupling; **b** elastic coupling with coil springs and special buffers to take up the force in the case of overloading

It is hoped that the principle of self-help based on self-reinforcing, self-balancing and self-protecting solutions will encourage designers to examine every conceivable arrangement in an effort to arrive at an effective and economical solution.

7.4.4 Principles of Stability and Bi-Stability

We know the concepts of stable, neutral and unstable equilibrium from mechanics, as illustrated in Figure 7.60. When elaborating solutions, designers must always consider the effect of disturbances and try to keep the system stable by devising means whereby the disturbances can be made to cancel out, or at least to mitigate one another. If disturbances are self-reinforcing, we have unstable or bi-stable behaviour. This effect is desirable in certain solutions, where we speak of "planned instability".

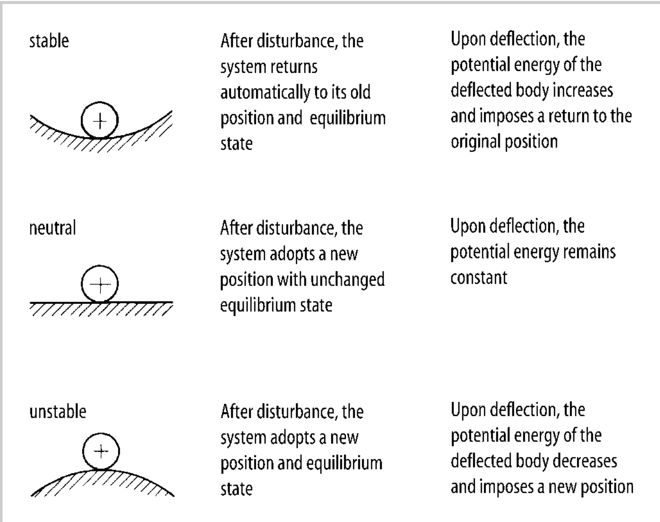


Figure 7.60. Characteristics of equilibrium states

1. Principle of Stability

By applying this principle, designers try either to ensure that disturbances cancel themselves out or to reduce their particular effects. Reuter [7.225] has discussed this subject at length and we shall now look at some of his examples.

In the design of pistons for pumps or regulating devices, the main objective is to achieve stable behaviour and minimum friction. Figure 7.61a shows the layout of a piston with unstable characteristics. Disturbances due to, say, inaccuracies in the cylinder bore can tilt the piston slightly and produce pressure distributions over the piston that encourage further tilting (unstable behaviour). Stable behaviour is

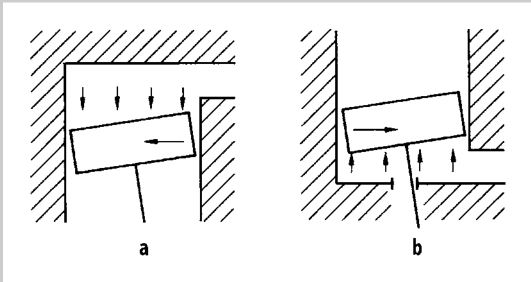


Figure 7.61. Piston in cylinder, tilted due to a disturbance, after [7.225]: **a** resulting pressure distribution produces an effect that increases the disturbance (unstable behaviour); **b** resulting pressure distribution produces an effect that opposes the disturbance (stable behaviour)

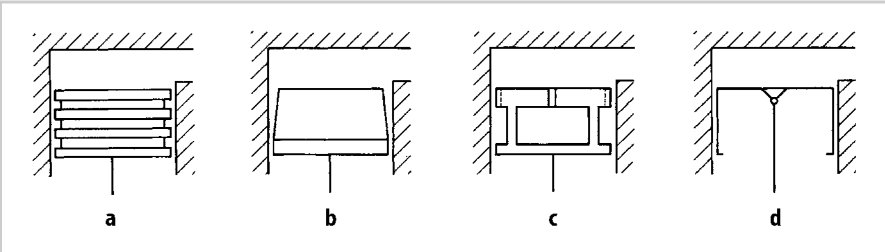


Figure 7.62. Measures for improving the resulting pressure distribution, after [7.225]: **a** unstable behaviour mitigated by pressure-equalising grooves; **b** stable behaviour through conical piston; **c** through pressure pockets; **d** through joint fitted above centre of gravity of the piston

ensured by the layout shown in Figure 7.61b, which, however, has a disadvantage: the piston rod inlet has to be sealed off on the pressure side.

According to [7.225], the layout shown in Figure 7.61a can be stabilised by the measures shown in Figure 7.62a-d. They ensure that a disturbance will itself initiate pressure distributions that tend to correct the misalignment.

Another example is the well-known case of hydrostatic bearings with oil pockets distributed around the periphery. When the bearing is loaded, the leakage path below the load is reduced, with the result that pressure builds up in the affected oil pocket and decreases in the opposite one. Thanks to the combined effect, the bearing can take up the load with very small shaft displacement.

The stuffing boxes and seals of turbomachinery must always be designed for thermostable behaviour [7.225]. The seal of a turbocharger shown in Figure 7.63 is a case in point. In the thermo-unstable layout (see Figure 7.63a), most of the frictional heat generated by contact forces will flow into the rotor, which will heat up further, expand, and hence increase the contact forces. In the stable arrangement (see Figure 7.63b), in contrast, the frictional heat will cause the contact forces to be reduced. A disturbance thus produces a self-limiting effect.

A similar approach is used in the design of taper roller bearings. Thus, in the layout shown in Figure 7.64a, heating of the shaft, by excessive loading for instance, will tend to increase the load even further because of the expansion of the shaft due to the increased frictional heat. The arrangement shown in Figure 7.64b, in contrast,

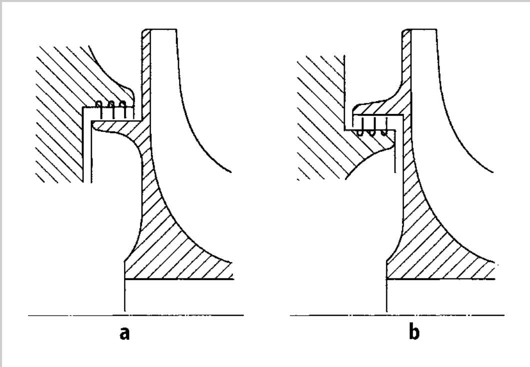


Figure 7.63. Seal in turbocharger, after [7.225]

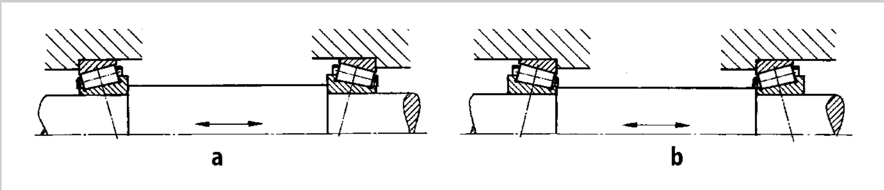


Figure 7.64. Taper roller bearings in which the shaft heats up more than the housing: **a** thermal expansion leads to increased loading and hence to unstable behaviour; **b** thermal expansion leads to reduced loading and hence to stable behaviour

will lead to a load reduction. In the case under consideration, this reduction must not, however, be allowed to reach the point where one of the bearings becomes unloaded, because the shaft at that point would then not be located radially and the bearings would be easily damaged.

Another interesting example of thermostable behaviour is provided by the double-helical gears used in marine gearboxes [7.322].

2. Principle of Bi-Stability

In some cases, unstable or bi-stable behaviour is positively welcome. This happens when, upon reaching a limit, a clearly distinct state or position is required and no intermediate state is acceptable. The requisite instability is initiated when a selected physical quantity reaches a limiting value and then introduces self-reinforcing effects which cause the system to jump into a second stable state. This bi-stable behaviour is required for switches and protective systems (see Section 7.3.3).

A well-known application of this is in the design of safety and alarm valves [7.225], which, upon reaching a limiting pressure, will spring from a completely closed to a completely open position. This avoids undesirable settings with a low flow rate or flutter and wear of the valve seat. Figure 7.65 illustrates the solution principle.

Up to the limiting pressure $p = p_1$, the valve remains closed under the preload of the spring. If this pressure is exceeded, then the valve head will lift off very slightly.

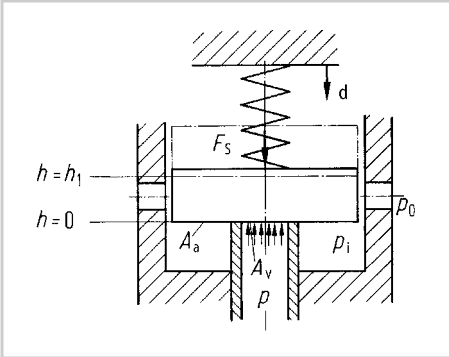


Figure 7.65. Solution principle for a valve with an unstable opening mechanism: d = precompression of spring; s = stiffness of spring; F_S = spring force; h = lift of valve head; p = pressure on valve; p_l = limiting pressure just sufficient to open the valve; p_i = intermediate pressure upon opening of valve; p' = pressure after opening of valve; p_0 = atmospheric pressure; A_v = surface area of valve opening; A_a = additional surface area.
Valve closed: $F_S = s \cdot d > p \cdot A_v$, $h = 0$
Valve just open: $F_S = s \cdot d \leq p_l \cdot A_v$, $h \approx 0$
Valve opening fully: $F_S = s(d + h) < p \cdot A_v + p_i \cdot A_a$, $h \rightarrow h_l$
Valve fully open: $F_S = s(d + h_l) = p'(A_v + A_a)$, $h = h_l$ (new equilibrium position)

The result is an intermediate pressure p_i, because the valve head throttles the outlet. This intermediate pressure acts on the additional surface A_a of the valve head and produces a supplementary opening force that offsets the elastic force of the spring F_S to such an extent that the valve head lifts rapidly. In the open state, a different intermediate pressure p' is set up and keeps the valve open. To close the valve, the pressure must be reduced considerably below the limiting opening pressure, because, the pressure is applied to a greater working surface area in the open state.

One application of this is the pressure switch for monitoring bearing oil pressure shown in Figure 7.66. If the bearing oil pressure drops below a certain value, the piston jumps open and the pressure inside the safety system is reduced with consequent shut-off of the endangered machinery.

The principle of bi-stability is also applied to the design of quick shut-off devices in which a striker pin under a spring preload has its centre of gravity slightly offset from the centre of rotation (see Figure 7.67). Once a limiting angular speed is reached, the striker pin begins to move against the spring preload. The resulting increase in the eccentricity of the centre of gravity leads to an increase in the centrifugal inertia force acting on the pin, which is flung out even without any further increase in the angular speed. For this to happen, however, the rate of increase in the centrifugal inertia force with x must be greater than that of the opposing spring force when the centre of gravity of the pin begins to move, see Figure 7.68. The forces must be equal in the limiting state ($\omega = \omega_l$). This can be achieved provided that:

$$dF_c/dx > dF_s/dx \quad \text{or} \quad m \cdot \omega_l^2 > s$$

Once it has been displaced to the outside, the pin strikes a catch which, in turn, activates the quick shut-off mechanism.

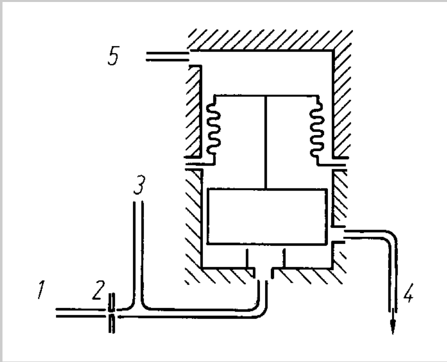


Figure 7.66. Diagrammatic sketch of a pressure switch used to monitor bearing oil pressure, after [7.225]. *1* Main oil system pressure: *2* orifice; *3* safety system activating quick shut-off valves; *4* drainage (no pressure); *5* bearing oil pressure

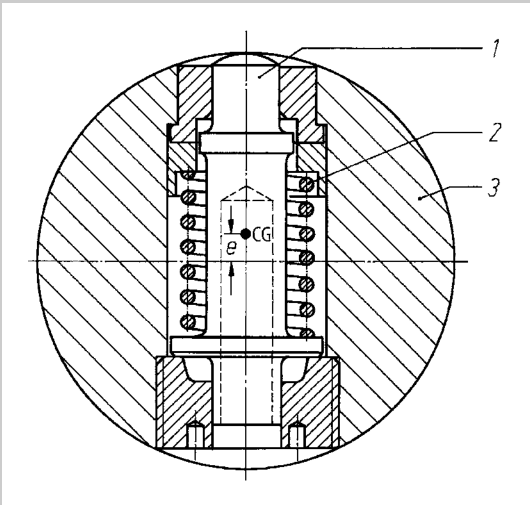


Figure 7.67. Quick shut-off pin *1* in shaft *3* with centre of gravity CG offset by *e*, and spring *2* holding the pin in the normal position, after [7.225]

7.4.5 Principles for Fault-Free Design

In high-precision products, in particular, but also for other technical systems, an embodiment should be sought in which the number of potential faults is minimised. This can be achieved by:

- designing a simple structure with simple components that have few close tolerances
- adopting specific design measures to minimise the causes of faults
- selecting working principles and working structures whose functions are largely independent of any disturbing effects, or which only have a low interdependency (see Section 7.3.1: basic rule of clarity)

- ensuring that any potential disturbing factors influence two parameters that compensate each other at the same time (see Section 7.4.1: principle of balanced forces).

Examples of this important principle [7.159,7.241,7.315] that result in simpler production and assembly and maintain product quality are: the elastic and adjustable

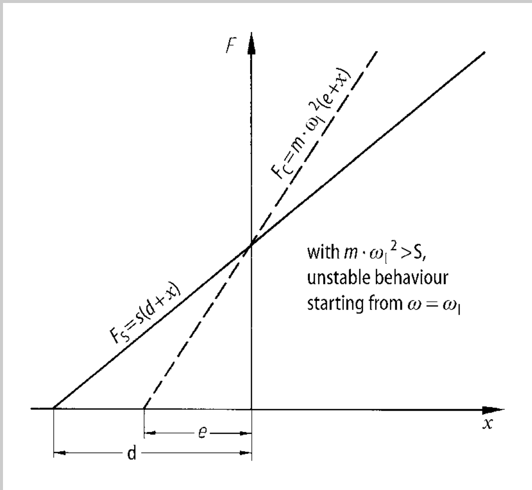


Figure 7.68. Graph of spring force and centrifugal inertia force against the displacement x of the centre of gravity of the quick shut-off pin (see Figure 7.67) . e = eccentricity of centre of gravity; d = spring precompression; ω_l = limiting angular speed beyond which the pin lifts off

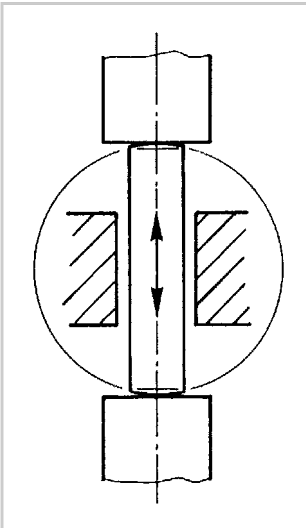


Figure 7.69. Link that is independent of play for the precise transfer of position [7.159]

configurations used in multigear gearboxes to balance out tooth tolerances (see Figures 7.45 and 7.47); the low stiffness of bolts and springs used to reduce the production tolerances in prestressed bolted connections and suspension systems; simple structures with few parts, low tolerances, and few toleranced joints; the possibility of adjusting and resetting to allow lower tolerances on individual components; the principle of stability (see Section 7.4.4).

Figure 7.69 shows a simple example: a compression link for the precise transfer of position. By making the ends of the link dome-shaped based on a shared spherical surface, the distance between the driving component and the receiving component remains the same despite any tilting of the plunger caused by any play in the guides [7.159].

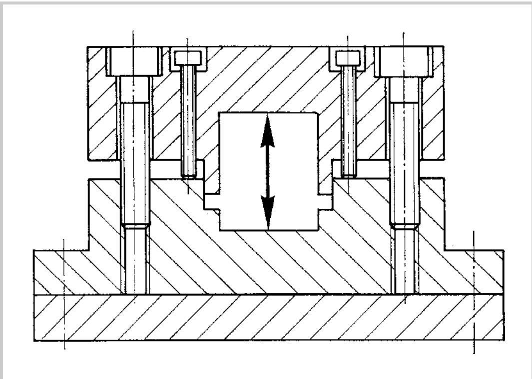


Figure 7.70. Continuous adjustment provided to maintain tight tolerances

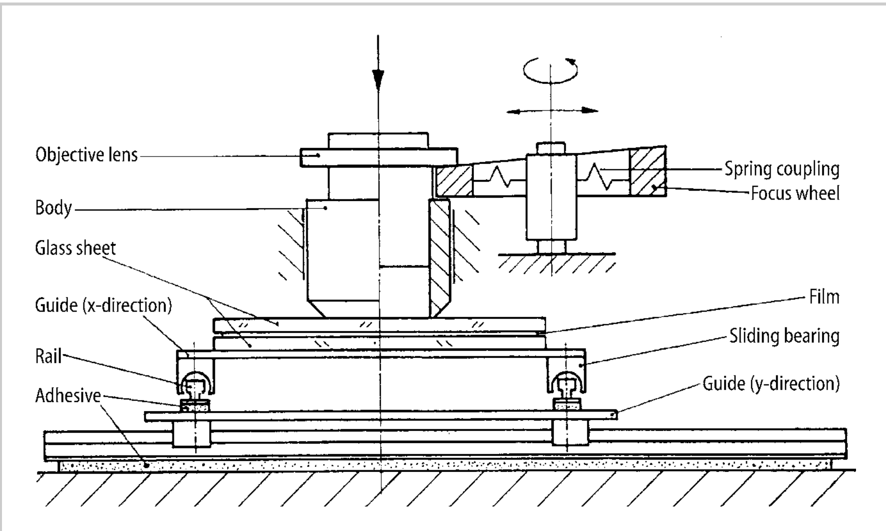


Figure 7.71. Automatically adjusting function chain in a microfiche reader

The example in Figure 7.70 illustrates how continuous adjustments can be incorporated to make it easier to maintain a volume with very tight tolerances in, for example, a split mould.

Figure 7.71 shows a further example. In a microfiche reader it is important to keep the objective lens perpendicular to the microfiche, which is held between glass plates. The usual solution is to mount the lens in a cylindrical body with tight tolerances, with its axis perpendicular to the glass surface. The solution in Figure 7.71, however, locates the cylindrical body directly on the glass plate and therefore automatically maintains it perpendicular to the surface of the glass.

7.5 Guidelines for Embodiment Design

7.5.1 General Considerations

In addition to the three basic rules of clarity, simplicity and safety derived from the general objectives (see Section 7.3), designers should also follow a number of embodiment design guidelines based on the general constraints set out in Section 2.1.7 and the checklist in Figure 7.3. These guidelines are internationally known as *Design for X*. They support the basic rules and help designers meet the specific requirements and constraints.

In what follows we cover the most important guidelines, without making any claims to completeness. Detailed discussions are dispensed with whenever summaries or special accounts have been published, to which the reader is referred.

This is the case for *design for durability* (stress requirements), and designers should refer to the literature covering the selection and design of machine elements [7.157, 7.165, 7.198, 7.275]. Special attention should be paid to changes in loading conditions with time and to the correct estimates of the level and type of the resulting stresses, as well as to the selection of the most suitable failure criteria. Damage-accumulation criteria help to improve service life predictions [7.16, 7.113, 7.116, 7.126, 7.247]. When determining stresses, stress concentrations and/or multiaxial stress conditions should be taken into account [7.193, 7.276, 7.284]. Assessments of durability should be based on the material properties and the appropriate failure criteria [7.192, 7.274, 7.276, 7.298, 7.299].

When *designing to allow for deformation, stability and resonance*, designers should refer to the appropriate calculations in mechanics and machine dynamics: mechanics and strength problems [7.17, 7.165]; vibration problems [7.155, 7.176]; stability problems [7.217]; and Finite Element methods [7.335]. In Section 7.4.1 we dealt briefly with the problems of designing with due allowance for the deformation caused by the transmission of forces.

This book discusses in some detail the following embodiment design guidelines. Design to allow for expansion and creep—that is, temperature phenomena—are discussed in Sections 7.5.2 and 7.5.3; design against corrosion in Section 7.5.4; and design to minimise wear in Section 7.5.5. Design for ergonomics is discussed in Section 7.5.6 and for aesthetics in Section 7.5.7; design for production and assembly, including quality control and transport, is dealt with at some length in Sections 7.5.8 and 7.5.9; and design for maintenance in Section 7.5.10. Design for recycling is discussed in Section 7.5.11; design for minimum risk in Section 7.5.12; and design to standards in Section 7.5.13.

7.5.2 Design to Allow for Expansion

Materials used in technical systems tend to expand when they are heated. The resulting problems must be taken into consideration, not only in the design of thermal devices in which higher temperatures must be expected as a matter of course, but also in high-performance engines and devices in which frictional heating can occur and special cooling is employed. As a result, several areas will be affected by local heating. Moreover, devices whose environmental temperature fluctuates significantly will only work properly if the physical effects of thermal expansion have been allowed for in the design [7.202, 7.206].

Apart from the thermal effects of linear expansion, designers must also consider the purely mechanical extension of parts subjected to heavy loading. In principle, the guidelines also apply to this type of change of length.

1. Expansion

Expansion has been the subject of a host of special studies. For solid bodies, the coefficient of linear expansion is defined as:

$$\alpha = \Delta l/(l \cdot \Delta\vartheta_m)$$

where Δl = change in length (expansion) due to a temperature rise of $\Delta\vartheta_m$, l = the length of the component under consideration, and $\Delta\vartheta_m$ = mean temperature difference to which the body is subjected.

The coefficient of linear expansion defines the expansion of a solid along one coordinate axis only, while the coefficient of cubical expansion defines the relative change in volume per degree of temperature rise. For homogeneous solids, its value is three times that of the coefficient of linear expansion. Coefficients of expansion should be understood as mean values over a particular temperature range; they depend not only on the material but also on the temperature. At higher temperatures, the coefficient usually increases.

Figure 7.72 gives the coefficients of linear expansion for distinct groups of engineering materials. It shows that with commonly used combinations of metals, for example 35C carbon steel with austenitic (10C/18% Cr-Ni-Nb) steel, or grey cast iron with bronze or aluminium, great care must be taken to allow for relative expansions because of the significant differences in the coefficients of thermal expansion between the materials. With large dimensions, even relatively small differences between, say, 35C carbon steel and 13% chromium steel (10C/13% Cr) can cause serious problems.

Metals with a low melting point, such as aluminium and magnesium, have greater coefficients of thermal expansion than metals with a high melting point, such as

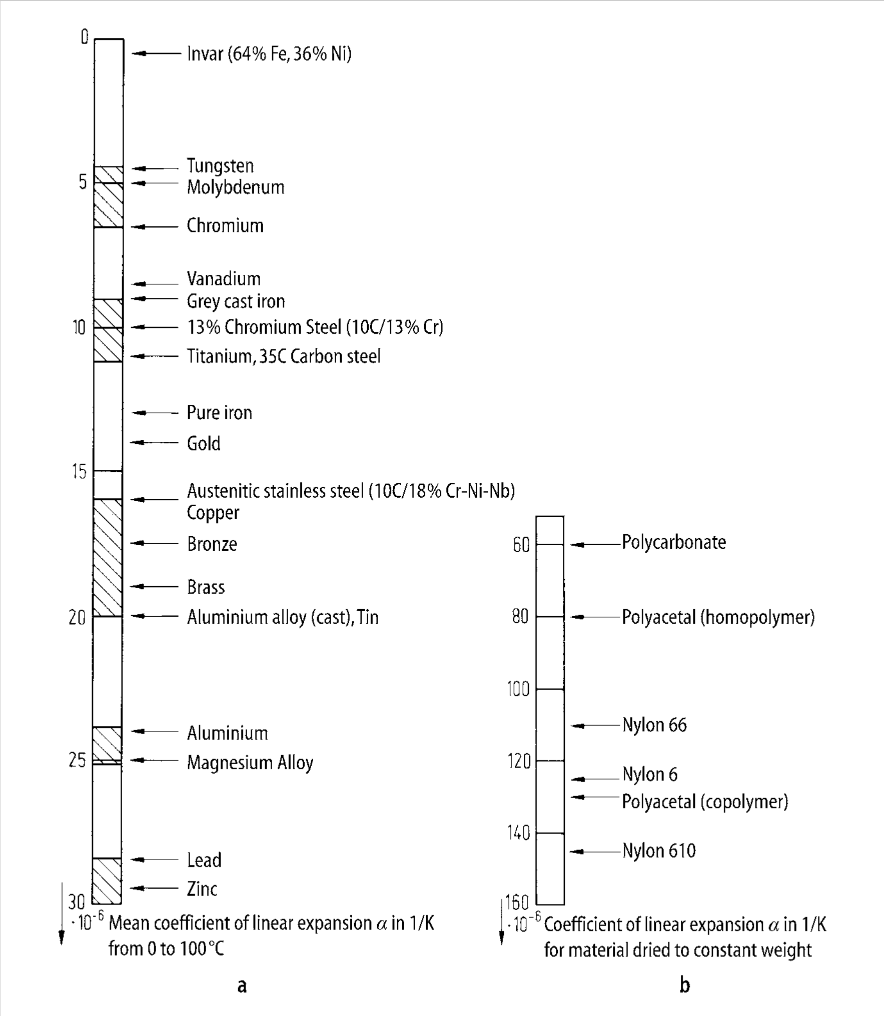


Figure 7.72. Mean coefficients of linear expansion for various materials: **a** metallic; **b** synthetic

tungsten, molybdenum and chromium. Nickel alloys have different coefficients depending on their nickel content. Very low values occur in the range of 32–40% by weight, with 36% Ni-64% Fe, known as "Invar", having the lowest coefficient. Synthetic materials have significantly higher coefficients of expansion than metals.

2. Expansion of Components

To calculate changes in length Δl, designers must know the temperature distribution (position and time) in the component and hence the mean temperature change with respect to the initial value. If the temperature distribution does not change with time, we speak of a *steady* or fixed expansion. If the temperature distribution changes with time, we speak of an *unsteady* or fluctuating expansion.

In the case of steady expansion, the physical quantities upon which the expansion of a component depends is obtained from the basic equations:

$$\Delta l = \alpha \cdot l \cdot \Delta\vartheta_{\mathrm{m}} \qquad \Delta\vartheta_{\mathrm{m}} = \frac{1}{l}\int_0^l \Delta\vartheta(x)\,\mathrm{d}x$$

The change in length Δl is therefore dependent on:

- the coefficient of linear expansion α
- the length l of the component
- the mean temperature change $\Delta\vartheta_{\mathrm{m}}$ over this length.

The value thus determined has a direct bearing on the design: every component must be clearly located and must only have as many degrees of freedom as are necessary for its proper functioning. In general, a point is fixed and the requisite translational and rotational movements are set by appropriate guides, for example slides, bearings, etc. A body in space (a satellite or helicopter) has three translational degrees of freedom in the x, y and z directions and three rotational degrees of freedom about the x, y and z axes. A sliding pivot (for example the nonlocating bearing of a shaft) provides two degrees of freedom: one translational and one rotational. A body clamped at one point (for example a built-in beam), on the other hand, has no degrees of freedom. Layouts based on these considerations alone do not, however, allow for expansion automatically, as we shall now demonstrate.

Figure 7.73a shows a body clamped at one point with no degrees of freedom. Upon thermal expansion it can expand freely from this point along the various axes. Figure 7.73b shows a plate that can be rotated about the z axis and thus has one degree of freedom. As shown in Figure 7.73c, this single degree of freedom can be simply removed by means of a slide. Were this plate to expand under uniform temperature increases, it would have to rotate about the z axis, for the slide does not lie in the direction of the expansion that results from the change of length in the x and y directions. If the slide allowed only translational movement and did not also act as a pivot, then jamming would occur. By fitting the slide in the direction of one of the coordinates (see Figure 7.73d), it is possible to avoid the rotation of the component.

After deformation due to thermal expansion, geometric similarity will only be maintained if the following conditions are met:

- The coefficient of expansion α must be constant throughout the component (isotropy), which can be taken for granted in practice provided that only one kind of material is used and that the temperature differences are not too great.
- The thermal strains ε along the x, y, z axes must be such that:

$$\varepsilon_{\mathrm{x}} = \varepsilon_{\mathrm{y}} = \varepsilon_{\mathrm{z}} = \alpha \cdot \Delta\vartheta_{\mathrm{m}} \text{ [7.196]} .$$

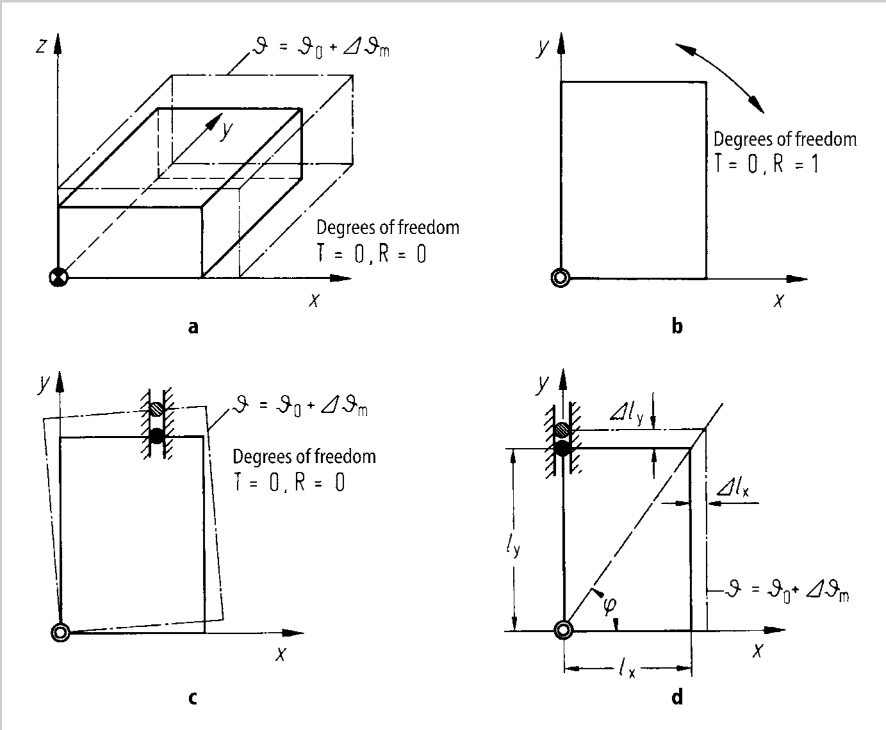


Figure 7.73. Expansion due to steady uniform temperature distribution. *Continuous line*: initial state; *broken line*: higher temperature state. **a** Body attached to a fixed point; **b** plate can rotate about the *z* axis, that is, it has one degree of freedom; **c** plate as in **b** but with single degree of freedom removed by an additional sliding pivot; **d** plate as in **b** but allowing for expansion without rotation. It would also be possible to use simple slides which might equally well be arranged along the *x* axis as along a line through the *z* axis inclined at $\tan\varphi = l_y/l_x$

If α is constant throughout a component, then the mean temperature increase must be the same for all three axes, so that we have:

$$\Delta l_x = l_x \cdot \alpha \cdot \Delta\vartheta_m$$
$$\Delta l_y = l_y \cdot \alpha \cdot \Delta\vartheta_m$$
$$\Delta l_z = l_z \cdot \alpha \cdot \Delta\vartheta_m$$

and for the *x* and *y* axes:

$$\tan\varphi = \frac{\Delta l_y}{\Delta l_x} = \frac{l_y}{l_x} \qquad \text{(See Figure 7.73d)}$$

- The component must not be subjected to additional thermal loads, which will not happen if, for instance, it is completely surrounded by a source of heat [7.183].

As a rule, however, different temperatures are measured in a single component. Even in the simplest case, with the temperature distribution changing linearly along the x axis (see Figure 7.74a), a change in angle is produced which, again, can only be taken up by a guide with a sliding as well as a pivoting movement. A simple slide, which allows translational movement with one degree of freedom, can only be used if the guide lies along the line of symmetry of the deformation (see Figure 7.74b). If this condition is not fulfilled, a further degree of freedom must be allowed.

Hence we obtain the rule that guides that take up thermal expansion and have one degree of freedom must only lie on a line through the fixed point, and this line must be the line of symmetry of the deformed state. The deformed state can be caused by load-dependent and temperature-dependent stresses, in addition to the expansion itself.

Since the stress and temperature distribution also depend on the shape of the component, the required symmetry line of the deformed state should, in the first instance, be sought both along the symmetry line of the component and also along that of the superimposed temperature field. However, as Figure 7.74b shows, this line of symmetry may not be easily identifiable from the component shape and the temperature distribution, so that the ultimate state of deformation must also be taken into account. That state, as we said earlier, may also be caused by external loads. To that extent, our remarks also apply to guides for components subject to large mechanical deformations. An example can be found in [7.8].

The following examples serve as further illustrations. Figure 7.75 is the plan view of a device whose temperature decreases from the centre to the periphery. It is supported on four feet. In Figure 7.75a, one of the feet is chosen as the fixed point. If the device is not to rotate or jam, the guide may only be placed along the line of symmetry of the temperature field, that is, on the opposite foot. Figure 7.75b shows a method of providing guides along lines of symmetry without a designated fixed point. The intersection of the lines through the guides constitutes an imaginary fixed point from which the device can expand evenly in all directions. In that case, two guides, for example *1* and *2*, could be omitted.

Figure 7.76 shows the location of inner casings in outer casings when a common centre must be maintained, as occurs, for instance, in turbines. If the deformed

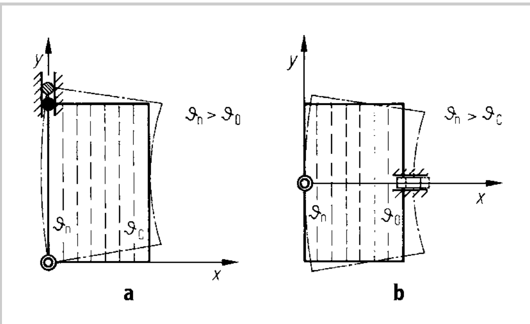


Figure 7.74. Expansion under nonuniform temperature distribution, here decreasing linearly along the x axis: **a** plate corresponding to Figure 7.73d, nonuniform temperature distribution produces deformation shown by *broken line*, sliding pivot required; **b** guide placed on symmetry line of deformed state so that a simple slide can be used

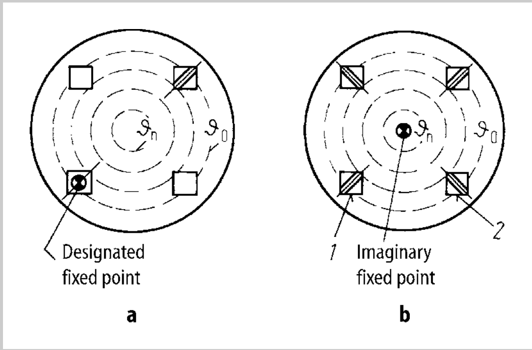


Figure 7.75. Plan view of a device mounted on four feet, whose temperature decreases from the centre to the periphery: **a** designated fixed point on one foot; simple slide along a line that is also the symmetry line of the temperature field; **b** imaginary fixed point in the centre of the device formed by the intersection of the lines of expansion

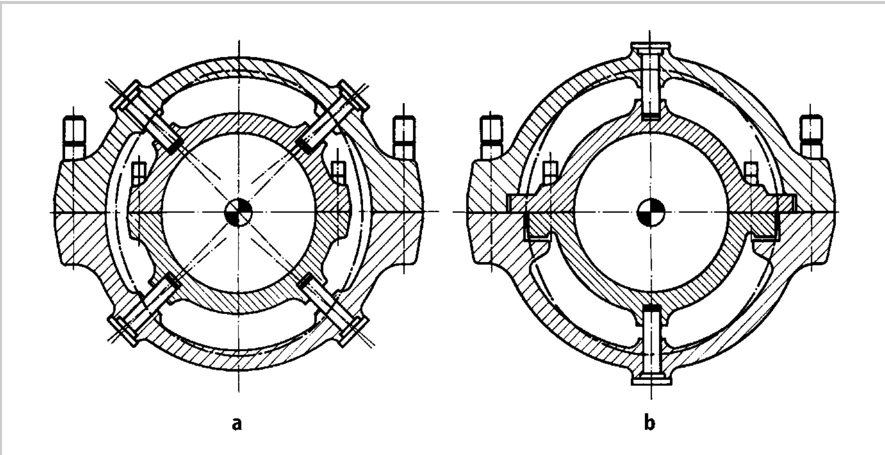


Figure 7.76. Location of inner casings in outer casings: **a** arrangement of guides does not allow for expansion: oval deformation of the housings can cause guides to jam; **b** arrangement allowing for expansion: guides lie along symmetry lines, no jamming with oval deformation

shape of these components is not completely rotationally symmetrical, then the guides must be placed on the lines of symmetry to prevent jamming of the guides due to, say, oval deformation of the casings (see Figure 7.76b). Such oval deformation is caused by temperature differences in the housing wall and flange, especially during the warm-up phase. The imaginary fixed point lies on the longitudinal axis of the casing or shaft.

Figure 7.77 shows an austenitic steel high-temperature steam inlet pipe *a* which must be fitted into a ferritic steel outer casing *b* while protruding into a ferritic steel inner casing *c*. Because of marked differences in the two coefficients of expansion and also because of the considerable temperature differences between the components, particular attention must be paid to relative expansion. An imaginary fixed

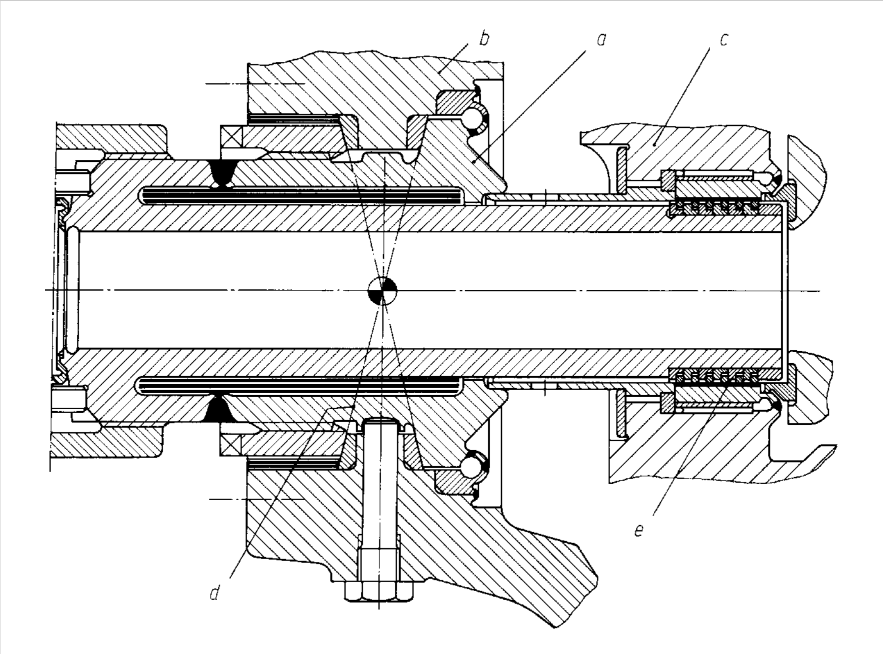


Figure 7.77. Inlet pipe *a* of a steam turbine made of austenitic steel that takes the steam through the ferritic steel outer casing *b* to the inner casing *c*. Expansion planes through guideways *d* determine an imaginary fixed point. Piston ring seals at *e* permit the axial and radial expansion of the end of the inlet pipe (BBC)

point is provided by the rotationally symmetrical guides *d*, an arrangement ensuring the unimpeded expansion of the austenitic component along any line through the imaginary fixed point. Because the temperature distribution at that point is fairly uniform, the respective radial and axial expansions produce an expansion along the indicated lines.

By contrast, the insertion of the inlet pipe into the inner casing must allow independent expansion along two axes, because the fixed point of the inlet pipe and the fixed point of the inner casing are not identical and no definite temperature distributions can be assigned to the components. The double degree of freedom is obtained with the help of the piston-ring seal *e*, which permits independent axial and radial movements of the inlet pipe.

3. Relative Expansion of Components

So far we have considered expansion in a relatively stable environment. Very often, however, the relative expansion of two (or more) components has to be taken into account, especially in the case of mutual loadings or when certain clearances must be maintained. If in addition the temperature varies with time, then designers are faced with a very difficult problem. The relative expansion of two components is:

$$\delta_{\mathrm{rel}} = \alpha_1 \cdot l_1 \cdot \Delta\vartheta_{\mathrm{m}_1(\mathrm{t})} - \alpha_2 \cdot l_2 \cdot \Delta\vartheta_{\mathrm{m}_2(\mathrm{t})}$$

Steady-State Relative Expansion

If the relative mean temperature difference does not vary with time, and if the coefficients of linear expansion are identical, then all that has to be done to minimise the relative expansion is to even out the temperature or else to select materials with different coefficients of expansion. Often both are necessary.

This can be seen in the case of a flanged connection consisting of a steel stud and an aluminium flange [7.200]. Because the aluminium has a higher coefficient of expansion, a temperature rise will increase the load on the stud, which may lead to failure (see Figure 7.78a). This can be prevented, on the one hand, by increasing the length of the stud and using a sleeve and, on the other hand, by using components with appropriate coefficients of expansion (see Figure 7.78b). If relative expansion is to be avoided altogether, then we must have:

$$\delta_{\text{rel}} = 0 = \alpha_1 \cdot l_1 \cdot \Delta\vartheta_{\text{m}_1} - \alpha_2 \cdot l_2 \cdot \Delta\vartheta_{\text{m}_2} - \alpha_3 \cdot l_3 \cdot \Delta\vartheta_{\text{m}_3}$$

With $l_1 = l_2 + l_3$ and $\lambda = l_2/l_3$, the relative length of sleeve to flange becomes:

$$\lambda = \frac{\alpha_3 \cdot \Delta\vartheta_{\text{m}_3} - \alpha_1 \cdot \Delta\vartheta_{\text{m}_1}}{\alpha_1 \cdot \Delta\vartheta_{\text{m}_1} - \alpha_2 \cdot \Delta\vartheta_{\text{m}_2}}$$

With steady-state expansion, $\Delta\vartheta_{\text{m}_1} = \Delta\vartheta_{\text{m}_2} = \Delta\vartheta_{\text{m}_3}$, and with steel ($\alpha_1 = 11 \times 10^{-6}$), Invar ($\alpha_2 = 1 \times 10^{-6}$) and aluminium alloy ($\alpha_3 = 20 \times 10^{-6}$) as the chosen materials (see Figure 7.78b), we have $\lambda = l_2/l_3 = 0.9$.

Designers will be familiar with the complicated expansion problems associated with the pistons of internal combustion engines. Here, the temperature distribution over and along the piston differs even in the near-steady

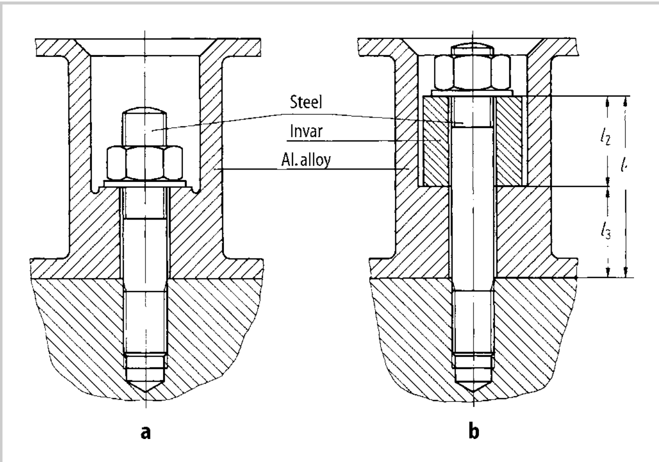


Figure 7.78. Connection by means of a steel stud and aluminium flange [7.200]: **a** stud endangered because aluminium flange has greater expansion; **b** incorporation of Invar expansion sleeve with a coefficient of expansion close to zero helps to balance the relative expansion of flange and stud

state and, what is more, differences in the coefficients of expansion of piston and cylinder must also be taken into account. One solution is the use of an aluminium–silicon alloy with a relatively small coefficient of expansion (smaller than 20×10^{-6}); of expansion-inhibiting inserts that are also good heat conductors; and of a flexible piston skirt. The bimetallic effect provided by steel inserts also helps to match the shape of the piston skirt to that of the cylinder [7.178] (see Figure 7.79). A further possibility is to make the piston oval-shaped.

If, on the other hand, the choice of materials is restricted in practice, then designers must rely on temperature adjustments. In high-power generators, for instance, large lengths of insulated copper rod must be embedded in the steel rotors. For insulation purposes alone, the absolute and relative expansions must be kept as small as possible. Here the only solution is to keep the temperature to a minimum by cooling [7.163, 7.317]. However, if these fast-running rotors have large dimensions, thermal imbalances may occur even though the temperature distribution is relatively uniform. The rotor, because of its complicated structure and the various materials that have gone into it, may not always (and at every point) display the same temperature-dependent behaviour. This can only be remedied if the expansions are kept under control by the carefully planned introduction of appropriate cooling or heating.

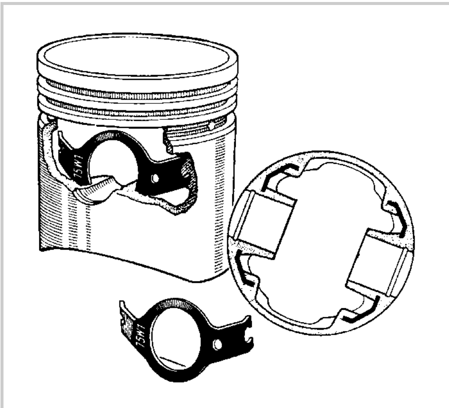


Figure 7.79. Piston of internal combustion engine made of aluminium–silicon alloy with steel inserts which inhibit circumferential expansion; moreover the bimetallic effect ensures optimum adaptation of the piston skirt to the cylinder (Mahle), after [7.178]

Unsteady Relative Expansion

If the temperature changes with time, for instance during heating or cooling processes, we often obtain a relative expansion that is much greater than that found in the final steady state. This is because the temperatures of the individual components can differ considerably. In the common case, where the components are of equal length and have equal coefficients of expansion, we have:

$$\alpha_1 = \alpha_2 = \alpha \text{ and } l_1 = l_2 = l$$
$$\delta_{\text{rel}} = \alpha \cdot l \left(\Delta\vartheta_{\text{m}_1(\text{t})} - \Delta\vartheta_{\text{m}_2(\text{t})}\right)$$

The heating of components has been examined by, among others, Endres and Salm [7.99, 7.236]. No matter whether we assume a step or linear change in temperature in the heating medium, the heating curve will be characterised by a time constant. If, for instance, we consider the temperature change $\Delta\vartheta_\text{m}$ of a component during a sudden temperature increase $\Delta\vartheta^*$ of the heating medium, then, under the admittedly approximate assumption that the surface and mean temperatures of the components are equal—which, in practice, is approximately true only for relatively thin walls and high thermal conductivity—we obtain the curve shown in Figure 7.80, with:

$$\Delta\vartheta_\text{m} = \Delta\vartheta^* \left(1 - \text{e}^{-\text{t/T}}\right)$$

Here t is the time and T is the time constant such that:

$$T = \frac{c \cdot m}{h \cdot A}$$

where:

c = specific heat of the component

m = mass of the component

h = heat transfer coefficient of the heated surface of the component

A = heated area of the component.

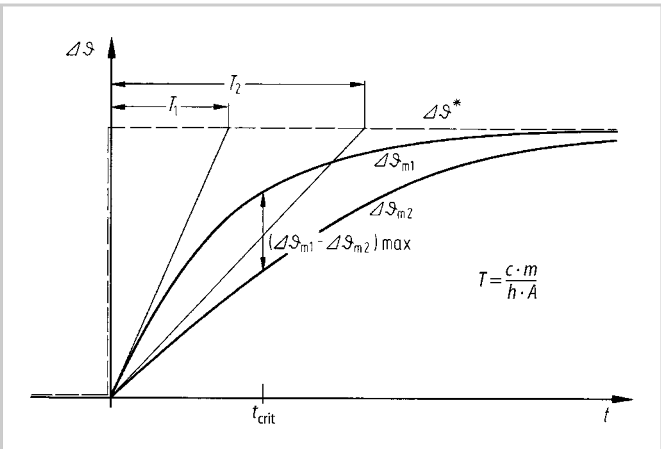


Figure 7.80. The effects on two components with different time constants of a step change in temperature $\Delta\vartheta^*$ in the heating medium

Despite the simplification involved, this approach may be considered to be fundamental. With two components that have different time constants, we obtain temperature curves that, at a given critical time, will have a maximum difference. At this point we have maximum relative expansion, and must provide clearances to accept the expansion or run the risk of excessive stresses beyond the yield point. Two identical temperature curves appear if the time constants of the two components can be equalised. In that case, there is no relative expansion. This objective cannot always be achieved, but in order to render the time constants approximately equal—that is, to reduce the relative expansion—the following relationship:

$$T = c \cdot \varrho \cdot \frac{V}{A} \cdot \frac{1}{h}$$

where V = volume of the component and ϱ = density of the component, can be used by designers to:

- adapt the ratio of the volume V to the heated surface area A, or
- adjust the heat transfer coefficient h by means of, say, lagging or selecting different cooling airflow rates.

Figure 7.81 gives the relationship V/A for a number of simple but representative bodies.

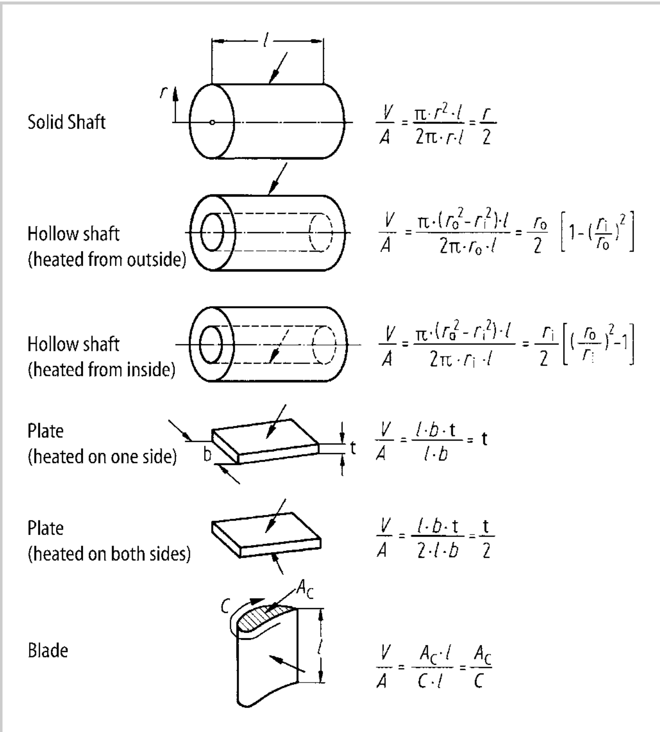


Figure 7.81. Volume–surface area relationships for various geometrical bodies; *arrows* point to heated surfaces

An example is shown in Figure 7.82. Here, the problem is to ensure adequate clearance for a valve spindle so that it can move safely and smoothly in its sleeve, even during temperature changes. In Figure 7.82a, the sleeve has been incorporated into the housing. When heated, the spindle will quickly expand radially, while the sleeve, which transfers its heat readily to the housing, remains cooler for a longer time. As a result, the clearance between the spindle and the sleeve will diminish dangerously.

In Figure 7.82b, the sleeves are sealed axially but can expand freely radially. Moreover, the volume-to-area ratio of the sleeves is such that the spindle and sleeves have the approximately equal time constants. As a result, the clearance remains more or less uniform at all temperatures and can therefore be kept small. The surface of the valve spindle and the inner surfaces of the sleeves are heated by steam leaks, so that we have:

$$(V/A)_{\text{spindle}} = r/2$$

$$(V/A)_{\text{sleeve}} = \left(r_o^2 - r_i^2\right)/2r_i$$

with $r_i = r$ and $(V/A)_{\text{spindle}} = (V/A)_{\text{sleeve}}$, we have

$$r/2 = (r_o^2 - r^2)/2r, \text{ and so}$$

$$r_0 = r \cdot \sqrt{2}$$

Figure 7.83 shows various steam turbine housings. With appropriate design, it is possible to adapt the volume-to-area ratios of the housings, the heat transfer coefficients and sizes of the heated surfaces to the time constants of the shafts and thus keep the blade clearances approximately constant when starting (heating) the

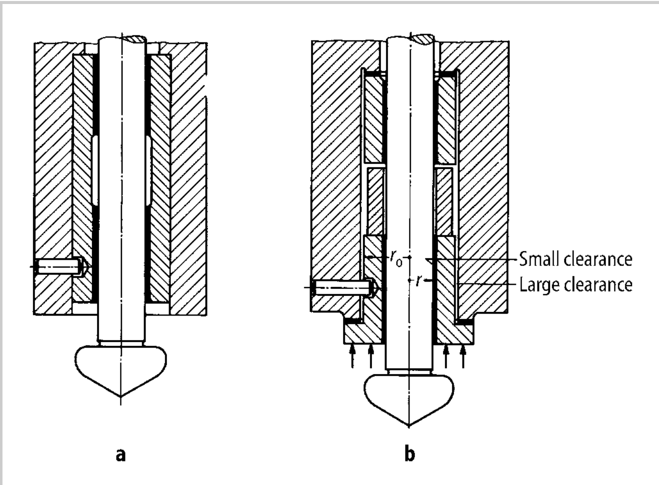


Figure 7.82. Spindle seals of steam valves: **a** fixed sleeve requires relatively large spindle clearance because it has not been designed to allow for expansion; **b** radially free and axially sealed sleeves permit small spindle clearance because spindle and sleeves have been designed to have the same time constant

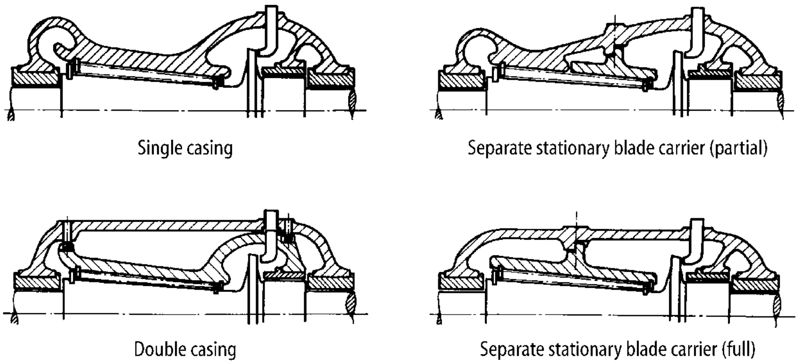


Figure 7.83. Steam turbine housings with different time constants

turbines. Another approach is to ensure that the relative expansion is such that the clearance increases rather than decreases during start-up.

There are several well-known methods for reducing the heat transfer coefficient of a component (for example by insulation), and thus for slowing down the heating and reducing the relative expansion.

The ideas we have just put forward are applicable wherever temperatures change with time, and particularly wherever relative expansion goes hand-in-hand with clearance reductions that are likely to endanger the functioning of turbines, piston engines and machines operating in hot environments.

7.5.3 Design to Allow for Creep and Relaxation

1. Behaviour of Materials Subject to Temperature Changes

When designing components subject to temperature changes, we must take into account not only the expansion effect but also the creep properties of the materials. The temperatures involved need not necessarily be very high, although they usually are. However, there are some materials that will, even at temperatures well below 100°C, behave in much the same way as metals do at very high temperatures.

Beelich [7.4] has examined this subject at some length, and in what follows we shall base ourselves largely on his findings.

Materials in common use, pure metals as well as alloys, have a polycrystalline structure and a temperature-dependent behaviour. Below a *critical temperature*, the stability of the intercrystalline bonds is largely independent of time, and the yield point can be used to determine the strengths of components. Components at temperatures above the critical temperature are strongly influenced by the time-dependent behaviour of the material. In this temperature range, materials will, under the influence of load, temperature and time, experience a gradual plastic deformation that, after a given period, may lead to fracture. The ensuing time-dependent fracture stress is much lower than the 0.2% proof stress at the same temperature determined by short-term experiments (see Figure 7.84). Critical temperature and creep strength depend largely on the materials used and both must be taken into consideration. With steels, the critical temperature lies between 300°C and 400°C.

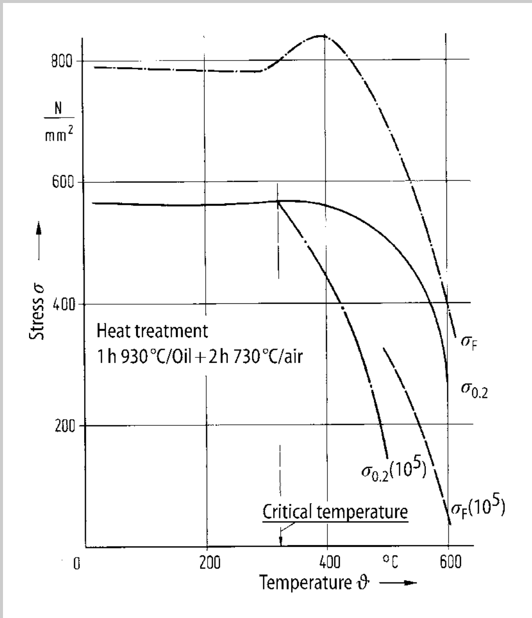


Figure 7.84. Characteristic values determined by high-temperature tensile strength and creep experiments with 21C/1.5% Cr-Mo-V steel at various temperatures; the critical temperature is the intersection of the curves of 0.2% proof stress and stress for 0.2% creep strain in 10^5 hours

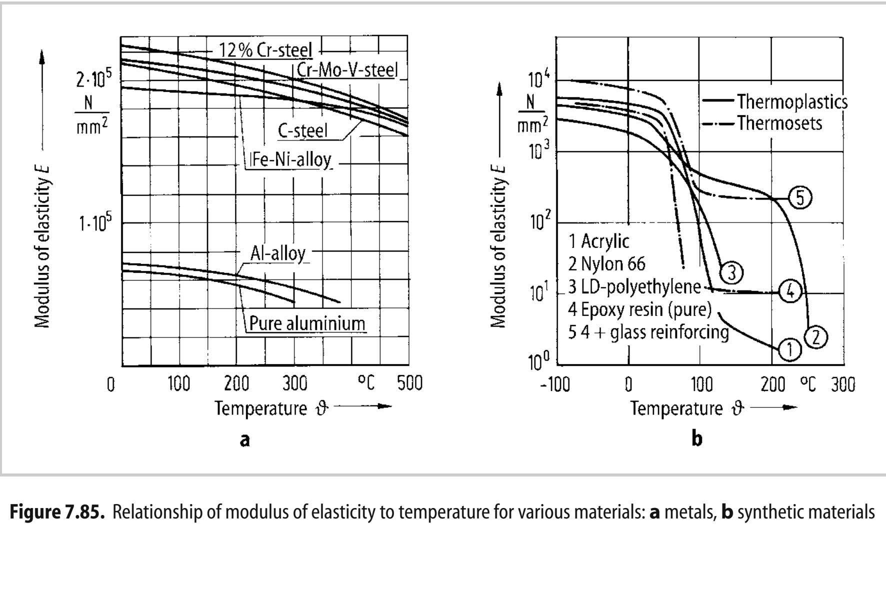


Figure 7.85. Relationship of modulus of elasticity to temperature for various materials: **a** metals, **b** synthetic materials

When working with synthetic materials, designers must allow for their viscoelastic behaviour, even at temperatures below 100°C.

In general, the modulus of elasticity changes inversely with the temperature (see Figure 7.85a). The smallest changes occur with nickel alloys. As the modulus of elasticity drops, so does the stiffness of the components—synthetic components in particular (see Figure 7.85b). In this case, designers must know the temperature at which the modulus of elasticity suddenly drops to relatively low values.

2. Creep

Components that are put under loads for long periods at high temperatures will, in addition to the strain given by Hooke's Law ($\varepsilon = \sigma/E$), also experience plastic deformation ($\varepsilon_{\text{plast}}$) with time. This property of materials, which is known as *creep*, depends on stress, the effective temperature ϑ and time.

We say a material creeps if the strain of a component increases under constant load or stress [7.4]. The creep curves of various materials are well known [7.110, 7.136].

Creep at Room Temperature

Before we can design components loaded to near the yield stress, we must know how they react in the transition region between the elastic and the plastic states [7.136]. With persistent static loads in this transition region, we can expect primary creep in metals even at room temperature (see Figure 7.86). The resulting plastic deformations are small and merely affect the dimensional stability of a particular component. In general, steels show little creep when subject to stress $\leq 0.75 \cdot \sigma_{0.2}$ or $\leq 0.55 \cdot \sigma_{\mathrm{F}}$, whereas, in the case of synthetic materials, a reliable assessment of the mechanical behaviour can only be made by considering the temperature and time-dependent characteristics.

Creep Below the Critical Temperature

Previous studies [7.136, 7.147] of metals have shown that the customary calculations, based on high-temperature yield strength as the maximum permissible stress for short-term loads, additional thermal loads and load variations, suffice up to the critical temperature.

With components that must have high dimensional stability, however, the characteristics of the material determined by creep experiments must also be taken into account, even at moderately high temperatures. Unalloyed and low-alloy boilermaking steels and even austenitic steels show varying degrees of creep depending on length of operation and working temperature.

Synthetic materials experience structural changes, even at slightly elevated temperatures. These transformations may lead to a marked temperature and time dependence of the properties of the materials, which is not the case with metals. In specific cases these changes are irreversible and referred to as thermal ageing [7.156, 7.185].

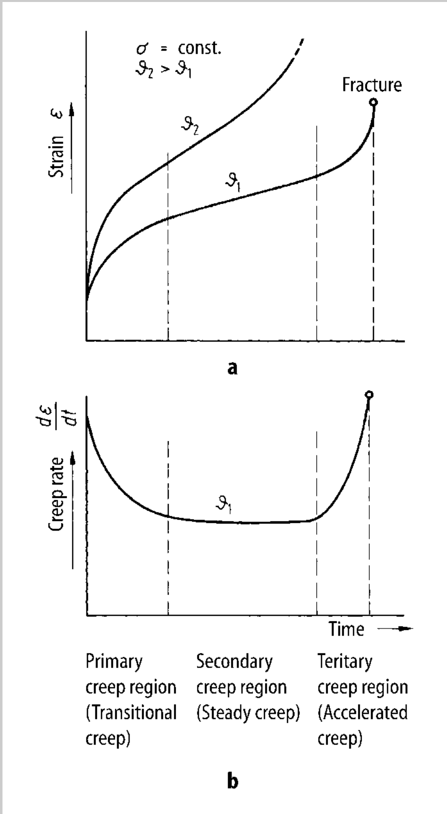


Figure 7.86. Strain **a** and creep rate **b** with duration of load (schematic representation); characteristics of the various creep phases

Creep Above the Critical Temperature

In this temperature region, mechanical loads will cause deformations in metals at far below the appropriate high-temperature yield strength, that is, the materials will creep. This creep leads to gradual deformation of components and can lead to loss of function and possible failure. In general, this process can be divided into three phases [7.136, 7.147] (see Figure 7.86). For components affected by temperature changes, the beginning of the tertiary creep phase must be considered dangerous. This region begins at approximately 1% permanent strain. Figure 7.87 shows the 10^5-hour creep strengths $\sigma_{1\%/10^5}$ at 500°C for various steels.

3. Relaxation

In loaded systems (springs, bolts, tension wires, shrink fits), the necessary preload produces an overall strain ε (elongation Δl). Because of creep and settling of the material due to plastic flow at the bearing surfaces and split lines, the ratio of

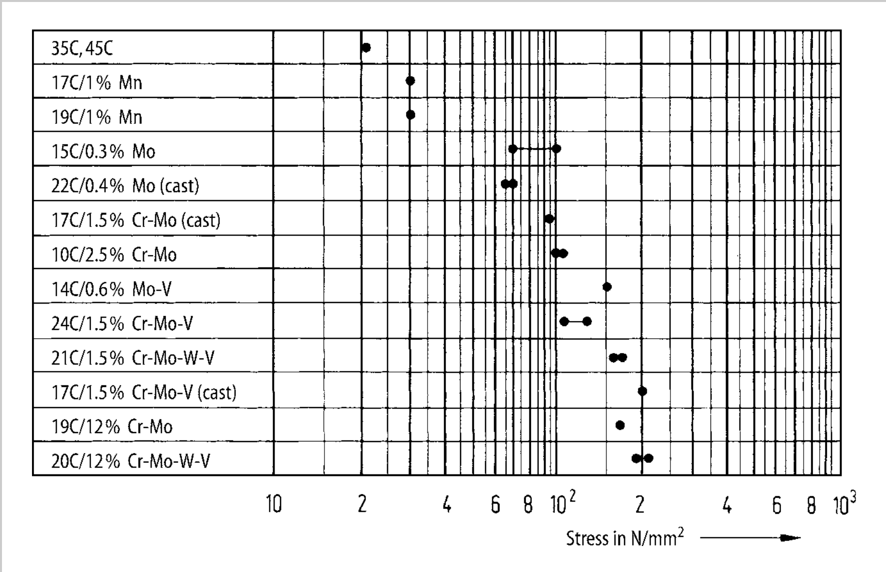


Figure 7.87. Stresses corresponding to a 1% permanent strain for various steels after 10^5 hours at 500°C [7.213]

plastic to elastic deformation gradually increases. The phenomenon of decreasing elastic strain at constant overall strain is called *relaxation* [7.100, 7.326, 7.327].

Loaded components are usually preloaded at room temperature. Because the modulus of elasticity varies with the temperature (see Figure 7.85), the preload decreases at higher temperatures, even without a change in length of the loaded system.

The preload remaining at operational conditions, though reduced, will lead to creep at high temperatures and hence to a further drop in the preload (relaxation). The residual clamping force is also affected by production and operation determined factors; for instance by the assembly preload, the design of the loaded system, the nature of the contact surfaces, and the influence of superimposed stresses (normal or tangential to the surface). Studies of the relaxation of bolted flanges [7.100, 7.326, 7.327] have shown that plastic deformation also occurs at the split lines and bearing surfaces (settlement) and in the threads (creep and settlement).

To sum up, we can say that, with metallic components:

- The drop in preload depends on the relative stiffness of the parts loaded against each other. The more rigid the connection, the greater the drop in the preload due to plastic deformation (creep and settlement).
- Although settlement can be appreciably offset during the tightening of bolted flanges or the assembling of shrink fits, designers should, where possible, provide for few but accurately machined surfaces (split lines, bearing surfaces).
- There is a temperature limit beyond which the material cannot be properly used. In addition, designers should always choose materials in which the ap-

propriate high-temperature yield point is not reached, even with superimposed operational stresses.

- In the short term, high initial preloads (initial clamping forces) give rise to higher residual clamping forces. In the longer term, the residual clamping forces become relatively independent of the initial preload.
- Joints that have already undergone relaxation can be tightened up if the toughness of the material permits. As a rule, creep of about 1%, which leads to the tertiary creep region, must not be exceeded.
- If joints are subjected to an alternating load in addition to the static preload, then, as experiments have shown, the amplitudes tolerated during relaxation-dependent decreases in the mean stress are considerably greater than those tolerated at constant mean stress. However, relaxation-dependent decreases in the mean stress will often lead to a loosening of the joint.

When using bolted joints made of synthetic materials, designers try to take advantage of their low electrical and thermal conductivity, their resistance to corrosion, their high mechanical damping, their small specific weight, etc. In addition, such joints must, of course, have the appropriate strength and toughness. Special attention must also be paid to preload decay, otherwise the functioning of the joints can be seriously impaired. Special studies [7.190, 7.191] have shown that in synthetic (unlike in metallic) materials:

- The preload remaining after a given time and at room temperature is determined by the material itself and its tendency to absorb moisture.
- Continual changes in the absorption and release of moisture have a particularly deleterious effect.

4. Design Features

In order to increase the potential life of components subject to long-term loads, designers must familiarise themselves with the behaviour of the materials involved over time. According to [7.136], it is dangerous to use short-term values to predict load responses for periods of 10^5 hours or longer.

It is impossible to avoid thermal stresses in all components by specifying the use of highly alloyed materials. Appropriate design features are often more useful than changing the materials.

The design must enable creep to be kept within permissible limits, which can be done by means of:

- a high elastic strain reserve, which helps to minimise additional loads due to temperature fluctuations (see Figure 7.88)
- insulation or cooling of components, as in double-casing steam turbines and gas turbines (see Figure 7.89)
- the avoidance of mass concentrations which, in unsteady processes, may lead to increased thermal loading

- the prevention of creep in unwanted directions, which can cause functional failure (for instance the jamming of valve spindles) or dismantling problems (see Figure 7.90).

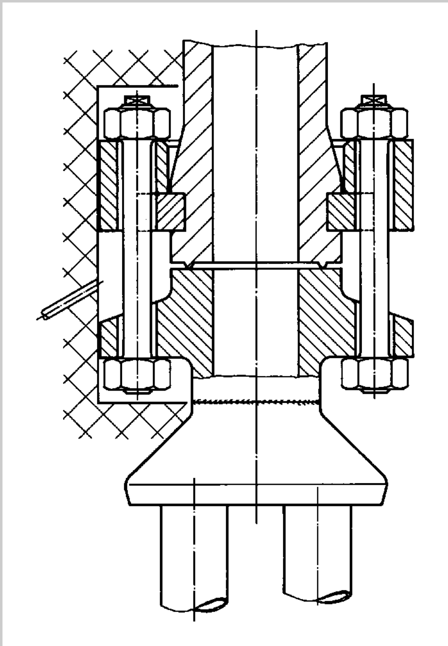


Figure 7.88. Austenitic–ferritic steel flanged joint intended for operating temperatures of 600°C [7.265]

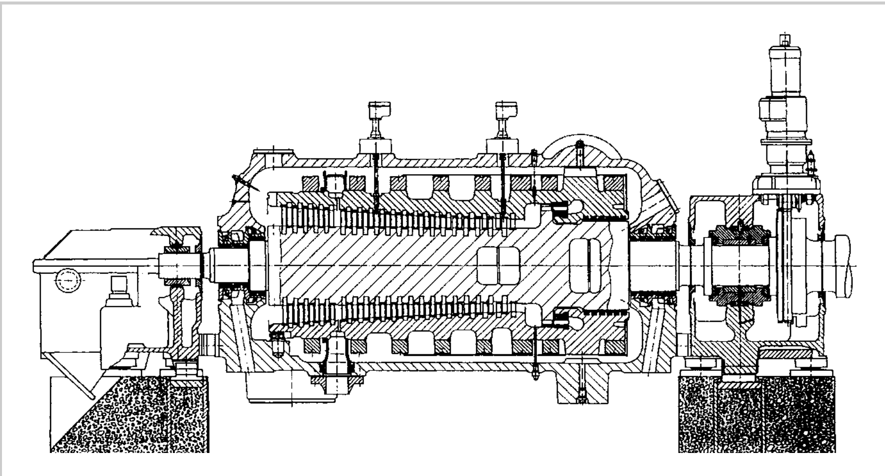


Figure 7.89. Double-casing steam turbine with shrink rings that hold the inner casing together. Relaxation of the shrink rings is reduced by cooling with exhaust steam. As the machine increases its output, the shrink rings exert an increasing pressure thanks to growing temperature difference between the steam inlet and outlet. The shrink rings are seated on heat-inhibiting segments which, with the help of shims, permit the original shrink fit to be restored after relaxation (ABB)

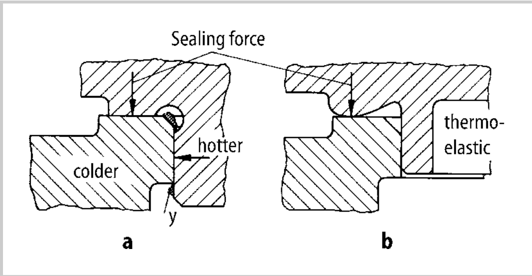


Figure 7.90. Centering and sealing of a cover [7.206]. **a** Dismantling is impeded because the material creeps into the relief groove and at y; **b** convex sealing edge provides a better seal with smaller sealing force. Creep does not impede dismantling thanks to improved design

In Figure 7.90a, the material of the cover creeps into the relief groove. The cover, which heats up more quickly, presses against the centering surface and also creeps at point y. The cover shown in Figure 7.90b is a better design since, despite the creep, it can be dismantled easily. In addition, the cover has been made hollow so that it cannot exert a significant radial force on the centering surface. In other words, the part which is moved during dismantling should not project axially beyond the fixed part [7.206].

7.5.4 Design Against Corrosion

It often happens that corrosion can only be reduced, not completely avoided. Rubo [7.235] emphasises the use of components with the same corrosion resistance in a machine. The use of corrosion-proof materials throughout may not be economic, in which case suitable embodiments can be used to retain functionality despite corrosion. This suggests a shift from focusing on corrosion protection to designing machines and their components to be corrosion tolerant. It follows that designers must tackle corrosion with appropriate concepts or special embodiment design features. The measures they take will depend on the type of corrosion anticipated. An extensive description of the types of corrosion and many useful design features are provided in the guidelines on design against corrosion in [7.158]. Spähn, Rubo and Pahl [7.212, 7.261] describe various types of corrosion and their remedies, and the following remarks are largely based on their findings. In order to provide a systematic categorisation from a design viewpoint, these are set out slightly differently from DIN 50900 [7.80, 7.81].

1. Causes and Effects of Corrosion

While the formation of metal oxide layers in dry environments and at higher temperatures tends to increase chemical resistance to corrosion, relatively weak electrolytes are formed in conditions below the dew point, and these generally lead to electrochemical corrosion [7.260]. Corrosion is also fostered by the fact that different components have contacting surfaces with different properties, for instance due to the inclusion of various noble or base metals, to differences in crystalline structure, and to residual stresses set up, for instance, by heat treatment and welding. In addition, wherever the design calls for slits or holes, local differences in electrolyte concentration appear, even in the absence of clear differences in electric potential, due to the use of different materials.

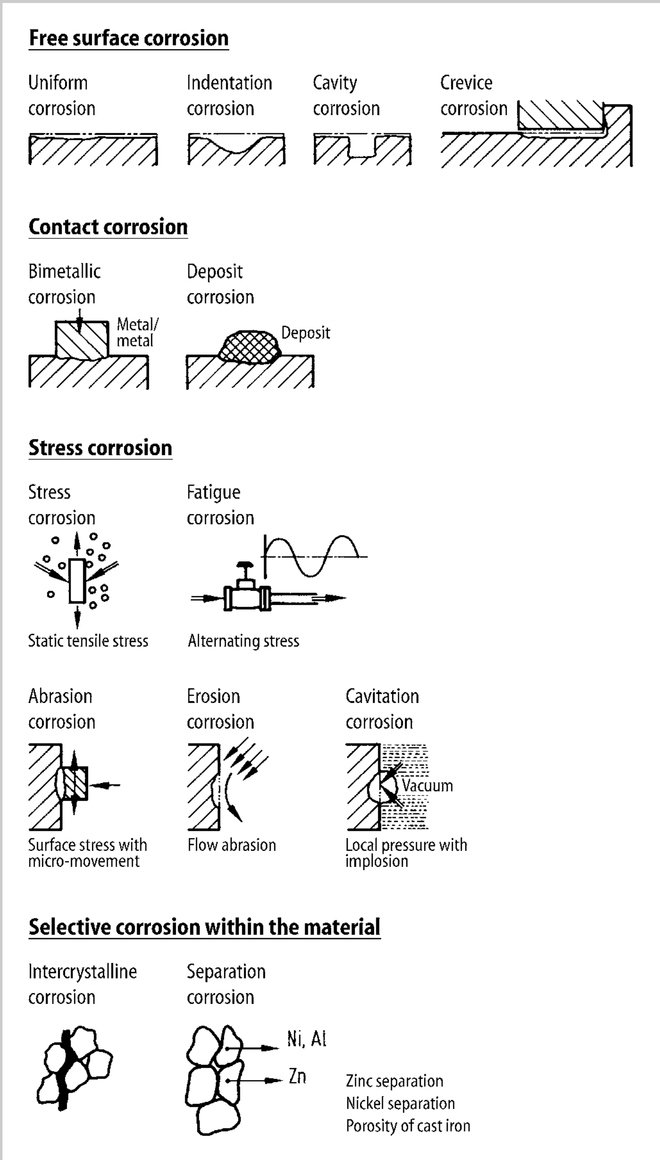


Figure 7.91. Types of corrosion

According to [7.80, 7.212] (see Figure 7.91) we must distinguish between:

- free surface corrosion
- contact corrosion
- stress corrosion
- selective corrosion within the material.

The preventive measures depend on the respective causes and effects. Various examples are given in the following sections.

2. Free Surface Corrosion

The corrosion of free surfaces can be uniform or locally concentrated. The latter is particularly dangerous because, in contrast to uniform corrosion, it leads to high stress concentrations and is often difficult to predict. It is, therefore, necessary to pay particular attention right from the start to potential danger zones.

Uniform Corrosion

Cause:
The presence of moisture (weak electrolytes) combined with oxygen from the air or the contacting medium, particularly below the dew point.

Effects:
Extensive uniform corrosion of the surface—in steel, for instance, approximately 0.1 mm per annum in a normal atmosphere. Sometimes more pronounced locally, especially in zones frequently kept below the dew point and hence subject to moisture concentration. Uniform corrosion is fostered by a more aggressive medium, higher flow velocity, and local heat transmission.

Remedies:

- Provide uniform service life by means of appropriate wall thicknesses and materials.
- Select a concept that obviates corrosion or makes it economically acceptable (see Example 1 below).
- Use small and smooth surfaces involving geometrical shapes with a maximum volume-to-surface area ratio (see Example 2 below).
- Avoid moisture traps (see Figure 7.92).
- Avoid temperatures below the dew point by good insulation and prevent hot or cold bridges (see Example 3 below).
- Avoid flow rates greater than 2 m/s.
- Avoid areas of high and differing thermal loads on heated surfaces.
- Apply a protective coating [7.82], possibly in conjunction with cathodic protection.

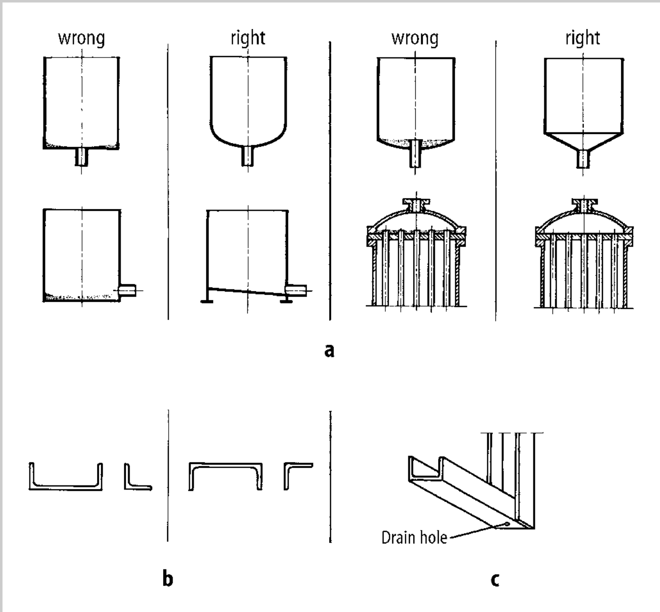


Figure 7.92. Drainage of components susceptible to corrosion: **a** design of bases encouraging and impeding corrosion; **b** wrong and right arrangement of steel sections; **c** brackets made of channel section with drain-hole

Indentation Corrosion

This type of corrosion is not uniform over the surface.

Cause:

There are components [7.81] with anodic and cathodic areas that cause differences in the rate of corrosion. These differences are usually caused by inhomogeneous material, by a medium with varying concentrations, and by local influences such as temperature and radiation.

Remedies:

- Remove inhomogeneity and varying influences.
- Provide a protective coating. Damage to this coating, however, will cause strong local corrosion (see cavity corrosion).

Cavity Corrosion

Cavity corrosion is concentrated on small surfaces with relatively deep indentations, with the depth being at least as great as the width. A clear distinction between indentation and cavity corrosion is not always possible.

Cause:

Similar to indentation corrosion, but its occurrence is more localised.

Remedies:

Basically the same as for indentation corrosion, although particular attention should be paid to reduction and prevention.

Crevice Corrosion

Cause:

Most often, the accumulation of acidic electrolytes (moisture, aqueous medium) following the hydrolysis of corrosion products in crevices etc. In rust and acid-proof steels, there is a breakdown of passivity due to depletion of oxygen in a crevice. Typically this type of corrosion is caused by insufficient ventilation.

Effects:

Increased corrosion in hidden areas. Increased stress concentrations in areas that are, in any case, under greater stress. Danger of fracture or separation without prior warning.

Remedies:

- Provide smooth, crevice-free surfaces and connections.
- Provide weld seams without permanent crevices; use butt seams or through-welded fillet seams (see Figure 7.93).
- Seal crevices, for instance by providing protruding parts with moisture-proof sleeves or coatings.
- Enlarge crevices so that throughflow prevents the accumulation of moisture.

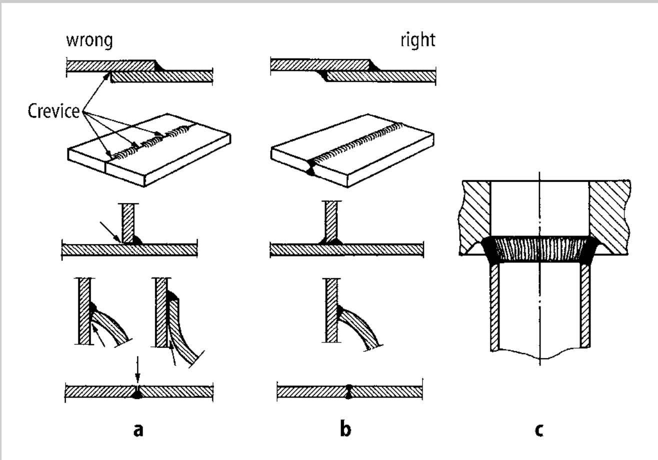


Figure 7.93. Examples of welded joints: **a** susceptible to crevice corrosion; **b** correct design, after [7.260]; **c** crevice-free welding of pipes, also improves resistance to stress corrosion cracking

3. Contact Corrosion

Bimetallic Corrosion

Cause:
The contact of two metals with different potentials in the presence of an electrolyte, that is, a conductive fluid or vapour [7.259].

Effects:
The baser of the two metals will corrode more rapidly than the nobler round the contact area, and this will occur more quickly for a smaller surface area (galvanic corrosion). Once again, the stress concentration is increased and corrosion products may be deposited. Such deposits have secondary effects of various kinds; for instance the production of sludge, contamination of the medium, etc.

Remedies:

- Use combinations of metals with small potential differences and hence a small contact current.
- Prevent action of electrolytes on the contact area by providing local insulation between the two metals.
- Avoid electrolytes altogether.
- If necessary, resort to planned corrosion by introducing still baser materials in the form of sacrificial anodes.

Deposit Corrosion

Cause:
Unwanted materials become deposited on the surface or in crevices and cause potential differences at particular locations. These deposits can come from existing corrosion, the surrounding medium, vaporisation residues, excess sealing material, etc.

Remedies:

- Avoid, filter, or collect the deposits.
- Prevent water traps, aim at smooth flow, maintain reasonable speed and self-drainage (see Figure 7.92a).
- Rinse or clean the components.

Transition Zone Corrosion

Cause:
Changes in the state of the medium or its components from the liquid to the gaseous phase and vice versa tend to increase the danger of corrosion of metallic surfaces in the transition zone. That danger may be increased further by encrustations in the transition zone [7.260].

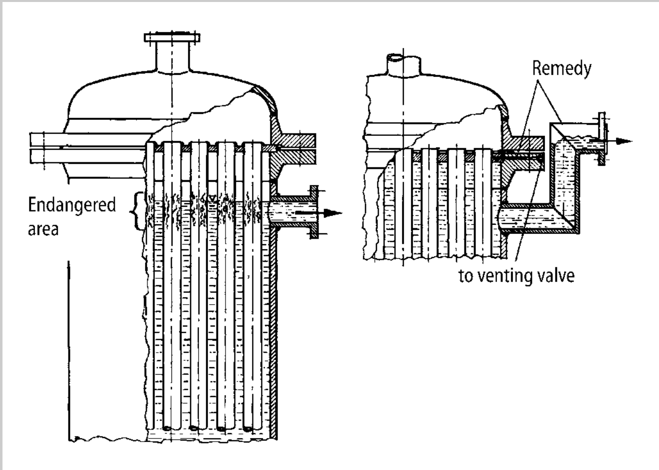


Figure 7.94. Increased corrosion at the transition from the gaseous to the liquid state, after [7.260] due to concentration of the medium in the region of the water line of a vertically arranged condenser. This can be remedied by raising the water level

Effects:

This type of corrosion is concentrated in the transition zone and is more pronounced with more sudden changes of state and more aggressive media [7.234].

Remedies:

- Gradually supply and remove heat using a heating or cooling element.
- Reduce turbulence, and hence heat transfer coefficients at the inlet of the affected medium, for instance by means of guide vanes.
- Provide corrosion-resisting jackets at critical points (see Examples 3 and 4).
- Avoid transition zone problems by appropriate design features (see Figure 7.94).
- Continuously change fluid level, for example by stirring.

4. Stress Corrosion

Components susceptible to corrosion are often mechanically loaded, either statically or dynamically. The mechanical stresses produced by these loads can cause several serious corrosion phenomena.

Fatigue Corrosion

Cause:

Corrosive attacks on a component subjected to mechanical fatigue loading appreciably reduce its strength. The greater the loading, the more intense the corrosion and the shorter the life of the component.

Effects:
Fracture without distortion, as in fatigue failure. Because the corrosion products, especially in slightly corrosive media, can only be seen under a microscope, this type of corrosion is often mistaken for normal fatigue failure.

Remedies:

- Minimise alternating mechanical or thermal stresses and especially avoid oscillatory stresses due to resonance phenomena.
- Avoid stress concentrations.
- Provide compressive stresses on the surface by shotblasting, roller burnishing, nitriding, etc., to increase the working life.
- Avoid contact with corrosive media (electrolytes).
- Provide surface coating (for example rubber, baked enamel, hot dip galvanisation, aluminium, etc.).

Stress Corrosion

Cause:
Certain sensitive materials tend to develop transcrystalline or intercrystalline cracks if static tensile stresses are combined with a specific trigger.

Effects:
Depending on the medium [7.260], various very fine and rapidly developing transcrystalline or intercrystalline cracks appear in the component. Adjacent parts are not affected.

Remedies:

- Avoid sensitive materials, which may not, however, be possible because of other requirements. These materials are: unalloyed carbon steels, austenitic steels, brass, magnesium, aluminium alloys and titanium alloys.
- Substantially reduce or completly avoid tensile stresses on the attacked surfaces.
- Introduce compressive stresses on the surface, for instance by shrink fits, by preloaded multilayer materials, or by shotblasting.
- Reduce residual tensile stresses by annealing.
- Apply cathodic coatings.
- Avoid corrosive influences by lowering the concentration and temperature.

Strain-Induced Corrosion

Cause:
Under repetitive large extensions or compressions, any protective outer layer cracks and opens repeatedly. This removes the protection and local corrosion will occur.

Remedy:
Reduce the magnitude of any extensions and compressions.

Erosion and Cavitation Corrosion

Corrosion may accompany erosion and cavitation, in which case the breakdown of the material is accelerated. The basic remedy is the avoidance or reduction of erosion and cavitation by hydrodynamic means or special design features. Only when this is not possible should such hard surface treatments such as metal spraying or hard chrome coating be considered.

Abrasion Corrosion

Cause:
Abrasive corrosion can be caused by relatively small movements between two surfaces subject to contact stresses (see also Section 7.4.1). Abrasion spots can appear, for instance, as a result of thermal expansion, or of pipes vibrating against their guides, etc. In either case, the oxidic protection layer on the surfaces of the rubbing parts may become damaged. Exposed metallic areas have a more negative electrochemical potential than those covered with a protective layer. If the fluid medium is an electrolyte, these relatively small exposed areas will be broken down electrochemically unless the protective layer can be regenerated.

Effects:
The affected surfaces form hard oxidation products (so-called abrasion rust) that speed up the process. At the same time, stress concentrations increase.

Remedies:
The most effective remedy is the removal of the abrasive movement, for example through elastic suspensions or hydrostatic bearings.
If the abrasive movement cannot be removed, the following measures should be adopted:

- Reduce the vibration of the pipes by reducing the flow velocity inside them and/or change the distances between the guides.
- Increase the gaps between the pipes and their guides so that no rubbing contact takes place.
- Increase the wall thicknesses of the pipes, thus increasing their stiffness and the tolerable corrosion rate.
- Use pipe materials that readily accept protective coatings.

5. Selective Corrosion within a Material

In the case of selective corrosion, only certain interfaces in the material matrix are affected. Of importance are:

- intercrystalline corrosion of stainless steels and aluminium alloys
- so-called "spongiosis"—graphite corrosion of cast iron when iron particles separate out
- dezincification of brass (zinc separation).

Cause:
Many material constituents or intercrystalline areas are less corrosion-resistant than the bulk material matrix.

Remedy:
Suitable selection of materials and their processing, such as adopting welding procedures which avoid producing a corrosion-sensitive material structure. Designers need to consult a materials expert when this type of corrosion is thought to be likely.

6. General Recommendations

In general, designers should aim at ensuring maximum and uniform lives for all components [7.234,7.235]. If it should prove economically impossible to meet these requirements with the appropriate choice of materials and layout, then designers must provide for the regular monitoring of all areas and components particularly prone to corrosion. This can done by visual inspection and regular measurements of wall thicknesses, directly by mechanical or ultrasonic methods and/or indirectly by means of corrosion probes that can be scrutinised and replaced at regular intervals.

Corrosion should never be allowed to proceed to the point where it threatens safety (see Section 7.3.3).

Finally, the reader is referred back to the principle of the division of tasks (see Section 7.4.2), with can enable even difficult corrosion problems to be solved. Thus, one component might provide protection against corrosion and provide a seal, while another provides support or transmits forces. As a result, the combination of high mechanical stresses with corrosion stresses is avoided, and the choice of materials for any one component becomes easier [7.207].

7. Examples of Design against Corrosion

Example 1

Lye is used to absorb CO_2 from a gaseous mixture under pressure, and the CO_2-enriched lye is then forced to surrender much of its CO_2 by expansion (regeneration). The position of the expansion chamber in a gas-washing plant is determined by the following factors.

If the lye were expanded immediately behind the washing tower (see Figure 7.95, point *A*) the pipework to *B* would have to withstand lower pressures and would therefore permit a saving in wall thickness. However, because of the release of

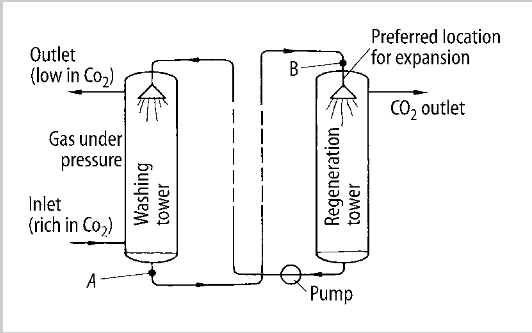


Figure 7.95. Influence of the point chosen for the expansion of CO_2-enriched lye on the choice of material for the pipework from A to B

CO_2, the aggressiveness of the lye permeated with CO_2 bubbles would increase to such an extent that the cheap unalloyed pipe steel commonly used would prove inadequate and hence have to be replaced with a more expensive rust- and acid-proof material. For that reason, it is far better to keep the CO_2-enriched lye under pressure until it enters the regeneration tower (point *B*).

Example 2

In this example, designers have to choose between two methods of storing compressed gases (see Figure 7.96): (a) 30 cylindrical containers, each with a capacity of 50 litres and a wall thickness of 6 mm; and (b) one spherical container with a capacity of 1.5 m^3 and a wall thickness of 30 mm. Solution (b) is less prone to corrosion for two reasons:

- The surface exposed to corrosive attack is approximately 6.4 m^2, which is about five times smaller than that in (a). In other words, less material is lost through corrosion to the same depth.

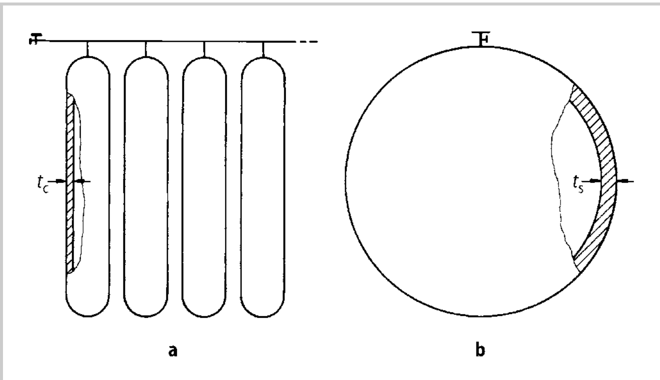


Figure 7.96. Influence of container shape on corrosion [7.234] for gases stored at 200 bar: **a** in 30 cylinders with a capacity of 50 litres each; **b** in a sphere with a capacity of 1.5 m^3

- For an anticipated corrosion depth of 2 mm in ten years, the loss of strength in (a) is such that the walls of the containers must be increased to a thickness of 8 mm, while corrosion to a depth of 2 mm in the 30 mm wall of container (b) is relatively insignificant. The spherical container can be dimensioned by considering strength requirements only and is therefore the better design of the two.

Example 3

Figure 7.97a shows a container holding a mixture of superheated steam and CO_2 [7.234]. The outlet is not insulated and cooling leads to the formation of a condensate with strong electrolytic properties. Corrosion will attack at the transition zone between the condensate and the gases, with the result that the outlet may break away.

Figure 7.97b shows a solution using insulation and Figure 7.97c one using separate components made of more durable materials.

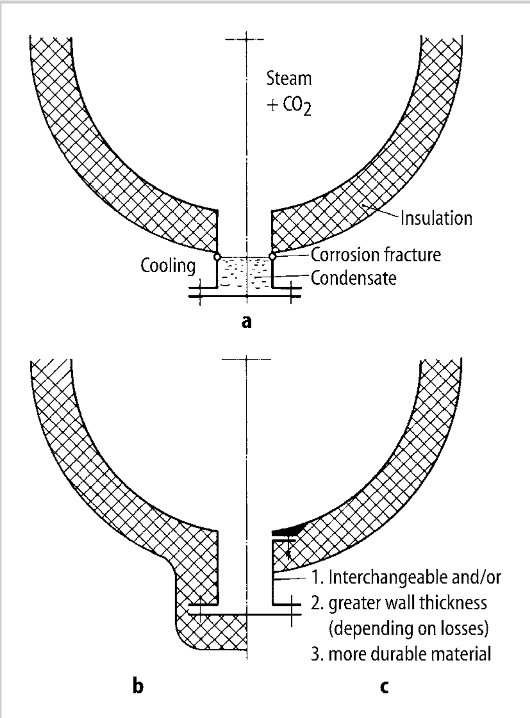


Figure 7.97. Outlet of a container for superheated steam and CO_2 under pressure: **a** original design; **b** insulated outlet avoiding condensation; **c** other corrosion-resistant variants with separate components

Example 4

In a heated pipe carrying moist gases, the inlet to the heated area is particularly prone to corrosion (see Figure 7.98a). A less sudden transition (see Figure 7.98b) or an extra protective sleeve (see Figure 7.98c) offer remedies.

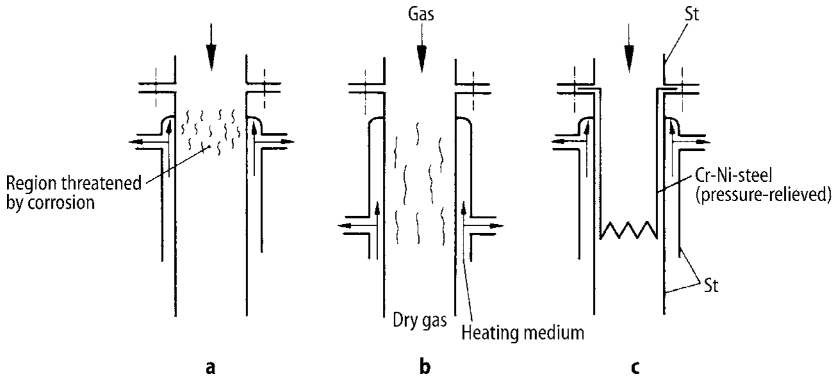


Figure 7.98. Corrosion in a heated pipe [7.234]: **a** severe corrosion at the inlet due to sudden transition; **b** sudden transition avoided; **c** protective sleeve covers critical zone and mitigates sudden transition

7.5.5 Design to Minimise Wear

1. Causes and Effects

The causes and effects of wear are many and complex. For a deeper understanding see the following literature: [7.28,7.121,7.153,7.258,7.314]. DIN 50320 [7.79] defines the types of wear and wear mechanisms. The main consequences of wear are shorter component lives, reduced functional performance, and higher losses. The most common and fundamental wear mechanisms affect component surfaces, in particular at the micro level. These mechanisms are described below.

Adhesive Wear

Adhesive wear is caused by high loading between two moving surfaces, which leads to microwelds (localised atomic binding) that are continuously broken by the relative movement between the surfaces. This results in surface damage and wear particles.

Abrasive Wear

Abrasive wear between components is caused by hard particles in the surface of one component (or in the medium) that micromachine (grind) the surface of the other component. This results in grooves and scoring in the direction of the relative movement. Mild abrasive wear can lead to a smoother surface and better surface mating; stronger wear leads to unacceptable surface damage.

Surface Disruption Wear

Surface disruption wear is caused by alternating mechanical stresses in the surface layers of the components. The effects are cracks, pitting, tears and wear particles.

Tribo-Chemical Wear

Tribo-chemical wear is caused by a chemical reaction between two components involving elements of the lubricant and/or the environment activated by friction (temperature increase). The effects are surface changes, such as hardened zones or wear particles. The latter in turn again increase the wear.

2. Design Features

Designing to minimise wear involves the application of tribological measures (system: material, working geometry, surface, lubricant/fluid) or material-related measures to minimise wear between loaded components subject to relative surface movements.

As with other effects, such as corrosion, the first step is to try to avoid the causes of the particular wear mechanism (*primary measures*); for example, by applying tribological measures to provide fluid friction between the moving surfaces and avoid dry or mixed friction. The elastohydrodynamic effect can, for example, provide fluid friction for sliding movements with the appropriate conditions for fluid viscosity, sliding speed and surface loading. If the layout or operating constraints do not allow this approach, a hydrostatic or magnetic solution might be chosen. In the case of small relative movements, the use of elastic joints should be considered.

When primary measures to remove the cause are not possible, *secondary measures* involving the materials and lubrication have to be applied to reduce the rate of wear. To reduce the wear rate, the local energy input due to the friction power per unit area, $p \cdot v_R \cdot \mu$, should be minimised by reducing the surface pressure (stress) p, the relative velocity v_R, and/or the coefficient of friction μ. Friction coefficients and wear coefficients for many common combinations of materials are provided in [7.28]. The wear coefficient is defined as:

$$\text{Wear coefficient} = \frac{\text{Sliding displacement} \times \text{Wear volume}}{\text{Normal force}}$$

When wear cannot be avoided, the following measures can prove helpful:

- Filter wear particles out of the fluid flow to avoid particle build-up and increased wear.
- Use the principle of the division of tasks (see Section 7.4.2) for structures with working surfaces that are in danger of wear; that is, the wear zones should be easy and economical to replace or be made out of a wear-resistant material.
- Allow the wear rate to be measured by using wear indicators and hence ensure operational safety and timely maintenance (see 7.5.10).

7.5.6 Design for Ergonomics

Ergonomics deals with the characteristics, abilities and needs of humans and, in particular, the interfaces between humans and technical products. A knowledge of ergonomics can lead to an embodiment that [7.173, 7.300]:

- adapts technical products to humans
- matches humans to products or activities by selection based on education and experience.

The range of technical products also includes domestic products and those used for hobbies and leisure.

The emphasis of ergonomic research is moving its focus from conventional physical activities in production facilities to working conditions in electronic industries and the design of user-friendly human–machine interfaces [7.56, 7.311]. This has, among other results, led to software tools for ergonomic workplace assessment and design [7.164].

1. Fundamentals

The starting point is the human being, where he or she is the operator, user or recipient. Humans can work with or be affected by technical products in many different ways (see Section 2.1.6). In this context it is helpful to address biomechanical, physiological and psychological issues.

Biomechanical Issues

The operation and use of products requires specific *body postures and movements*. These result from the spatial situation resulting from the embodiment of a product (for example the position and movement of controls) combined with the *body dimensions* [7.67]. This relationship can be represented and evaluated using templates of body dimensions [7.70] (see Figure 7.99).

The maximum forces humans can exert are given in [7.71]. To find the acceptable forces for a particular situation, however, also requires knowledge of frequency, duration, age, gender, experience and fitness, as well as knowledge of the methods used to calculate these influences [7.25, 7.127].

Physiological Issues

Body postures and movements resulting from the operation and use of technical products involve static and dynamic muscle actions. Muscle action requires the circulatory system to supply blood to the muscles based on the external loading. For static muscle action (for example when supporting a load), the blood throughput is throttled and recovery of the muscle is postponed. For this reason, large loads can only be sustained for short periods.

From an ergonomic point of view, it is important to distinguish between *loads, stresses and fatigue*. Loads are external influences. Loads produce stresses related to individual characteristics, such as age, gender, fitness, health, and training. The result of stress is fatigue, which depends on the intensity and duration of the stress. Recovery is achieved through *relaxation*. Fatigue-like situations, such as *monotony*, however, are not recoverable through relaxation but require a change of activity.

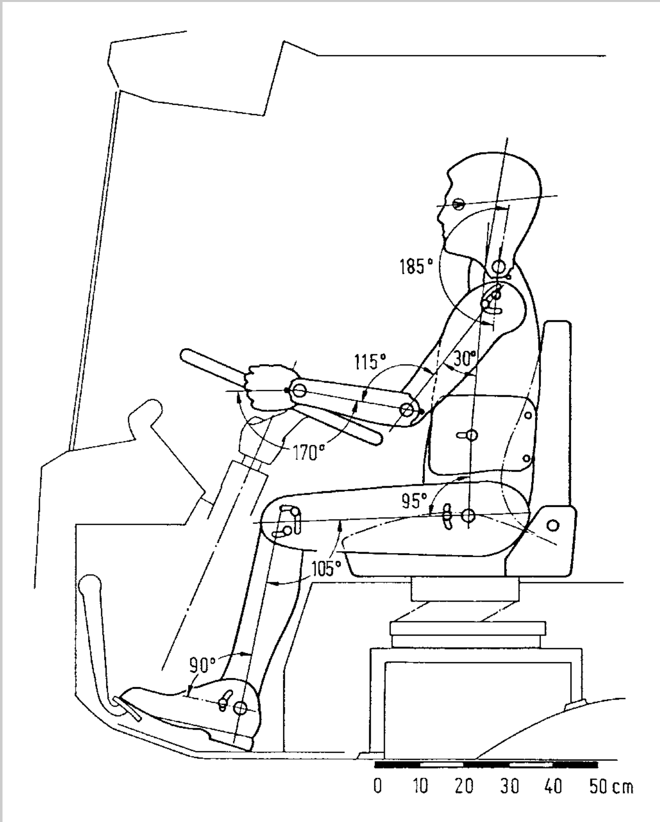


Figure 7.99. Application of a body template to evaluate the sitting position in a truck, after [7.70]

A further physiological requirement for human life and work is a normal body temperature of between 36°C and 38°C. Despite external *hot* and *cold situations*, and continuous heat generation within the body (increased during heavy work), the body temperature in the brain and other parts remains nearly constant because of the heat transferred by the blood. Working requirements and climatic influences have to be matched through technological measures, for example ventilation or organisational changes such as work breaks [7.68].

The *senses* also play an important role in work and leisure activities. Physiological variables involved in vision, for example, are minimum, optimum and maximum light density and light contrasts [7.44,7.45,7.69]. The variables related to hearing are noise level and noise differences [7.306], and these must be taken into account when designing acoustic warning signals in a noisy environment [7.69]. The relevant signals must be based on the sensor characteristics of human beings. No clear models exist for the processing of these signals in the nervous system and the brain. However, it is known that humans filter inputs to each of the sensors according to experience, interest, etc.

Psychological Issues

Several psychological issues have to be considered in the design of technical products. The use of sensors, for example, implies that the processing of the signals involves a series of steps that can be influenced in a variety of ways. Examples include optical illusions, not hearing or seeing unimportant things, and different interpretations. Guiding attention is therefore an important embodiment design guideline. This is true for the embodiment of control rooms [7.54], as well as for the placing of indicators and signs on products.

The process of sensing, deciding and acting usually proceeds undisturbed. When this process, which is partially unconscious, is disturbed, conscious thinking is used to bring certainty back into the process of perceiving, deciding and acting. In products where the structure and functionality cannot be seen from the outside, the cause of and remedy for unusual phenomena, such as disturbances, cannot be clarified by thinking. It is therefore necessary to convey the required information through sufficient and clear signs and through operating manuals. A well-designed product should minimise the thinking required for its operation so that thinking capacity can focus on the actual task. The requirement is for an obvious configuration; that is, one which during operation avoids thought processes that can be easily disturbed or are susceptible to errors. For example, the relation between the movement of a control and the resulting response should be obvious and simple.

Perception and thought focus on the actual action. Learning is defined as storing successful actions and knowledge for later use. For the operation and use of products, for example, one has to take into account that a series of actions learnt earlier may be reintroduced out of habit. Subsequent versions of a similar product should, therefore, avoid introducing unnecessary changes in operation or use; in particular, opposite movements or different positions for similar control actions should be avoided. Such changes must never be introduced if the consequences of an error could lead to direct or indirect safety risks.

Directing and constraining human activities excessively using technical or organisational systems can have a negative effect on motivation and behaviour, especially over a long period. All such activities should therefore leave space for free actions.

2. Human Activities and Ergonomic Constraints

Humans can be involved or affected by technical processes, either actively or passively. In an active relationship they can act and are deliberately involved in the technical product. That is, they execute certain functions such as activating, controlling, monitoring, loading, removing, registering, etc. In general, the following *repetitive* activities are undertaken in an activity cycle:

- Preparing for the activity, e.g. going to work.
- Gathering and processing information, e.g. observing and orienting, drawing conclusions, deciding on an action.

- Undertaking the activity, e.g. activating, connecting, separating, writing, drawing, talking, giving signs.
- Checking results, e.g. identifying status, checking measured values.
- Stopping the activity or starting a new one, e.g. cleaning, closing, going away, starting a new activity cycle.

When the involvement of human beings is functional—in other words, deliberate—then this involvement should be planned carefully and suitable arrangements made. This should start early in the design process, even when clarifying the task (see Chapter 5). It is often necessary to represent this involvement in the function structure (see Section 6.3).

Active Human Involvement

Whether it is sensible and useful to involve humans in technical systems has to be assessed from the viewpoints of *effectiveness, efficiency and humanity* (dignity and appropriateness). This initial and basic consideration influences and determines, to a large extent, the involvement of humans and thereby the solution principle. The following ergonomic aspects can be useful in the generation of solutions and as evaluation criteria [7.300] (see Table 7.4):

- Is human involvement necessary or desirable?
- Will the involvement be effective?
- Is involvement easy to achieve?
- Can the involvement be sufficiently precise and reliable?

Table 7.4. Ergonomic aspects for the requirements list and the evaluation criteria [7.300]

	Active human involvement in a technical system intended to fulfil a task:
•	necessary, desired
•	effective
•	simple
•	fast
•	precise
•	reliable
•	error-free
•	clear, sensible
•	learnable
	Active or passive involvement through disturbing effects and side-effects on humans:
•	tolerable stress
•	low fatigue
•	low annoyance
•	no physical danger, safe
•	no health risk or loss
•	stimulation, change, holding attention, no monotony
•	personal development

- Is the activity clear and sensible?
- Can the activity be learnt?

Only when the answers to these questions are positive should the involvement of humans in technical systems be considered.

Passive Human Involvement

Not only those actively involved, but also those passively involved will experience *disturbing effects* and *side-effects* from technical systems (see terminology in Section 2.1.6). The effects of energy, material and signal flows and the environment, such as vibrations [7.292], light [7.43–7.45], climate [7.68] and noise [7.306] are very important. These effects have to be identified early on so that they can be considered during the selection of the working principle and the development of the embodiment. The following questions can be useful and can also serve as evaluation criteria (see Table 7.4):

- Are distresses tolerable, and is the emerging fatigue recoverable?
- Has monotony been avoided, and is stimulation, change and attention ensured?
- Are annoyances or disturbances few or nonexistent?
- Has physical danger been avoided?
- Has health risk or loss been excluded?
- Does the work allow for personal development?

When these questions cannot be answered satisfactorily, then another solution should be selected, or the existing solution considerably improved.

3. Identifying Ergonomic Requirements

In general it is not easy for designers to immediately find satisfactory answers to the questions listed above. As described in Guideline VDI 2242 [7.300], the problem of identifying the most important influences and suitable measures can be approached in two ways, as discussed below.

Object-Based Approach

In many cases the technical system (object) that has to be ergonomically embodied is known and documented, e.g. a control panel, a driver's seat, a piece of office equipment, or an item of protective clothing. In such cases it is useful to apply the checklist in Part 2 of Guideline VDI 2242 [7.301]. It is also important to be sensitive to the particular requirements of the system under consideration and to make use of Table 7.5. Just reading the guidelines can be very instructive and can help clarify the issues. Concrete design features can be based on the insights acquired or obtained from the literature listed below.

Table 7.5. Characteristics used to identify ergonomic requirements [7.300]

Characteristics	Examples
Function	Division of functions, type of functions, type of activities
Working principle	Type and intensity of the physical or chemical effects, consequences such as vibration, noise, radiation, heat
Embodiment	
• Type	Type of elements, configuration, type of operation
• Form	Ergonomic overall form and elements, division based on symmetry and proportion, aesthetically pleasing
• Position	Configuration, arrangement, distance, direction of effect and visibility
• Size	Dimensions, working area, contact surfaces
• Number	Amount, division
Energy	Adjustment force, adjustment direction, resistance, damping, pressure, temperature, moisture
Material	Colour and surface finish, contact properties such as safe to touch, easy surface to hold
Signals	Labelling, text, symbols
Safety	Free of danger, avoiding danger sources and spots, inhibit dangerous movements, protective measures

Effect-Based Approach

In new situations—that is, when no system has been defined—it is useful to adopt the following approach instead. The effects related to existing and thus known energy, material and signal flows of technical systems are identified and compared with the ergonomic requirements. When there are limitations, intolerable loads or even dangers to safety, other solutions must be sought. Effects such as mechanical forces, heat and radiation are derived from the individual types of energy and the forms in which they appear. The material flow has to be checked to identify whether the suggested materials are flammable, easy to ignite, poisonous, cancer-causing, etc. For this purpose, Guideline VDI 2242 Part 2 [7.301] provides a checklist of effects, which gives an indication of existing problems and refers to literature with possible solutions.

The following literature are also useful:

Design of work space	[7.65, 7.72, 7.127, 7.172, 7.243, 7.300]
Work physiology	[7.53, 7.231]
Illumination	[7.20, 7.37, 7.43–7.45, 7.55]
Computer workplace	[7.52, 7.83, 7.84]
Climate	[7.68, 7.246]
Operation and handling	[7.24, 7.65–7.67, 7.70, 7.71, 7.78, 7.140, 7.195]
Vibration and noise	[7.31, 7.306, 7.310]
Monitoring and control	[7.69, 7.73, 7.74].

7.5.7 Design for Aesthetics

1. Aims

Technical products should not only fulfil the required technical functions as defined by the function structure (see Section 6.3) but also be aesthetically pleasing to their users. A considerable change has occurred recently in user expectations and in the way that products are judged.

VDI Guideline 2224 [7.296] focuses on the aesthetics of products. Starting with a technical solution, the guideline provides rules for the external form or shape; for example, it should be compact, clear, simple, unified, in line with function, and compatible with materials and with production processes.

In many products nowadays, aesthetics are as important as technical functionality. This is particularly true for products aimed at large markets and used directly by users in their daily lives. In such cases the emphasis is not only on aesthetics and use, but also on factors such as prestige, fashion and lifestyle. The forms—or better the embodiments—of consumer products are determined primarily by industrial designers, artists and psychologists. While ensuring the technical functionality, they select the forms, shapes, colours and graphics—that is, the overall appearance—based on human feelings and values. Expression and style play an important role; for example a military appearance may be applied to radio products, a space age look to lights, a safari image to cars, or a nostalgic feel to telephones. The body of a car, for example, is strongly based on artistic and psychological criteria and not only on technical criteria such as low air resistance and transportation efficiency.

It is clear that all of the requirements regarding function, safety, use and economy have to be fulfilled. The aim of designers, however, is to create products that appeal to customers. Given this aim, industrial design lies between engineering and art, and has to address ergonomic and visual issues in the same way as engineering design has to address function and safety issues. In addition, the company image has to be promoted in order to underline the individuality of its products. Such complex requirements suggest that the involvement of industrial designers should not be left until the end of the design process. They should be part of the design team and involved from the beginning of the task clarification phase. In special circumstances they can even help formulate the task or undertake preliminary design studies.

The result of this approach is a design process that proceeds from "outside to inside". Continuous collaboration between industrial designers and engineering designers is required to ensure that the requirements of appearance, expression and impression still allow the technical functions to be fulfilled within the forms and shapes created.

In this collaboration, engineering designers should not try to replace industrial designers, but should focus on developing the technical and economic aspects of the product. In the same way that technical solutions are developed, visual variants have to be proposed and evaluated, and models and prototypes made to decide on the final appearance of the product. When searching for solutions, the same methods as those described for the engineering design process can be used, such as brainstorming, stepwise development of variants through sketches, and systematic variation of configuration, form and colour.

Tjalve [7.280] gives a very clear example of such a development (see Figure 7.100). This clarity is evident throughout his book and illustrates the way in which form

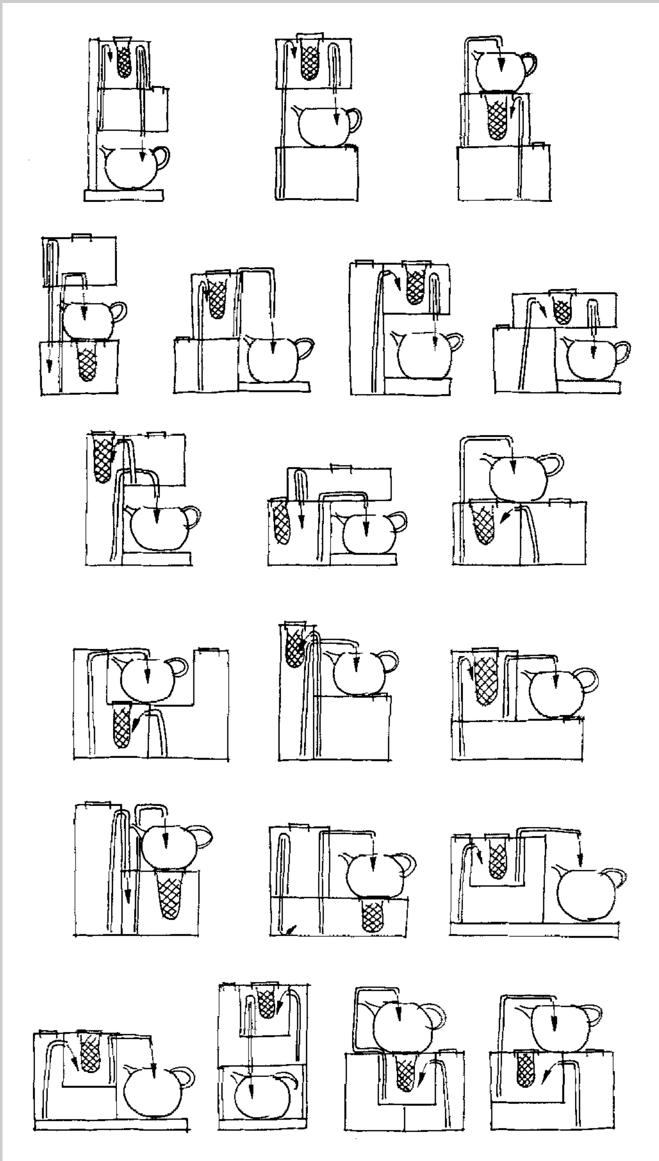


Figure 7.100. Systematic variation of the structure of an automatic teamaker (after [7.280]), investigating the configuration of the water kettle, the tea container and the teapot

and embodiment can be varied. He emphasises that the following factors influence each other and determine the appearance of the product:

- engineering (purpose, function, construction structure)
- production (process, assembly, cost)
- sales and distribution (packaging, transport, storage, company image)
- use (handling, ergonomics)
- disposal (recycling).

Seeger [7.251] underlines the close link between design for ergonomics and design for user-friendliness. Klöcker [7.152] focuses more on physiological and psychological aspects. In [7.252, 7.253] Seeger discusses the basic knowledge used for the development and embodiment of industrial products. Their appearance is developed from structure, form, colour and graphics. The impressions experienced by observers are of crucial importance. Information on this topic can be found in the literature from the partially overlapping areas of physiology, psychology and ergonomics. In his book entitled *Product Quality and Design* [7.111], Frick emphasises the importance of systematic collaboration between industrial and engineering designers in the context of an interdisciplinary development process. Using a series of examples, he proposes methods, procedures and tools to support such a collaboration.

2. Visual Information

In general, the technical function and the selected technical solution, together with its construction structure, determine the configuration and form and hence the appearance of assemblies and components. This results in a *functional embodiment* that is often difficult to change. An example of a simple functional embodiment is a spanner (lever arm and shape of bolt head), and a complex one is a dredger (kinematic requirements, shape of dredging buckets, power train, location of operator, etc.). Human beings not only see this functional embodiment but also other visual impressions, such as stability, compactness and a modern or striking appearance. They also expect information on operational procedures, safe areas, potential dangers, etc., which together form the *information presentation*.

In the embodiment design phase, the information presentation that is required or desired should be integrated with the functional embodiment. Based on Seeger [7.251], we list the essential information presentation areas and some related rules.

Market and User Information

When determining this type of information presentation, it is important to consider the type of user being addressed, such as technical expert, prestige seeker, nostalgia lover, and the avant-garde. In general, the overall appearance should be:

- simple, uniform and pure, and it should embody style
- structured and well-proportioned
- identifiable, definable and approachable.

Purpose Information

This information presentation should enable the purpose of the product to be easily recognised and understood. The outer shape, colouring and graphics should support identification of the functions and the actions involved, such as where a tool should be located and which parts exert forces.

Operation Information

The information presentation about the correct operation and intended use should:

- be centrally located and recognisable, for example control elements should have a function-related layout
- be ergonomically appropriate, in accordance with the action space of human limbs
- be labelled clearly, for example gripping and stepping areas
- identify the operational status
- use safety signs and colours [7.40, 7.42].

Manufacturer and Distributor Information

This information presentation expresses origin or house style. It contributes to continuity, confidence in known quality, involvement in the further development of successful products, and membership of a group. This can be achieved by easily recognisable and repeated elements, though the style and expression can be adapted to current fashions.

3. Guidelines for Aesthetics

Information presentation is achieved through specific and intended *expression*, such as lightness, compactness and stability, and by related *structure, form (shape), colour* and *graphics*. The following recommendations need to be considered (see Figures 7.101 to 7.103).

Select an Expression

- Provide a recognisable and uniform expression that creates an impression in the observer that is in accordance with the aim; for example an impression of being stable, light and compact.

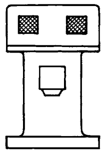
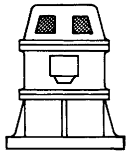
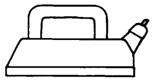
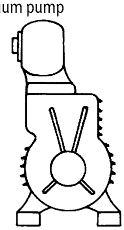
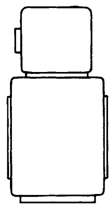
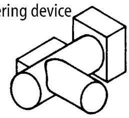
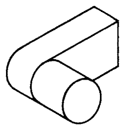
Embodiment Guidelines	Wrong	Right
Select an expression		
Provide a recognisable and uniform expression	Vertical three-phase AC motor  Unstable, top-heavy	 Stable, compact
	Iron:  Heavy, immobile	 Light, easy to handle
Structure the overall form		
Structure in a identifiable way	Vacuum pump  Unidentifiable	 Box shape
Divide into clearly distinguishable areas	Steering device  Unidentifiable	 Clear arrangement L-shape

Figure 7.101. Embodiment guidelines for aesthetics: expression and structure

Structure the Overall Form

- Structure in an identifiable way, such as in a block shape, a tower shape, an L-shape, a C-shape, etc.
- Divide into clearly distinguishable areas with identical, similar or adapted form elements.

Unify the Form

- Minimise variations in form and position; for example, use only circular shapes with horizontal orientation along the main axis, or only rectangular forms with vertical orientation.

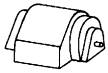
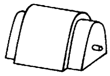
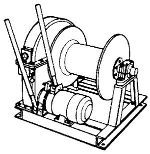
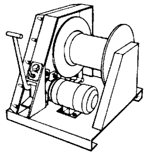
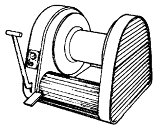
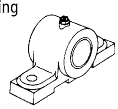
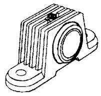
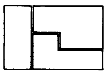
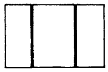
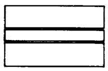
Embodiment Guidelines	**Wrong**	**Right**
Unify the form		
Minimise the number of different forms	Generator 	
	Rope winch 	 Open structure  Closed structure
Aim for similar forms and contours	Bearing 	
Adjust lines	Air conditioner  Confusing, inhomogeneous	 Block form  Layer form

Figure 7.102. Embodiment guidelines for aesthetics: form

- Introduce form elements and alignments appropriate to the basic form selected, for example, use the split lines of assemblies. Arrange the form by bringing several edges to one point or by running them parallel to one another. Support the intended expression with form elements and appropriate lines, such as horizontal lines to emphasise length. Keep an eye on the overall profile.

Support Using Colour

- Match colours to form.
- Reduce colours and material differences.

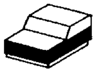
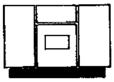
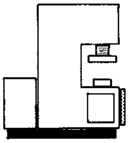
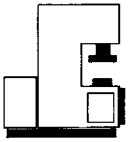
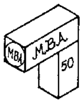
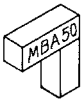
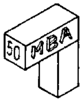
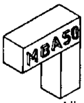
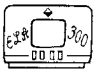
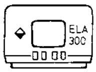
Embodiment Guidelines	**Wrong**	**Right**
Support using colour		
Match colour to form		
Reduce colours and material differences		
Choose one main colour supported by complementary colours		
Complement with graphics		
Use uniform styles for fonts and graphic symbols		
Unity expression		
Adjust size, form and colour of the graphics to the other forms and colours		

Figure 7.103. Embodiment guidelines for aesthetics: colour and graphics

- When using several colours, choose one main colour supported by complementary colours. For contrast use black and white, for example use black to contrast yellow, white to contrast red, green, blue, etc. (see also safety colours).

Complement with Graphics

- Use uniform styles for fonts and graphic symbols.
- Unify expression by using the same processes for the graphics, for example, etching, painting or embossing.
- Adjust size, form and colour of the graphics to the other forms and colours.

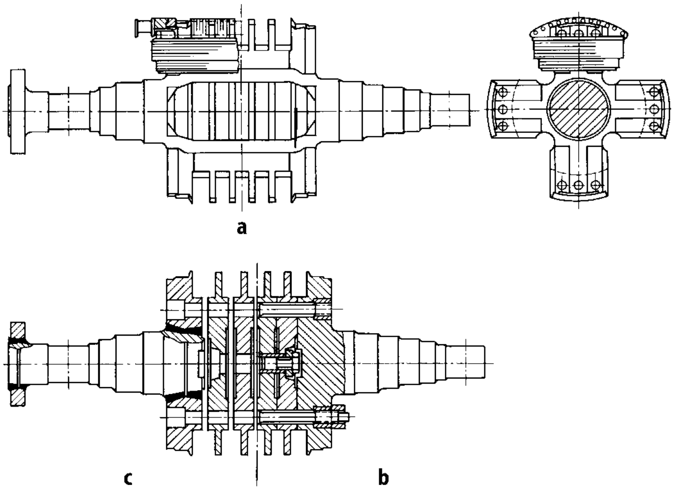


Figure 7.104. Rotor of a synchronous generator, after [7.8] (AEG-Telefunken): **a** as a forged part; **b** as a disc construction with forged flanges; **c** as for *b* but with welded flanges

7.5.8 Design for Production

1. Relationship Between Design and Production

The crucial influence of design decisions on *production costs*, *production times* and the *quality of the product* is described in [7.307,7.313]. *Design for production* means designing for the minimisation of production costs and times while maintaining the required quality of the product.

The term *production* usually refers to:

- the production of components in the narrow sense by accepted processes [7.49] (primary forming, secondary forming, material removal, joining, finishing, changing material properties)
- assembly, including transport of components
- quality control
- materials logistics
- operations planning.

Designers would therefore do well to consult the checklist (see Figure 7.3) under the headings *Production, Quality Control, Assembly* and *Transport*. In what follows we shall first concentrate on the design of components or assemblies in the narrower sense, while paying due regard to quality control and improvement of the overall production procedure. In Section 7.5.9 we shall then examine design features for improved assembly and transport.

Design for production is greatly facilitated if, from the earliest possible stage, decisions are backed up with data compiled by the standards department, the planning and estimating department, the purchasing department and the production manager. Figure 1.4 shows how the flow of information can be improved by systematic means, appropriate organisational measures and integrated data processing.

By observing the basic rules of simplicity and clarity (see Section 7.3), designers are already proceeding along the correct lines. The principles of embodiment design (see Section 7.4) can also lead them to a better and safer fulfilment of a given function and to the best solution from a production point of view. Another step in the same direction is the application of general and company standards (see Section 7.5.13).

2. Appropriate Overall Layout Design

The overall layout design, developed from the function structure, determines the division of a product into assemblies and components and:

- identifies the source of the components; that is, whether they are in-house, bought-out, standard or repeat parts
- determines the production procedure; for instance whether the parallel production of individual components or assemblies is possible
- establishes the dimensions and the approximate batch sizes of similar components, and also the means of joining and assembly
- defines suitable fits
- influences quality control procedures.

Conversely, production limitations such as the capacity of machines, assembly and transport facilities, etc., naturally have repercussions on the choice of the overall layout.

The appropriate subdivision of the overall layout can give rise to *differential*, *integral*, *composite* and/or *building-block* methods of construction.

Differential Construction Method

Differential construction refers to the breakdown of a component (a carrier of one or several functions) into several easily produced parts. This idea comes from lightweight engineering [7.135, 7.325], where this approach was introduced for the purpose of optimising load-carrying capacity. In both cases, we are entitled to speak of the "principle of subdivision for production".

To show an example of the differential method, let us consider the rotor of a synchronous generator (see Figure 7.104). The large forging shown at the top *a* is divided into several rotor discs consisting of simple forged parts and two considerably smaller flanged shafts *b*. Each of the latter can also be subdivided into shaft, disc support flange and coupling flange, in the form of a welded construction *c*.

The reason for this differential construction might be the market situation of large forgings (price, delivery date), and the easier adaptation of the generator to various output requirements (rotor sizes) and types of coupling. A further advantage is that the parts can be produced as stock and not necessarily to a specific order. However, the illustration also demonstrates the limitations of the differential approach—beyond a certain rotor length and diameter, the machining costs become too great and the stiffness of the joints too problematical.

Another example is shown in Figure 7.105. In the winding machine *a*, the winding head is integrated with the drive unit on a common shaft. The differential solution *b* was developed to facilitate the parallel production of drive units and winding heads to meet various customer requirements. In this way, a small number of standard drive units can be combined with a large number of winding heads.

The differential construction method also influences the production time. Figure 7.106 shows an example of the production procedure for a medium-powered electric motor. The times spent on acquiring the material and on producing the

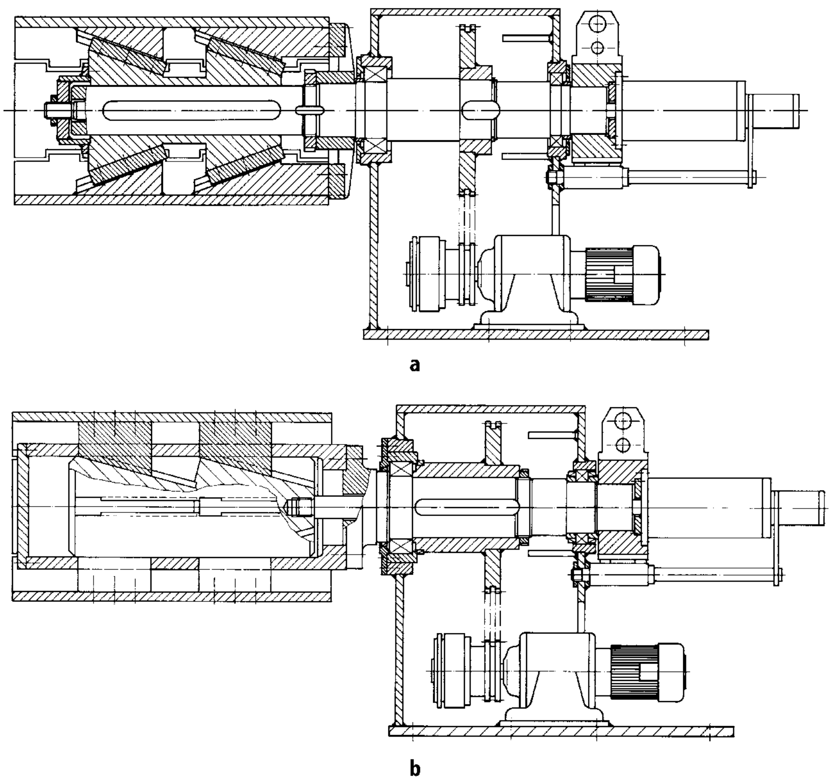


Figure 7.105. Winding machine (Ernst Julius KG): **a** winding head with integrated drive unit; **b** winding head with separate drive unit

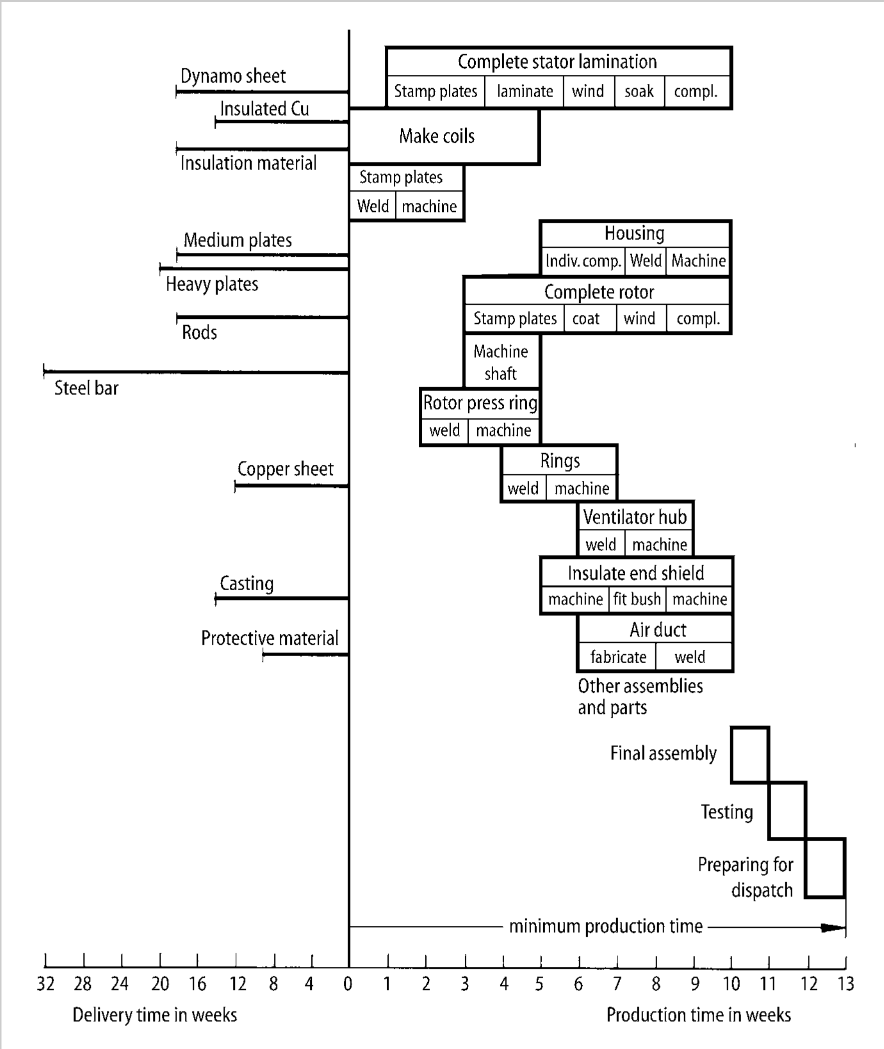


Figure 7.106. Production procedure for an electric motor from the series shown in Figure 9.17 (AEG-Telefunken)

components and assemblies are indicated by the lengths of the horizontal lines. The diagram not only makes clear where improvements can be made by choosing more quickly procurable raw and semi-finished materials or by keeping these materials in stock, but also where different production steps could be taken in parallel. Thus, by allowing the stator laminations to be built up in parallel with the construction of the housing (two time-consuming operations), a significant reduction in the overall production schedule is possible in comparison to older designs in which the stator laminations could only be inserted, followed by the windings, after the casing had been welded. All in all, differential designs have the advantages, disadvantages and limitations listed below:

Advantages:

- use of easily available and favourably priced semi-finished materials or standard parts
- easier acquisition of forged and cast parts
- easier adaptation to existing factory layout (dimensions, weight)
- increase in component batch sizes
- reduction in component dimensions allowing easier assembly and transport
- simpler quality assurance (more homogeneous materials)
- easier maintenance, for instance by simple replacement of worn parts
- easier adaptation to special requirements
- reduced risk of missing delivery dates
- reduced overall production time.

Disadvantages and limitations:

- greater machining outlay
- greater assembly costs
- greater need for quality control (smaller tolerances, necessary fits, etc.)
- limitations of function because of joints (stiffness, vibration, sealing).

Integral Construction Method

By the term *integral construction* we mean the combination of several parts into a single component. Typical examples are cast constructions instead of welded constructions, extrusions instead of connected sections, welded instead of bolted joints, etc. This method is often used for product optimisation because of the economic benefits of integrating several functions into one component. This method can indeed be an advantage for specific technical, production and procurement situations, particularly for labour-intensive production.

Figure 7.107 shows an example chosen from electrical engineering. Here, a cast and welded construction has been replaced with a single cast component. Though the casting is fairly complicated, it leads to a cost reduction of 36.5%. Naturally, this percentage will vary with the size of the batch and with market conditions.

Another example is the rotor of a hydroelectric generator (see Figure 7.108). Four different constructions with the same generator output and identical radial loads were investigated. Variant *a* has numerous individual support discs and may therefore be considered to be a differential construction. In variant *b*, the degree of division is reduced by the use of cast steel hollow shafts, two support rings and

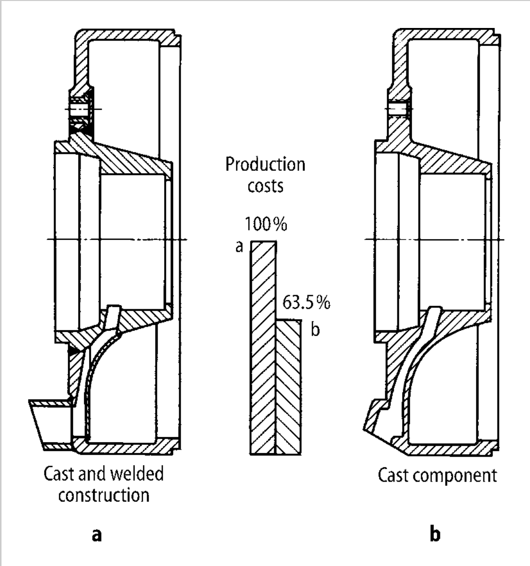


Figure 7.107. End cover of an electric motor, after [7.154] (Siemens): **a** composite construction; **b** integral construction

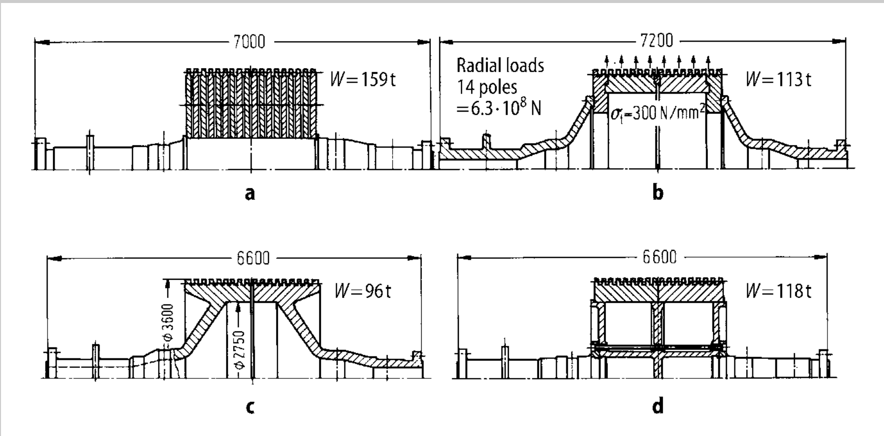


Figure 7.108. Rotor construction for a large-scale hydroelectric generator (Siemens)

end discs. Variant *c* is an integral construction in that two cast hollow bodies have been bolted together. In variant *d*, the cast construction is split up again (a cast central part, two forged shafts and two support rings). Weight comparisons show that the integral method saves material. In the end, however, variant *d* was chosen because of difficulties with procuring large castings.

The advantages and disadvantages of the integral construction method are easily determined through a reversal of the advantages and disadvantages of the differential method.

Composite Construction Method

By *composite construction* we mean:

- the inseparable connection of several, differently made, parts into a single component necessitating further work; for instance, the combination of cast and forged parts
- the simultaneous application of several joining methods for the combination of components [7.221]
- the combination of various materials for the optimal exploitation of their properties [7.290]

Figure 7.109 gives an example of the first method: the combination of cast steel components and rolled steel sheet into a welded construction.

Further examples are bogies with cast centres and welded arms, and also the welding of cast bar joints used in steel structures. Examples of the second method are combinations of adhesives and rivets or of adhesives and bolts. The combination of several materials into a single part is exemplified by synthetic components with cast-in thread inserts; by composite sound-absorption panels which have two plates separated by a plastic core; and also by rubber/metal components.

Another economical design of the composite type is the use of steel in prestressed concrete [7.120].

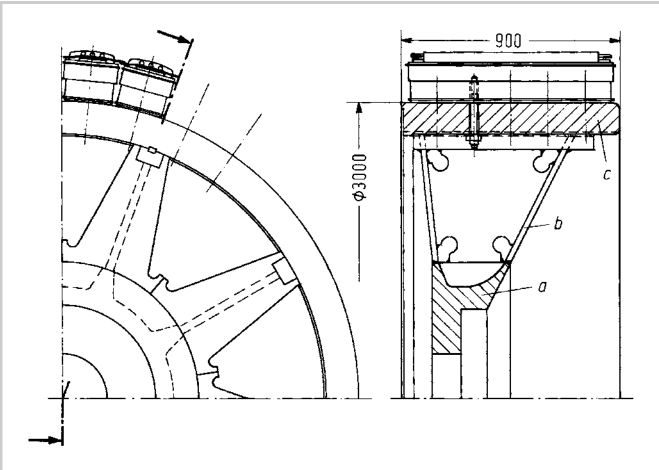


Figure 7.109. Magnet wheel of a hydroelectric generator of composite construction, after [7.15] (AEG-Telefunken): **a** Hub of cast steel; **b** Spoke of rolled steel sheet; **c** Support of cast steel

Building Block Construction Method

If the differential method is used to split a component in such a way that the resulting parts and/or assemblies can also be used in other products or product variants, then they can be considered to be building blocks. These are particularly useful if they are economical to produce. In a sense, the utilisation of repeat parts from stock can also be considered to be a building block construction method (see Section 9.2).

3. Appropriate Form Design of Components

During the form design of components, designers exert a great influence on production costs, production times and the quality of the product. Therefore, their choices of shapes, dimensions, surface finishes, tolerances and fits affect the selection of:

- production procedures
- types of machines, including tools and measuring instruments
- in-house components and bought-out components, preferably making use of repeat parts from within the company or suitable standard and off-the-shelf components
- materials and semi-finished materials
- quality control procedures.

Conversely, production facilities influence the design features. Thus, the available machine tools might limit the dimensions of components, necessitating that they be split up into several connected parts or that bought-out components be acquired. Many guidelines are available for the appropriate form design of components [7.19, 7.21, 7.123, 7.180, 7.198, 7.201, 7.262, 7.281, 7.283, 7.285, 7.287, 7.288, 7.291, 7.331–7.333]. Because of the importance of tolerances (geometry, dimension, position and surface) for the production and assembly of components, we specifically suggest the following literature: [7.36, 7.38, 7.39, 7.47, 7.143, 7.144].

It is important to use a *tolerancing basis* appropriate for the specific requirements [7.143]. A distinction is made between the *independent basis*, where dimensions are toleranced and checked individually, and the *envelope basis*, where geometrical features (such as a circle or pair of parallel surfaces) have an enveloping tolerance zone (maximum material condition) within which the dimension must lie. The latter cannot control deviations in position. For both tolerancing bases, deviations of position are independent of dimensional tolerances. The difference is whether deviations of geometry should be within the envelope. A fit has to remain within the envelope and, using the independent basis, this is indicated on the drawing with a fit specification, for example H7-j8. When the independent basis is used, *blanket tolerances* for geometry and position should be indicated. The envelope basis only requires a blanket tolerance for position [7.143, 7.144].

In keeping with the aims of this book, we shall present only essential design suggestions arranged systematically in the form of charts. Our classifying criteria will be production processes [7.48–7.50] with their individual *process steps*

(PS). In addition, we shall be assigning objectives—*reduction of costs* (C) and *improvement of quality* (Q)—to the various design guidelines. When designing components, designers should always bear these process steps and objectives in mind.

Form Design for Primary Shaping Processes

The form design of components to be shaped by primary processes, for example casting and sintering, must satisfy the demands and characteristics of the processes used.

In cast components (primary shapes obtained from the fluid state), designers must allow for the following process steps: *pattern* (Pa), *casting* (Ca) and *machining* (Ma). Figure 7.110 lists the most important design guidelines. The literature cited contains further information.

When designing *sintered* components (primary shapes obtained from the powder state), designers must allow for *tooling* (To) and *sintering* (Si). In particular, they must be guided by the latest findings in powder technology. The essential guidelines are shown in Figure 7.111.

Form Design for Secondary Shaping Processes

The form design of components to be shaped by secondary processes (hammer (free) forging, drop forging, cold extrusion, drawing and bending) must adhere to the guidelines listed below. Special considerations for the design of ferrous materials can be found in DIN 7521 to 7527 [7.46] and the design of nonferrous metals in DIN 9005 [7.51]. With *hammer forging*, designers need only allow for the actual forging process, since no complicated devices, for instance dies, are involved. The following design guidelines should be observed:

- Aim at simple shapes, if possible with parallel surfaces (conical transitions are difficult) and with large curvatures (avoid sharp edges). *Objectives*: reduction of costs, improvement of quality.
- Aim at light forgings, perhaps by separation and subsequent combination. *Objective*: reduction of costs.
- Avoid excessive deformations or excessive differences in cross-sections due, for instance, to the presence of excessively high and fine ribs or of excessively narrow indentations. *Objective*: improvement of quality.
- Try to place bosses and indentations on just one side. *Objective*: reduction of costs.

Design guidelines for *drop forging* have been collated in Figure 7.112. They allow for the process steps of: *tooling* (To), *forging* (Fo) and *machining* (Ma).

Figure 7.113 lists design guidelines for the cold extrusion of simple rotationally symmetrical solid and hollow bodies. They allow for the process steps of *tooling* (To) and *extrusion* (Ex). It must be stressed that only certain types of steel can

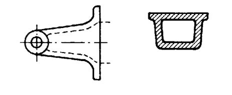
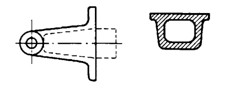
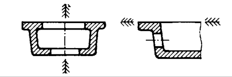
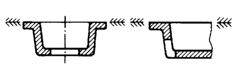
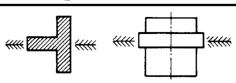
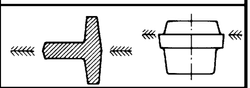
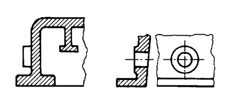
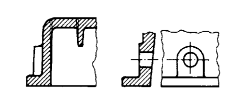
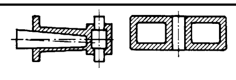
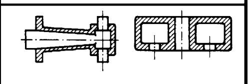
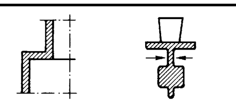
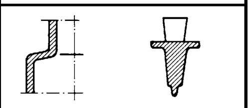
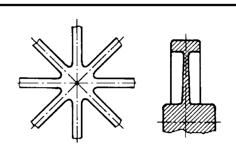
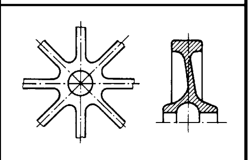
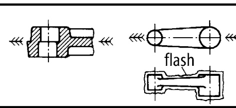
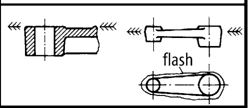
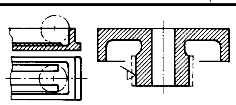
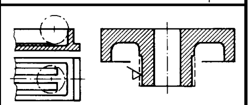
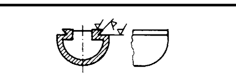
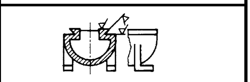
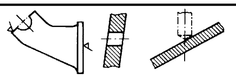
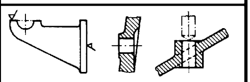
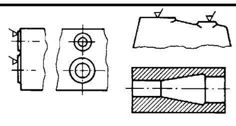
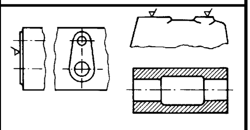
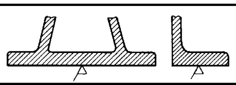
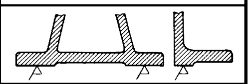
PS	Guidelines	Objectives	Wrong	Right
Pa	Choose simple shapes for patterns and cores (straight lines, rectangles).	C		
Pa	Aim at undivided patterns, if possible without cores (e.g. by means of open cross sections).	C		
Pa	Provide tapers from the split-line.	Q		
Pa	Arrange ribs so that pattern can be removed; avoid undercuts.	Q		
Pa	Ensure accurate location of cores.	Q		
Ca	Avoid vertical sections (bubbles, blowholes) and reduced cross-sections to the risers.	Q		
Ca	Aim at uniform wall thicknesses and cross-sections and gradual changes of cross-section; select material allowing adequate wall thicknesses and component sizes.	Q		
Ma	Set split-lines to avoid misalignment and to permit easy removal of the flash.	C Q		
Ma	Arrange castings to ease machining.	C Q		
Ma	Provide adequate support surfaces.	Q C		
Ma	Avoid sloping machining and boring surfaces.	C Q		
Ma	Combine machining processes by appropriate arrangement of machining and boring surfaces.	C		
Ma	Avoid unnecessary machining by breaking up large surfaces.	C		

Figure 7.110. Design guidelines with examples for cast components, after [7.123, 7.180, 7.198, 7.230, 7.331, 7.332]

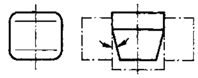
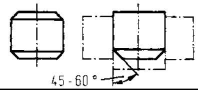
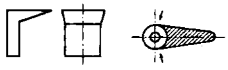
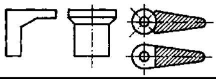
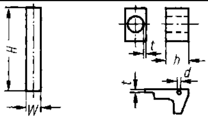
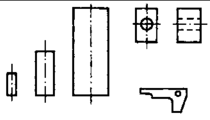
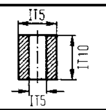
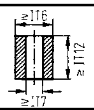
PS	Guidelines	Objectives	Wrong	Right
To	Avoid rounded edges and sharp angles.	C Q		
Si	Avoid sharp edges, sharp angles and tangential transitions.	Q		
Si	Observe dimensional limits and relations: Height H/Width $W < 2.5$ Wall thicknesses $t > 2$ mm Holes $d > 2$mm.	Q		
Si	Avoid small-toothed profiles.	Q		
Si	Avoid excessively small tolerances.	Q		

Figure 7.111. Design guidelines with examples for sintered components, after [7.106]

be used economically. Like all other cold forming methods, cold extrusion gives rise to work hardening, in which the yield strength is raised while the toughness of the material drops significantly. Designers must take this factor into consideration. The best materials for cold extrusion are case-hardening and heat-treatable steels.

For *drawing*, the following design guidelines are recommended in [7.230]:

- Allow for tooling (To): choose the dimensions in such a way that the smallest number of drawing steps possible are needed. Objective: reduction of costs.
- Allow for tooling and drawing (To/Dr): aim at rotationally symmetrical hollow bodies; producing the corners of rectangular hollow bodies leads to high loading of the materials and tools. Objectives: improvement of quality, reduction of costs.
- Allow for drawing (Dr): choose tough materials. Objective: improvement of quality.
- Allow for drawing (Dr): for the design of flanges see [7.201]. Objective: improvement of quality.

Bending (cold bending), as used for the production of sheet metal components in precision and electrical engineering as well as for casings, claddings and air ducts in general mechanical engineering, involves two separate steps: *cutting* (Cu) and *bending* (Be). Designers must therefore allow for both. The design guidelines shown in Figure 7.114 apply to the bending process alone; cutting is covered under the next heading.

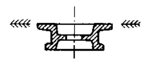
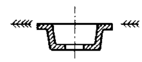
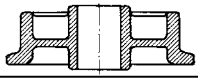
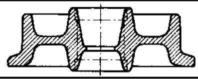
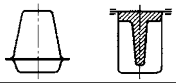
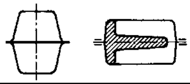
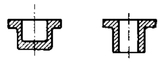
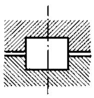
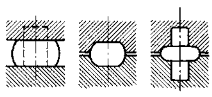
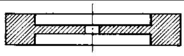
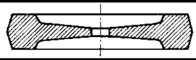
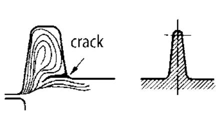
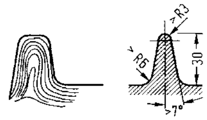
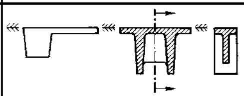
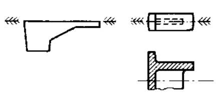
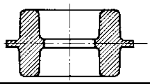
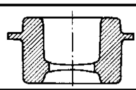
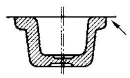
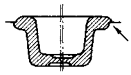
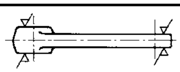
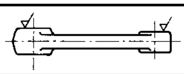
PS	Guidelines	Objectives	Wrong	Right
To	Avoid undercuts.	C		
To	Provide tapers.	C		
To	Aim for split lines at about half height perpendicular to smallest height.	C		
To	Avoid bent split lines.	C Q		
To Fo	Aim at simple, if possible rotationally symmetrical, parts. Avoid great protusions.	C		
Fo	Aim at shapes that occur during unrestrained pressing. For large numbers adapt to finishing shape.	C Q		
Fo	Avoid excessively thin sections.	Q		
Fo	Avoid large curvatures, excessively narrow ribs, fillets and excessively small holes.	Q		
Fo	Avoid sharp changes in cross-section and cross-sections that project excessively into the die.	Q		
Fo	Stagger split lines in the case of cup-shaped parts or large depth.	Q		
Ma	Select the split line so that misalignment is easily detected and removal of flash is simple.	C		
Ma	Arrange for surfaces to be machined to stand proud.	Q		

Figure 7.112. Design guidelines with examples for drop-forged parts, after [7.19, 7.145, 7.230, 7.238, 7.336]

Form Design for Separation

Of the separating procedures mentioned in DIN 8577 and 8580 [7.48, 7.49], we shall only consider "machining with geometrically defined cuts" (turning, boring, milling), "machining with geometrically undefined cuts" (grinding), and "separating" (cutting). In all separating processes, designers must allow for *tooling* (To), including clamping, as well as *machining* (Ma).

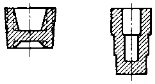
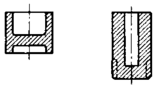
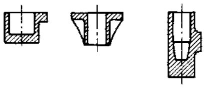
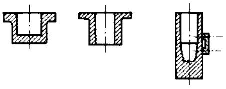
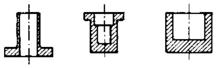
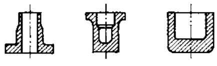
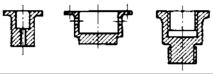
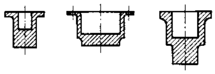
PS	Guidelines	Objec-tives	Wrong	Right
To Ex	Avoid undercuts.	Q C		
Ex	Avoid tapers and excessively small diameter diffrences.	Q		
Ex	Provide rotationally symmetrical parts without material protusions, otherwise split and join.	Q		
Ex	Avoid sharp changes in cross-section, sharp edges and fillets.	Q		
Ex	Avoid small, long or lateral holes and threads.	Q		

Figure 7.113. Design guidelines with examples for cold extrusions, after [7.108]

Design for tooling involves:

- The provision of adequate clamping facilities. *Objective*: improvement of quality.
- A preferential sequence of operations that does not necessitate the reclamping of components. *Objectives*: reduction of costs, improvement of quality.
- The provision of adequate tool clearances. *Objective*: improvement of quality.

Design for machining in all separating processes involves:

- The avoidance of unnecessary machining; that is, the reduction of machined areas, fine surface finishes and close tolerances to the absolute minimum (protruding bosses and cut-outs placed at the same height or depth are advantageous). *Objective*: reduction of costs.
- The location of machined surfaces parallel or perpendicular to the clamping surfaces. *Objectives*: reduction of costs, improvement of quality.
- The choice of turning and boring in preference to milling and shaping. *Objective*: reduction of costs.

Figure 7.115 represents the design guidelines for components machined by turning; Figure 7.116 shows them for components machined by boring; Figure 7.117 for components machined by milling; and Figure 7.118 for components machined by grinding.

In the design of *cut-out components*, the characteristics of the *tools* (To) and the *cutting method* (Cu) [7.19, 7.230] must be taken into consideration (see Figure 7.119).

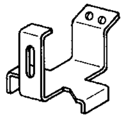
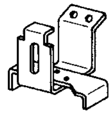
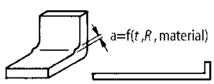
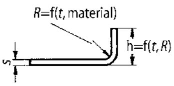
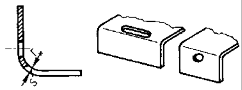
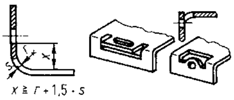
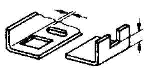
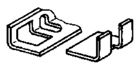
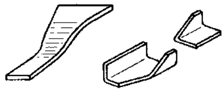
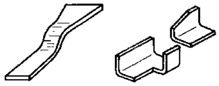
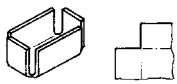
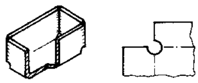
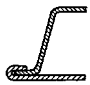
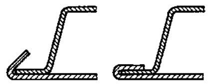
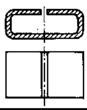
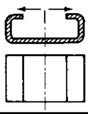
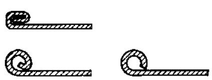
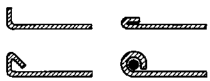
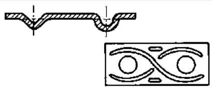
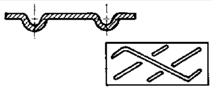
PS	Guidelines	Objec-tives	Wrong	Right
Be	Avoid complex bent parts (material waste); rather split and join.	C		
Be	Allow for minimum values of bending radii (bulging in the compression area and overstretching in tension area) flange height and tolerances.	Q		
Be	Provide sufficient distance between pre-pierced holes and bend.	Q		
Be	Aim at holes and notches to cross the bend when it is not possible to provide the minimum gap.	Q		
Be	Avoid sloping edges and tapers in the region of the bend.	Q		
Be	Provide clearances at the corners when all sides are to be bent up.	Q		
Be	Provide folded seam of sufficient width.	Q		
Be	Aim at large access openings for hollow shapes and undercut bends.	Q C		
Be	Provide stiffening at sheet edges.	A		
Be	Aim at identation forms.	A		

Figure 7.114. Design guidelines with examples for bent parts, after [7.1, 7.19]

Form Design for Joining

Of the joining methods discussed in DIN 8593 [7.50], we shall only consider welding under the above heading. For separable joints, the reader is referred to Section 7.5.9, "Design for Ease of Assembly".

Welding involves three process steps, namely *preparation* (Pr), *welding* (We) and *finishing* (Fi). The following design guidelines apply:

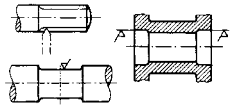
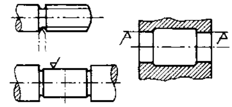
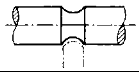
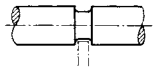
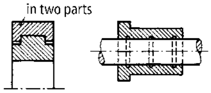
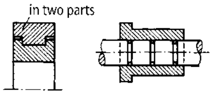
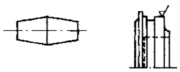
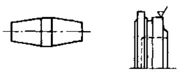
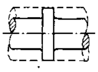
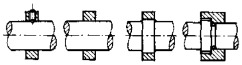
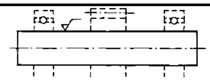
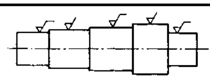
PS	Guidelines	Objectives	Wrong	Right
To	Provide adequate tool runout.	Q		
To	Aim for simple tool shapes.	C		
To	Avoid grooves and tight tolerances on inner surfaces.	C Q		
To	Provide for adequate clamping.	Q		
Ma	Avoid excessive machining, e.g. replace high collars by separate parts.	C		
Ma	Adapt working length and surface finish to the required function.	C		

Figure 7.115. Design guidelines with examples for components machined by turning, after [7.180, 7.230]

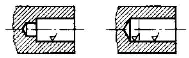
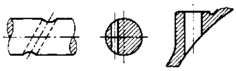
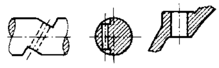
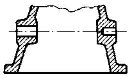
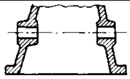
PS	Guidelines	Objectives	Wrong	Right
To Ma	Where possible, use boring tools on blind holes.	C Q		
To Ma	Provide starting and finishing flats for holes breaking through angled surfaces.	Q		
To	Aim for continuous holes, avoiding blind holes.	C		

Figure 7.116. Design guidelines with examples for components machined by boring, after [7.180, 7.198, 7.230]

- Pr, We, Fi: avoid the imitation of cast designs; preferably select standard, easily obtainable or prefabricated plates, sections or other semi-finished materials; make use of composite constructions (cast/forged components). *Objective*: reduction of costs.
- We: adapt the material, welding quality and welding sequence to the required strength, sealing and shape. *Objectives*: reduction of costs, improvement of quality.
- We: aim for short weld seams and small weld cross-sections to reduce damage through heating and to simplify handling. *Objectives*: improvement of quality, reduction of costs.

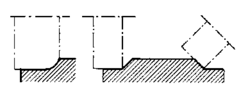
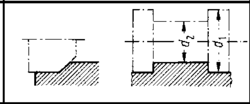
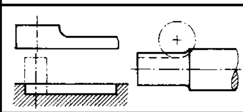
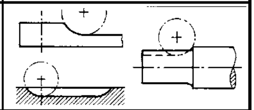
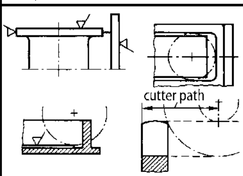
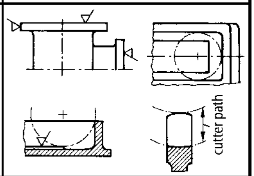
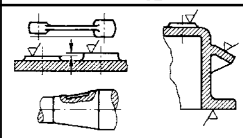
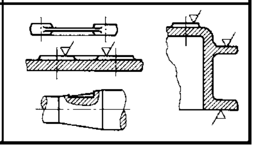
PS	Guidelines	Objec-tives	Wrong	Right
To	Aim for straight milling surfaces (form tools are expensive); select dimension for gang milling.	C		
To	Provide runouts for edge mills (edge milling is cheaper than end milling).	C Q		
To	Adapt runout to milling tool diameter. Avoid long milling cuts by selecting curved surfaces (e.g. slots).	C		
Ma	Arrange surfaces on one level and parallel to the clamping.	C Q		

Figure 7.117. Design guidelines with examples for components machined by milling, after [7.180, 7.230]

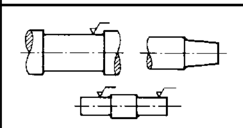
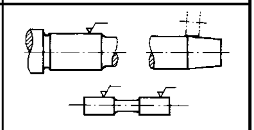
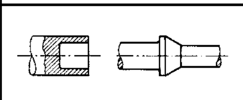
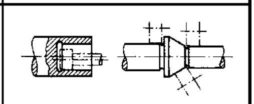
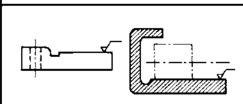
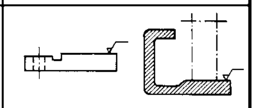
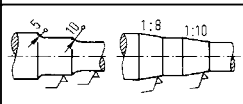
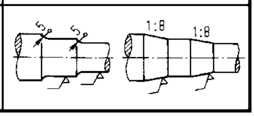
PS	Guidelines	Objec-tives	Wrong	Right
To	Avoid edge limitations.	Q C		
To	Provide runouts for grinding wheels.	Q		
To	Aim for unimpeded grinding by appropriate selection of surfaces.	C Q		
To Ma	Give preference to equal blend radii (if no runout possible) and to equal tapers.	C Q		

Figure 7.118. Design guidelines with examples for components machined by grinding, after [7.230]

- We,Fi: minimise the amount of welding (heat input) to avoid or reduce distortion and corrective work. *Objectives*: improvement of quality, reduction of costs.

Further guidelines are given in Figure 7.120.

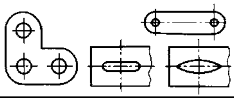
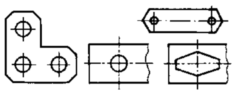
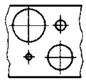
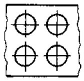
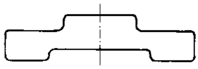
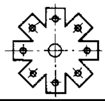
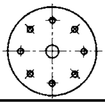
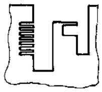
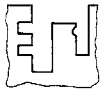
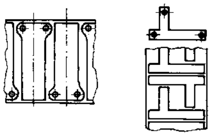
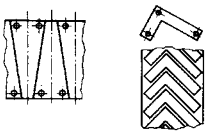
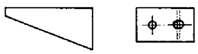
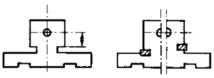
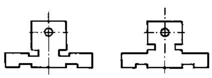
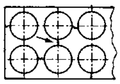
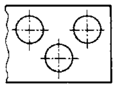
PS	Guidelines	Objectives	Wrong	Right
To	Aim for simple cuts, prefer angular corners, avoid curves.	C		
To	Aim for identical cut-out parts.	C		
To	Aim for sharp-edged transitions to facilitate the cutting of the template and to ensure easy grinding.	C Q		
To	Avoid complex contours.	C Q		
To	Avoid very narrow die cuts.	C Q		
Cu	Avoid waste by careful layout of cut-out parts on standard plate widths.	C		
Cu	Avoid sharp-angled shapes and excessively tight tolerances.	Q		
Cu	Prefer shapes permitting subsequent cuts without danger of damage.	Q		
Cu	Avoid very narrow spacing between holes.	Q		

Figure 7.119. Design guidelines for cut-out components, after [7.19, 7.230]

4. Appropriate Selection of Materials and of Semi-Finished Materials

An optimum choice of materials and semi-finished materials is difficult to make because of interactions between the characteristics of the function, working principle, layout and form design, safety, ergonomics, production, quality control, assembly, transport, operation, maintenance, costs, schedules and recycling. When the material costs of a proposed solution are particularly high, careful material selection becomes of the utmost economic importance (see Chapter 11). In general, designers are advised to consult the checklist (see Figure 7.3) and to evaluate the materials accordingly. The chosen materials and the resulting processing and

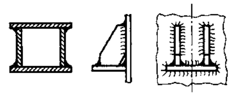
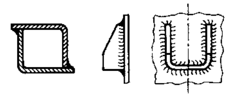
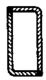
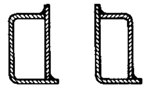
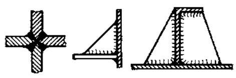
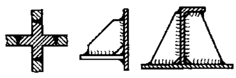
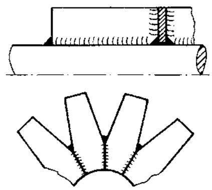
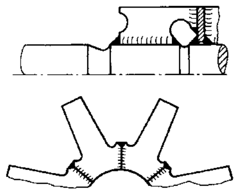
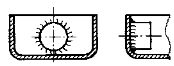
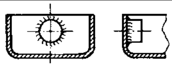
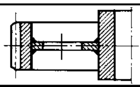
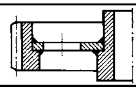
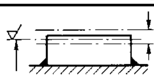
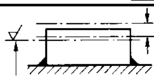
PS	Guidelines	Objectives	Wrong	Right
Pr	Prefer solutions with few parts and weld seams.	C		
Pr We Fi	Aim for easily weldable seams if loads permit.	C		
Pr We	Avoid build-up of weld material and intersecting weld seams.	C Q		
Pr We	Reduce residual stresses due to shrinkage by appropriate choice of weld seams and welding sequence, and of connecting sections of low stiffness (flexible tongues and corners).	Q		
We	Aim for good accessibility.	C Q		
We Fi	Ensure positive location of the components prior to welding.	Q		
Fi	Allow sufficient material for machining after welding.	Q	 Tolerance	 Tolerance

Figure 7.120. Design guidelines for welded components, after [7.19, 7.198, 7.220, 7.281]

machining of the components, their quality and the market conditions influence the selection of:

- production procedures
- types of machine, including tools and measuring instruments
- materials handling, for example, purchasing and storage
- quality control procedures
- in-house and subcontract production.

The close relationship between design, production procedures and materials technology calls for cooperation between designers, production engineers, materials experts and buyers.

The most important recommendations for the selection of materials for primary shaping processes (for example casting and sintering) and secondary shaping processes (for example forging, extrusion, etc.) have been set out by Illgner [7.137]. For production processes such as ultrasonic welding, electron-beam welding, laser technology, plasma cutting, spark erosion and electrochemical processes, see the following literature [7.27, 7.95, 7.133, 7.182, 7.240, 7.250, 7.262].

Closely connected with the selection of materials is the choice of *semi-finished materials* (for example tubes, standard extrusions, etc.). Because of the common method of costing by weight, designers tend to think that cost reduction invariably goes hand-in-hand with weight reduction. However, as Figure 7.121 makes clear, this belief is often mistaken.

The following example throws further light on this problem. Figure 7.122 shows a welded electric motor housing. The old layout required eight different plate thicknesses to achieve the required stiffness with minimum weight. In the mod-

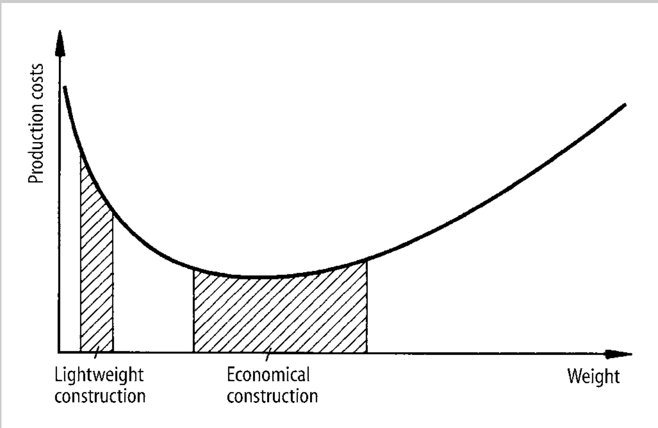


Figure 7.121. Cost areas for lightweight and economical constructions, after [7.297]

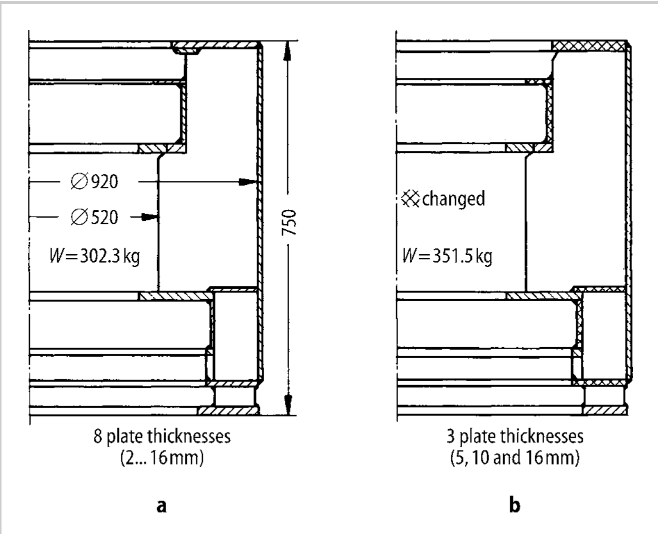


Figure 7.122. Electric motor housing of welded construction (Siemens): **a** current design; **b** proposed design

ified design, however, the number of plate thicknesses was deliberately reduced, although this increased the weight. This change in the design involved the replacement of standard flame cutting by numerically controlled machines. The extra outlay was to be justified by keeping the programming and re-equipping costs low, and through maximum utilisation of the plate material by stacking before cutting [7.5]. A cost analysis showed that, despite an increase in weight due to oversizing of some of the housing parts, the new design was cheaper than the old thanks to lower labour costs and lower production overheads. Admittedly, the actual saving was not very great, but this example serves to show that the minimisation of weight, which often involves a great deal of design and production effort, does not necessarily lead to minimum costs. Moreover, even when the calculated cost reductions resulting from the incorporation of semi-finished materials and simplification in production processes are not great, the actual savings may be much greater because of the consequent reduction in idle time and time spent on operations scheduling (see Chapter 11).

A further example of the economic use of semi-finished materials is given in Figure 7.123, which shows the plate-cutting plan for a welded motor housing. To allow the use of circular blanks for the end-wall bearing shields *d*, the end-walls are made from four parts *b* which are then welded together. The resulting aperture, even after machining, is smaller than the bearing shield made from the blank. In addition, this arrangement provides the support feet *c*.

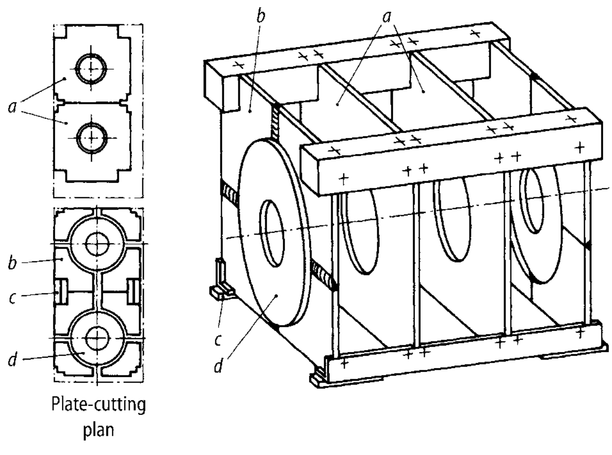


Figure 7.123. Electric motor housing. Welded construction with plate-cutting plan, after [7.162] (Siemens)

5. Appropriate Use of Standard and Bought-Out Components

Designers should always try to use components that do not have to be specially produced but that are readily available as *repeat*, *standard*, or *bought-out* parts. In this way, they can help to create favourable supply and storage conditions.

Easily available bought-out parts are often cheaper than parts made in-house. The importance of standard parts has already been stressed on several occasions.

The decision on whether components should be made in-house or bought-out depends on the following considerations:

- number (one-off, batch or mass production)
- whether production is for a specific order or for the general market
- market situation (costs, delivery dates of materials and bought-out parts)
- available production facilities
- utilisation of existing production facilities
- available or desired degree of automation.

These factors influence not only the decision on whether in-house production is to be preferred to subcontract production, but also the design approach. Unfortunately, most of the factors vary with time. This means that a particular decision may be justified at the time that it was made, but it may no longer be correct if the market situation and the production capacity change. Particularly in the case of one-off or batch products in the heavy engineering industry, the market and production situation needs to be re-examined at regular intervals.

6. Appropriate Documentation

The effect of production documents (in the form of CAD models, drawings, parts lists and assembly instructions) on costs, delivery dates, product quality, etc., is often underestimated. The layout, clarity and comprehensiveness of such documents have a particularly marked influence. They determine the execution of the order, production planning, production control and quality control.

7.5.9 Design for Assembly

1. Types of Assembly

Designers not only have a major influence on the costs (see Chapter 11) and the quality of the production of components, but also on the costs and quality of assembly [7.329].

By *assembly* we refer to the combination of components into a product and to the auxiliary work needed during and after production. The cost and quality of an assembly depend on the type and number of operations and on their execution. The type and number, in their turn, depend on the layout design of the product, on the form design of the components and on the type of production (one-off or batch production).

The following guidelines for design for assembly can therefore be no more than general hints [7.2, 7.32, 7.101, 7.102, 7.316, 7.318, 7.329]. The aims of the guidelines are to simplify, standardise, automate and ensure quality. In individual cases, they may be influenced or overridden by referring to the following headings in the checklist (see Figure 7.3): Function, Working Principle, Layout, Safety, Ergonomics, Production, Quality Control, Transport, Operation, Maintenance and Recycling. The particularities of specific cases must be checked [7.132, 7.170, 7.171, 7.223, 7.289].

According to Guideline VDI 3239 [7.309] and [7.3, 7.268], the following essential operations are involved:

- *Storing* (St) of parts to be assembled, if possible in a systematic manner. Automatic assembly further necessitates the programmed supply of parts and connecting elements.
- *Handling* (Ha) of components, including:
  - identifying the part by fitter or robot, e.g. by checking its orientation
  - picking up the part, if necessary in conjunction with individual selection and dispensing
  - moving the part to the assembly point, if necessary in conjunction with separation (removal, rejection, etc.), manipulation (rotation, inversion, etc.) and combining components.
- *Positioning* (Po) (placing the part correctly for assembly), and aligning (final adjustment of the position of the part before and possibly after joining).
- *Joining* (Jo) parts by the provision of appropriate connections. According to DIN 8593 [7.50], the following operations must also be included here:
  - bringing together, for example by inserting, superposing, suspending or folding
  - filling, for example by soaking
  - pressing together, for example by bolting, clamping or shrink-fitting
  - joining by primary processes, for example by fusing, casting and vulcanising
  - joining by secondary processes, for example by bending or via auxiliary components
  - joining by the combination of materials, for instance by welding, soldering or gluing.
- *Adjusting* (Ad) to equalise tolerances, to restore the required play, etc. [7.269].
- *Securing* (Se) the assembled parts against unwanted movements under operational loads.
- *Inspecting* (In). Depending on the degree of automation, various testing and measuring operations must be performed, possibly between individual assembly operations.

These operations are involved in every assembly process, their importance, sequence and frequency depending on the number of units (one-off assembly, batch assembly) and the degree of automation (manual, part automatic or fully automatic assembly).

According to [7.112], the linking of assembly operations or assembly cells can be divided into the following types: unbranched, branched, single level and multilevel assembly. The assembly process can be stationary or flowing.

It is also important to distinguish whether assembly takes place within the company or on site, by experts or by less well trained customer personnel. In general, the improvements one can make to automate assembly will also simplify manual assembly and vice versa. The selected type of assembly and the embodiment are closely related, that is, they influence one another.

2. General Guidelines for Assembly

In accordance with the steps of the embodiment design phase (see Section 7.1), it seems useful to start considering assembly even while working on the working structure and the layout. An easy-to-assemble layout can be achieved if the assembly operations are:

- structured
- reduced
- standardised
- simplified.

This will lead to a reduction in expenditure because the assembly process is improved and to increase in product quality because assembly is clearer and easier to control [7.94,7.105,7.257]. A layout that has been selected for these reasons could also lead a reduction in the number of components or at least the standardisation of components.

The embodiment guidelines that focus on ease of assembly are classified in Figure 7.124. The column *operation* contains the assembly operations that are primarily affected by the specific embodiment guidelines. The third column indicates whether the guideline leads to an improvement of *manual assembly* (MA) or *automated assembly* (AA), or both. This classification should ease the use and selection of embodiment guidelines for specific assembly situations.

3. Designing Assembly Interfaces

Another important aspect of improving assembly is the design of interfaces that are influenced by the layout. Improvements to the interfaces are achieved if they are:

- reduced
- standardised
- simplified.

These actions reduce the number of connecting elements and assembly operations, and minimise the quality requirements of the interfacing elements [7.2, 7.112, 7.273].

In Figure 7.125, the embodiment guidelines are again classified according to the aims and the affected assembly operations.

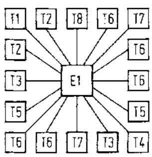
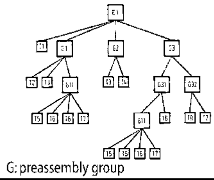
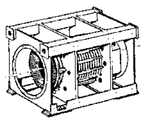
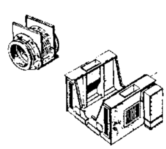
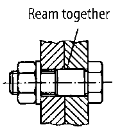
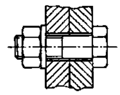
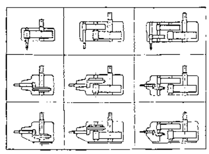
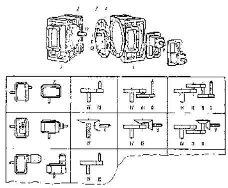
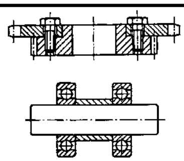
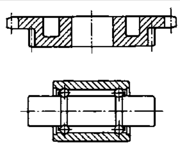
Oper.	Guidelines	Type	Wrong	Right
Arrange assembly operations				
St Ha Po Jo Ad Se In	Divide into assemblies to enable stepwise assembly with preassembly and final assembly.	MA AA		
Ha In	Arrange in independent assembly groups, e.g. to allow parallel assembly.	MA AA		
Jo	Avoid production operations during assembly.	MA AA		
Jo Ad In	Structure a variant product programme such that variants are created towards the end and at the same place in the assembly sequence.	AA		
In	Enable assembly groups to be inspected separately, especially for variant design.	MA AA	Balancing with a fully assembled machine.	Balancing the rotor on its own.
In	Aim at function testing of assembly groups or the whole product without testing individual parts.	MA AA	Measuring gear profiles on individual gears. Testing air-thigthness of components.	Measuring noise level of gearbox. Testing air-tightness of pipe system.
Reduce assembly operations				
St Ha Po Jo Ad Se In	Connect parts using integral and composite structures.	MA AA		

Figure 7.124. Embodiment guidelines for designing the layout for assembly

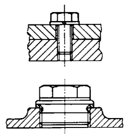
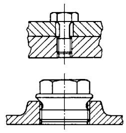
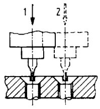
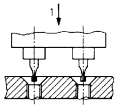
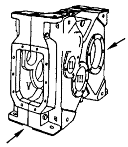
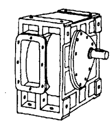
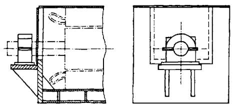
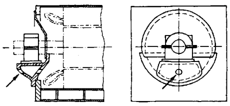
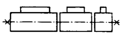
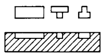
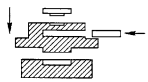
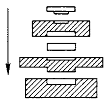
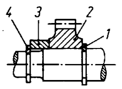
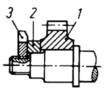
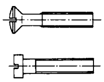
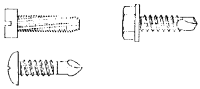
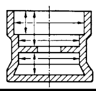
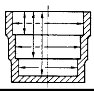
Oper.	Guidelines	Type	Wrong	Right
Reduce assembly operations				
St Ha Po Jo Ad Se In	Use function integration to reduce number of parts.	MA AA		
Jo	Execute assembly operations simultaneously.	AA		
Jo Ad Se	Reduce number of interfaces to be joined.	MA AA		
Ad Se In	Avoid disassembly to test functions of assembled groups and products.	MA AA	 Measuring of air gap not possible	 Measuring of air gap directly possible
Standardise assembly operations				
Po Jo In	Provide a basic component in every assembly group, e.g. to allow interlocking constructions.	AA		
Jo	Aim for uniform joining directions and procedures within an assembly group.	AA		
Simplify assembly operations				
Po Jo Ad Se In	Constrain assembly operations (clear assembly sequence).	MA	 or: *4 3 2 1*	
Jo	Combine production and assembly operations.	MA AA		
Ad In	Provide access for tests; enable visual inspection.	MA AA		

Figure 7.124. (continued)

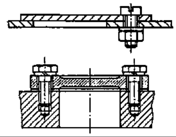
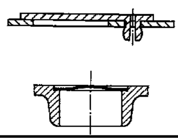
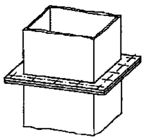
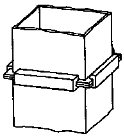
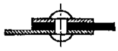
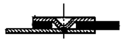
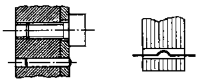
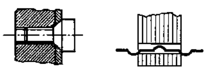
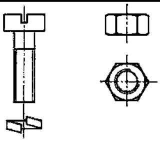
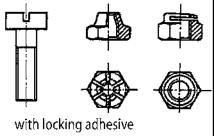
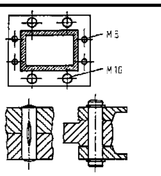
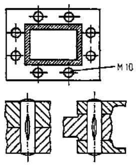
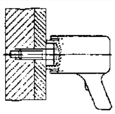
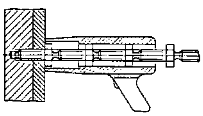
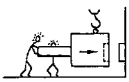
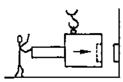
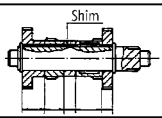
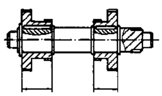
Oper.	Guidelines	Type	Wrong	Right
Reduce interfaces				
St Ha Jo Ad Se	Reduce connecting elements, e.g. by using clamp and snap connections.	MA AA		
St Ha Jo	Reduce connecting elements by using special connecting elements.	MA AA		
St Jo Se	Aim for direct connections without connecting elements.	MA AA		
Po	Aim for self-adjustment and positioning.	AA		
Se	Prefer self-locking connecting elements, e.g. through elastic-plastic deformation.	AA		
Standardise interfaces				
St Ha Jo	Use identical connecting elements, if possible even for different functions.	MA AA		
Simplify interfaces				
St Ha	Prefer connecting elements that can be delivered by belt or in a continuous flow.	AA		
Ha Jo		MA AA		
Po Jo	Avoid dimension chains with tight tolerances by splitting the dimension chain.	MA AA		

Figure 7.125. Embodiment guidelines for designing the interfaces for assembly

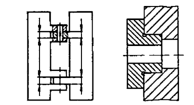
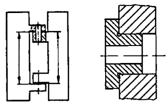
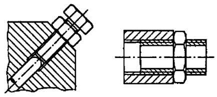
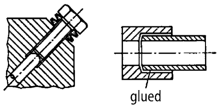
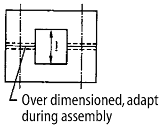
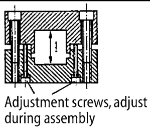
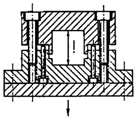
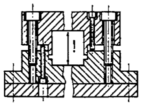
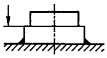
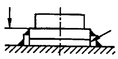
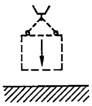
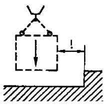
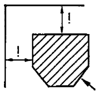
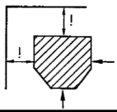
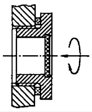
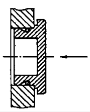
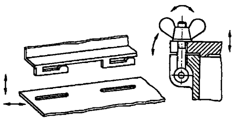
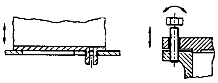
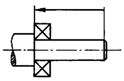
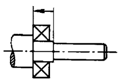
Oper.	Guidelines	Type	Wrong	Right
Simplify interfaces				
Po Jo	Avoid double fits to enable unambiguous positioning and to reduce tolerances on dimensions.	MA AA		
Po Ad	Prefer simple adjustments or provide positioning guides.	MA AA		
Po Ad	Enable continuous adjustment.	MA AA		
Po Ad	Aim for accessibility to allow adjustment without disassembling other parts.	MA AA		
Po Ad	Compensate tolerances by using compensation components.	MA		
Po Ad In	Provide reference surfaces, edges and points.	MA AA		
Po Ad In	Aim for unambiguous and independent adjustment operations.	MA AA		
Jo	Prefer translational joining motions.	AA		
Jo	Avoid joining motions involving multiple axes, in particular curves.	AA		
Jo	Avoid long joining paths.	MA AA		

Figure 7.125. (continued)

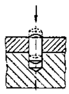
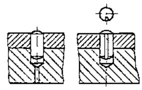
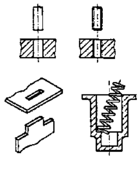
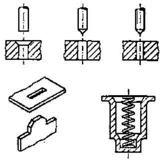
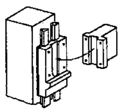
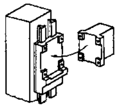
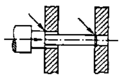
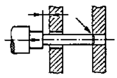
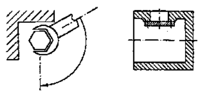
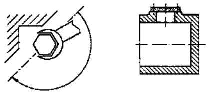
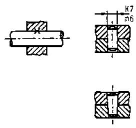
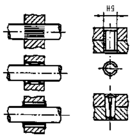
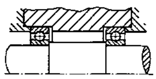
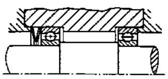
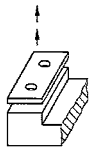
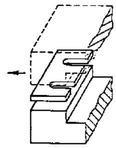
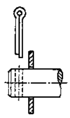
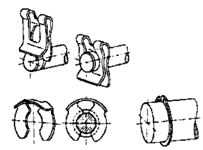
Oper.	Guidelines	Type	Wrong	Right
Simplify interfaces				
Jo	Avoid hindering caused by air pockets.	MA AA		
Jo	Provide tapering to ease joining.	MA AA		
Jo	Divide large interfaces into several smaller ones.	MA AA		
Jo Ad	Avoid simultaneous operations that influence each other.	MA AA		
Jo Ad	Provide access for assembly tools.	MA AA		
Jo Ad Se	Prefer connecting elements with elastic, elastic-plastic or material tolerance compensation.	MA AA		
Jo Se	Allow for large tolerances through assemply parts that are flexible.	MA AA		
Ad	Adapt using standardised matching parts without disassembling.	MA AA		
Se	Apply locking elements that are easy to assemble.	AA		

Figure 7.125. (continued)

4. Designing Interface Elements

Closely linked to the design of interfaces is the design of the interfacing elements. To improve automatic storage and handling, including the identification, ordering, picking-up and moving of the interfacing elements, these operations should be:

- enabled
- simplified.

This is particularly important for the application of automatic assembly machines (AA) [7.2, 7.103, 7.273, 7.289]. Figure 7.126 shows the design guidelines.

In summary, the essential guidelines can be derived from the basic guidelines of *simplicity* (simplify, standardise, reduce) and *clarity* (avoiding over constraining and under constraining) (see Sections 7.3.1 and 7.3.2). Further examples are given in [7.2, 7.104, 7.112, 7.114, 7.248, 7.249, 7.308].

5. Guidelines for Application and Selection

Design for assembly should, in line with the overall approach (see Section 7.1), involve the following five steps [7.112, 7.249] at appropriate stages of the design process.

Step 1: Draw-up demands and wishes for the requirements list that determine or influence assembly. This list will specify requirements such as:

- individually designed product or variant range
- number of variants
- safety and legal requirements
- production and assembly constraints
- test and quality requirements
- transport and packaging requirements
- assembly and disassembly requirements for maintenance and recycling
- requirements related to assembly operations undertaken by the user.

Step 2: Check for ways of easing assembly by using technical opportunities in the principle solution (working structure) and especially in the overall layout (construction structure), that is:

- Reduce the number of variants in a product range by using series and modular construction (see Chapter 9) or by concentrating on a few different types.
- Apply the embodiment guidelines shown in Figure 7.124 and use these to select layouts.

Step 3: Embody the assemblies, interfaces and interfacing elements that determine the assembly process, that is:

- Apply the embodiment guidelines shown in Figures 7.125 and 7.126 and use these to select embodiment variants.

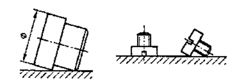
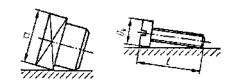
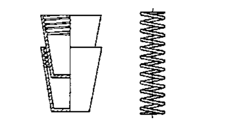
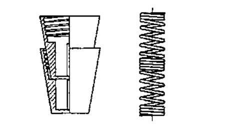
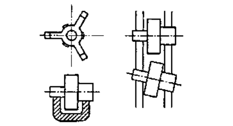
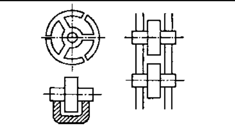
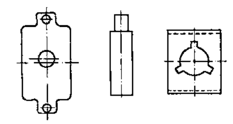
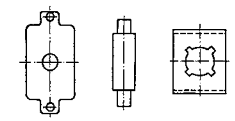
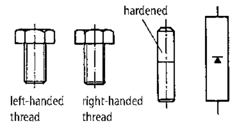
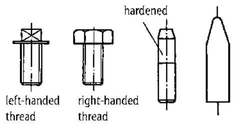
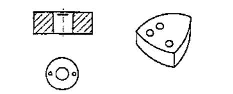
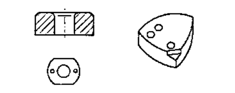
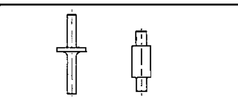
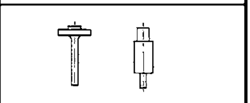
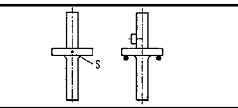
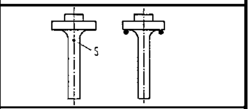
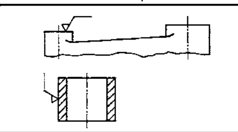
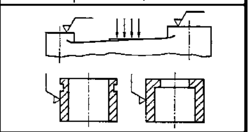
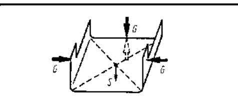
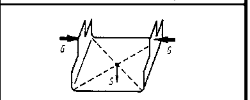
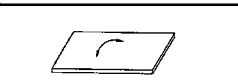
Oper.	Guidelines	Type	Wrong	Right
Enable and simplify automatic storage and handling				
St	Prefer interface elements that have a stable position.	AA		
St	Avoid identical interface elements that can interlock.	MA AA		
St Ha	Aim for interface elements that can roll.	AA		
Ha	Aim for symmetric contours when a specific position is not required.	AA		
Ha	Aim for geometric identifiers.	AA		
Ha	Prefer identifiers on the outer contour.	AA		
Ha	Avoid near symmetry when a specific position is required.	AA		
Ha	Ease handling by using interface elements that can be suspended and prefer a position based on the centre of gravity.	AA		
Ha	Provide features and surfaces outside functional surfaces to aid handling.	MA AA		
Ha	Position handling surfaces based on the centre of gravity.	MA AA		
Ha	Aim for interface elements with stable geometry.	MA AA		

Figure 7.126. Embodiment guidelines for designing interface elements for assembly

- Take into account special production and assembly restrictions (batch size; available machine tools; manual, semi-automatic or automatic assembly).
- Select connecting elements and processes not only based on functional requirements (strength, sealing, and corrosion resistance) but also based on requirements of assembly and disassembly (ease of loosening during disassembly, reuse, potential for automation).
- Consider production and assembly costs together.

Step 4: Evaluate embodiment variants technically and economically, paying particular attention to the required interfacing procedures, that is:

- Evaluate the ease of assembly of a design, preferably as soon as the principle solution is established. Designers should work together with the production planning department, because the assembly plan (assembly sequence and structure [7.112]) and the assembly processes and tools, including quality control, cannot be determined by designers alone. One way to aid the development of an assembly plan is to mentally divide the overall layout drawing into its individual elements; that is, to start by drawing up a disassembly plan. The inverse of this can then provide the basis for the assembly plan of the product. It can also be useful to simulate the assembly process using computer-supported production and assembly planning (CAP) and the production of prototypes.
- Assess the assembly process in terms of the supply of subcontract, bought-out and standard parts.
- Derive evaluation criteria from the goals and embodiment guidelines listed in Figures 7.124 to 7.126, adapting them where necessary to the particular situation.

Step 5: Prepare detailed assembly instructions together with the production documents. This includes overall layout drawings for subassemblies and the product (preassembly and final assembly), the assembly parts list and other assembly information.

7.5.10 Design for Maintenance

1. Goals and Terminology

Technical systems and products are subject to wear and tear, reduction of useful life, corrosion, contamination and changes in time-dependent material properties, such as embrittlement. After a certain period of time, whether in use or not, the actual condition of a system will no longer be the intended one. Deviations from the intended condition cannot always be recognised directly and can cause changes in performance, failures and dangerous situations. This can substantially reduce the functionality, economy and safety of a technical system. Sudden breakdowns disrupt normal operation, and because they are unexpected they require considerable cost to rectify. Not checking the condition of a system until damage has occurred, possibly involving injury, is unacceptable from both human and economic points of view.

Because systems and products have become more complex, the application of maintenance as a *preventative* measure has become increasingly important. Designers have a significant influence on maintenance costs and procedures through their selection of the principle solution and embodiment features, which according to [7.62, 7.304] strongly determine maintainability. We have already emphasised the importance of maintenance in our systematic approach; for example in the guidelines (see Sections 2.1.7 and 5.2.3, Figure 5.3, Section 6.5.2, and Figure 6.22) and their application in connection with the basic rules (see Section 7.3). More recent publications [7.139,7.151] emphasise the importance of an early consideration of maintainability and a systematic approach.

Maintenance is related to safety (see Section 7.3.3), ergonomics (see Section 7.5.6) and assembly (see Section 7.5.9). As the sections in this book addressing these topics already include suggestions and rules relevant to maintenance, this section focuses on what is necessary for a general understanding of maintenance and on design for ease of maintenance.

According to DIN 31051 [7.62], *maintenance* involves monitoring and assessing the actual condition of a system and maintaining or recovering the intended condition.

Possible measures are:

- *Service*, to maintain the intended condition
- *Inspection*, to monitor and assess the actual condition
- *Repair*, to recover the intended condition.

The type, extent and duration of service and inspection measures obviously depend on the type of system, its intended function, its required availability, its desired reliability, and on any potential dangers. The selected measures determine whether inspection and service has to take place after a fixed period of time, after a specific number of operating hours, or after a particular intensity of load.

The *maintenance strategy* is also influenced by the rate of deterioration of components, for example through wear that reduces operating life. The measures applied to recover the intended condition must be taken before components are predicted to fail. Accordingly, two types of repair are distinguished:

- *Failure repair* that takes place after a component has failed. This strategy is applied, and is often the only possibility, when failures cannot be predicted accurately. It is important that such failures do not cause danger. The disadvantage of this approach is the effect it has on planning. An example is the shattering of a car windscreen. This strategy is not suitable for production plant, and in situations where a function must be fulfilled or where danger is involved.
- *Preventive repair* that takes place before a component has failed. This can be determined by either *interval* or *condition*. Interval repair takes place after a fixed period of time, a specific distance or a set number of operations. An example is when the oil in an vehicle engine is changed after 10 000 km. Condition repair is based on actual performance measures, such as the loads or temperatures experienced in operation. When an unwanted condition is observed, the service

or repair measures must be carried out. An example of this is when the oil in a vehicle engine is changed after a certain number of cold starts, or the integrated average temperature of the oil reaches a certain value. Another example is when brake linings are replaced after a measured amount of wear. Whether the interval or condition strategy is applied depends on the operating conditions. A combination of the two strategies is also possible. A power station, for example, will use the time interval repair strategy to safeguard the base load. For components that can last several intervals, the condition repair strategy will be adopted.

More detailed discussions of maintenance strategies can be found in [7.282, 7.304]. Predictions of both the probability of failure and component reliability are discussed in [7.232].

2. Design for Maintenance

Maintenance requirements should have been included in the requirements list, see Figure 5.3 and VDI 2246, Part 2 [7.304]. When solutions have to be selected, easily maintained variants should be preferred. Examples are variants that require minimal servicing, include components that can be exchanged easily, and use components with similar life expectancies. During the embodiment phase, it is important to consider accessibility and ease of assembly and disassembly. However, design for maintenance should never compromise safety.

According to [7.282], a technical solution should, in principle, require as few preventative measures as possible. The aim is complete freedom from the need for service by using components with identical lives, reliability and safety. The chosen solution should thus incorporate features that make maintenance unnecessary or reduce it substantially.

Only when such features cannot be realised or are too costly should service and inspection measures be introduced. In principle, the following aims are important:

- Prevent damage and increase reliability.
- Avoid the possibility of errors during disassembly, reassembly and start-up.
- Simplify service procedures.
- Make the results of servicing checkable.
- Simplify inspection procedures.

Service measures usually concentrate on refilling, lubricating, conserving and cleaning. These activities should be supported by embodiment features and appropriate labelling based on ergonomic, physiological and psychological principles. Examples are easy access, nontiring procedures and clear instructions.

Inspection measures can be reduced to a minimum when the technical solution itself embodies direct safety techniques, see Section 7.3.3, and thus promises high reliability. Overloading, for example, can be avoided by using appropriate principles such as self-help that provide protection against failures and disturbing influences, see Section 7.4.3. When service and inspection measures cannot be avoided, embodiment guidelines, discussed earlier, should be applied [7.282]. In what follows, we limit ourselves to lists and short explanations.

Technical measures that can reduce the service and inspection effort, and should have been considered already in the conceptual phase, include:

- Prefer self-balancing and self-adjusting solutions.
- Aim at simplicity and few parts.
- Use standard components.
- Allow easy access.
- Provide for easy disassembly.
- Apply modular principles.
- Use few and similar service and inspection tools.

Service, inspection and repair instruction documents have to be prepared, and service and inspection points have to be labelled clearly. Guidance on developing maintenance manuals can be found in DIN 31052 [7.63], and guidance on determining maintenance intervals in DIN 31054 [7.64].

To facilitate the execution of service, inspection and repair measures, the following ergonomic rules, supported by appropriate technical embodiments, should be applied:

- Service, inspection and repair locations should be easily accessible.
- The working environment should follow safety and ergonomic requirements.
- Visibility should be ensured.
- Functional processes and supporting measures should be clear.
- Damage localisation should be possible.
- Exchange of components should be easy.

Instructive examples for each of these requirements can be found in [7.282].

Finally, maintenance should be part of the total concept. Maintenance procedures must be compatible with functional and operational constraints of the technical system, and must be included in the overall cost along with the purchase and operating costs.

7.5.11 Design for Recycling

1. Aims and Terminology

To save and reuse raw materials in order to move towards more sustainable development, the following possibilities can be considered [7.141, 7.142, 7.169, 7.186, 7.197, 7.218, 7.302, 7.305, 7.321]:

- *reducing material use* through better utilisation (see Section 7.4.1) and by reducing waste during production (see Section 7.5.8)
- *substituting materials* for those becoming rare and expensive [7.9]
- *recycling materials* by reusing or reprocessing production waste, products and parts of products.

In what follows, possible types of recycling and recycling processes are explained based on VDI Guideline 2243 [7.302]. They help us to understand the embodiment guidelines that support recycling (see Figure 7.127).

Production waste recycling involves reusing production waste in a new production process, for example offcuts (after they have been preprocessed).

Product recycling involves reusing a product or part of it, for example reusing a vehicle's engine (after it has been reconditioned).

Used material recycling is the reuse of old products and materials in a new production process, for example the reprocessing of materials from scrapped vehicles (after they have been preprocessed). These secondary materials or parts do not necessarily have a lower quality than new materials or parts, in which case they can be reused. When the quality is significantly reduced, they can only be used for other purposes.

Preprocessing and reconditioning make significant contributions to effective recycling.

The materials left over from the recycling system end up in waste dumps or in the environment. It is possible that in the future these materials will also be used as resources.

Various methods of recycling are possible within the recycling loops shown in Figure 7.127. Basically one can distinguish between *reusing* products and *reprocessing* products.

Reuse involves retaining the product shape whenever possible. This type of recycling represents a high level of utilisation and should therefore be aimed for. Two types of reuse can be distinguished. In the first, the product fulfils the same function, e.g. refillable gas cylinders, and in the second a different function, e.g. reusing car tyres as boat fenders.

Reprocessing destroys the product shape, and so this process leads to a lower utilisation value. Two types of reprocessing can be distinguished. In the first, reprocessing takes place for application in the same product production process, e.g. reprocessing the materials from scrapped vehicles; in the second, reprocessing takes place for a different application, e.g. converting old plastic into oil by pyrolysis.

2. Recycling Processes

Preprocessing

The reprocessing of production waste and scrap materials is influenced strongly by the necessary preprocessing methods [7.186, 7.197, 7.277, 7.302].

Compacting of loose scrap by *pressing* eases the process of charging in metal making, but does not allow the separation of materials in mixed scrap. It is therefore only suitable for the recycling of unmixed production waste and scrap metals, e.g. food cans.

Cutting heavy or large products can be done with *shears* or *flame cutting*. These methods are particularly suitable when the materials have to be separated afterwards.

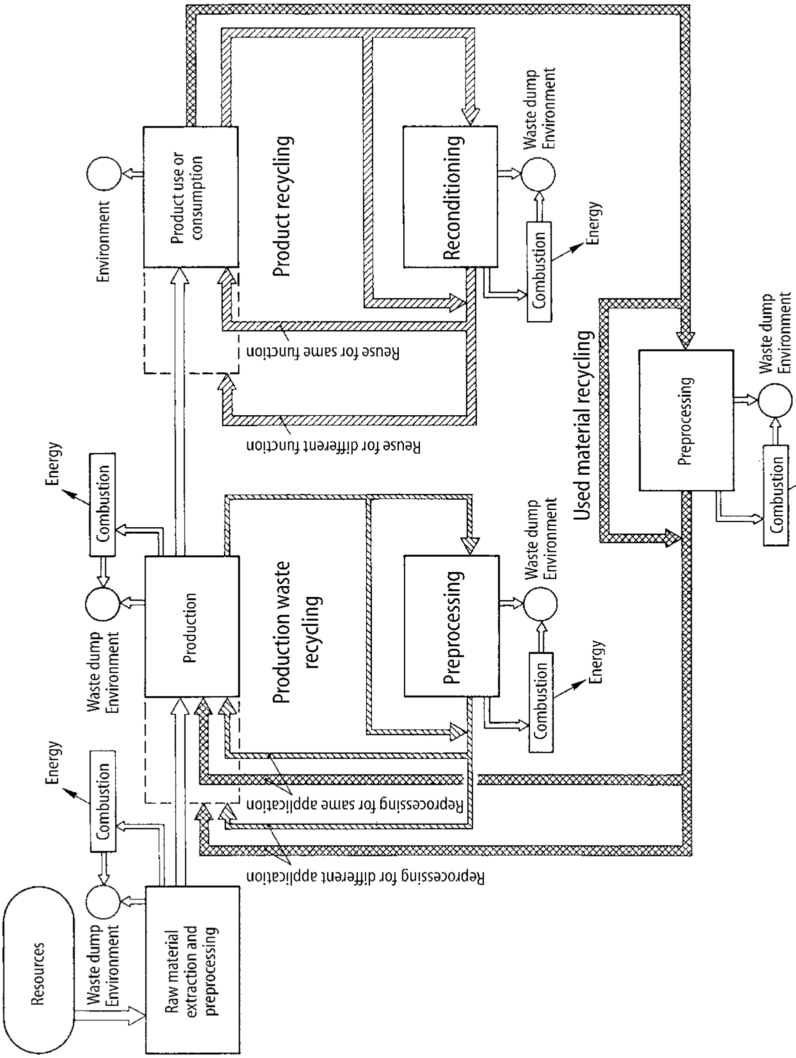


Figure 7.127. Possibilities for recycling, after [7.186, 7.302]

Separating can take place in a *shredding plant* based on the principle of a hammer pulveriser, in which a rotating hammer tears the product apart. In series with this pulveriser are other processes, such as dust removal, magnetic separation, size separation, and manual sorting of materials. Shredded scrap has high quality because of its high density, purity and uniform piece size. These technically complex and labour-intensive preprocessing methods are used for about 80% of scrapped vehicles and about 20% of scrapped domestic products, e.g. refrigerators. *Grinders* provide the same waste quality. They are just as technically complex, differing only in the method of pulverising used prior to material separation.

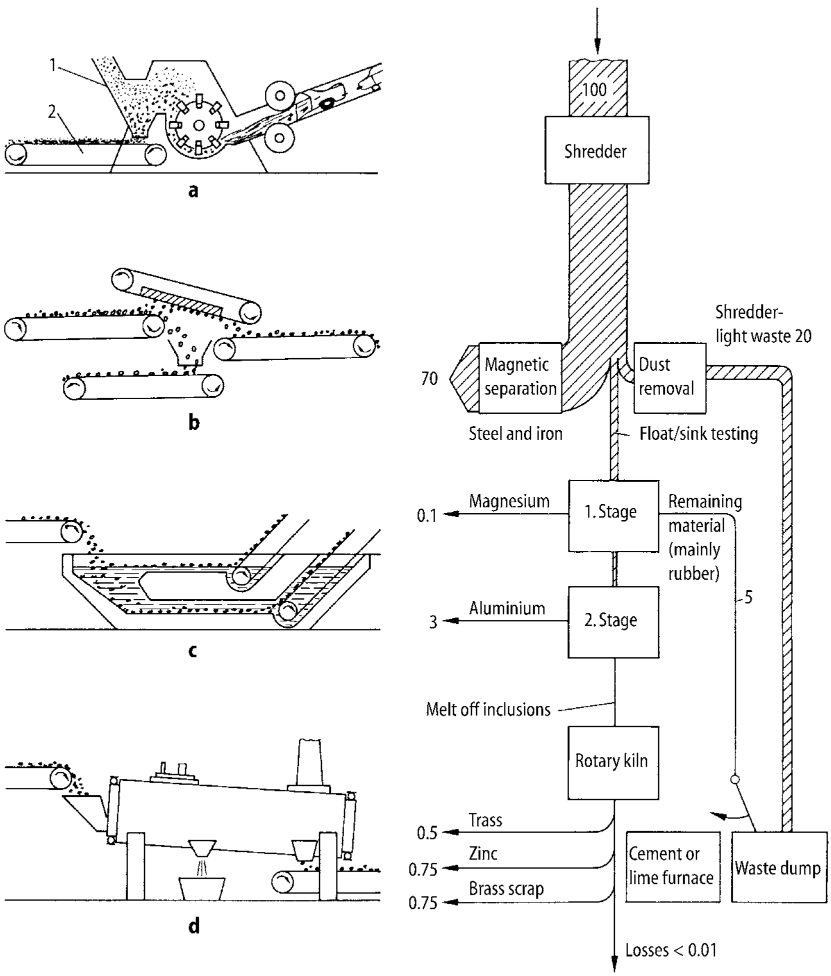


Figure 7.128. Operating principles and material flow in a shredding pant: **a** shredder, *1* dust removal, *2* sorting conveyors; **b** magnetic separation; **c** float/sink testing; **d** rotary kiln

Float/sink testing can be linked to shredders and grinders for improved separation of nonferrous and nonmetallic parts. *Dropping weights* can be used to reduce large grey iron castings with large wall thicknesses. *Chemical preprocessing* can be used to separate harmful materials and alloys before they are used again in metal making.

Figure 7.128 illustrates the material flow in a modern shredding plant [7.302].

Because *plastics* now make up a large proportion of scrap, it is becoming increasingly important to recycle these materials [7.18, 7.109]. The preprocessing of thermoplastics can be achieved through shredding, washing, drying and granulating, provided this waste has been presorted. This is difficult for household waste. The preprocessing of mixed plastic waste can be performed by mechanical separation, such as sorting, sizing and sieving, after it has been broken down into smaller parts. Other methods of separating include the use of electrostatics and floatation for density testing. Such preprocessing methods are still under development, so the sorting of plastics prior to collection would provide an economically viable alternative. Chemical preprocessing can be used for thermosetting plastics and elastomers [7.184].

The best waste and scrap quality—that is, the highest material reutilisation rate—is achieved by *disassembling* the product prior to preprocessing. Such disassembly into appropriate material groups can be undertaken by either specialist companies or by the product manufacturers themselves on dedicated disassembly lines.

The prerequisites for economic disassembly should be established by designers through the selected *embodiment features* and *assembly methods* (see Section 7.5.9). Economic preprocessing of scrap products and materials involves an appropriate combination of disassembly and preprocessing methods [7.186, 7.302].

Reconditioning

In order to be able to reuse products after they have been used for the first time, a reconditioning process that comprises the following steps is required [7.197, 7.266, 7.267, 7.302, 7.319]:

- complete disassembly
- cleaning
- testing
- reuse of worthwhile parts, repair of worn areas, reworking of parts to be adapted, replacement of unusable parts with new ones
- reassembly
- testing.

Two methods are used to recondition products, whether this is undertaken in special companies or by the product manufacturers [7.10]. The first method retains the identity of the original product; that is, while changing and reworking parts the configuration of the parts is retained and the tolerances are matched to each other.

For example, an engine reconditioned using this method will retain its original engine number. The second method breaks up the original product in such a way that all parts are treated as new ones along with their individual tolerances. The result is that the reconditioned original parts and the new parts are combined at the reassembly stage as if they were all new. This method has a promising future because the same production and assembly facilities can be used for both the reconditioned and original products.

3. Design for Recycling

To support preprocessing and reconditioning procedures, designers can introduce specific measures during product development [7.12–7.14, 7.22, 7.141, 7.142, 7.186, 7.187, 7.196, 7.302, 7.320, 7.321, 7.323]. These measures, however, must not conflict with the other goals and requirements of the task (see Figure 2.15). In particular, the cost effectiveness of production and operation must be guaranteed.

Recycling Considerations During the Design Process

Recycling possibilities should be considered during all stages of the design process (see Figures 1.9, 6.1, 7.1, 8.1). Figure 7.129 shows which recycling related design tasks should be undertaken in each of the design phases set out in VDI Guideline 2221 [7.270, 7.323].

Embodiment Guidelines for Preprocessing

The following guidelines relate to the overall product and the individual assemblies. They can be applied singly or in combination, with the aim of improving preprocessing or direct reprocessing.

Material compatibility. It is very difficult to design products made from a single material that can be reprocessed easily. For indivisible units, therefore, the aim should be to use materials that are compatible with regard to reprocessing. This results in an output from the process that is more economical and has higher quality.

To fulfil this aim, the production requirements for reprocessing need to be known. Here it is useful to define so-called scrap material groups or base materials to which compatible materials are assigned. Until such generally applicable groups are identified by materials scientists and the materials processing industry, designers should check the material compatibility in each case with experts. This is particularly important for large batch production with high recycling potential. Figure 7.130 shows a sample compatibility table for plastics.

Material separation. When material compatibility cannot be realised for inseparable parts or assemblies, additional interfaces should be introduced to break products down in such a way that the incompatible materials can be separated during preprocessing, for example through disassembly.

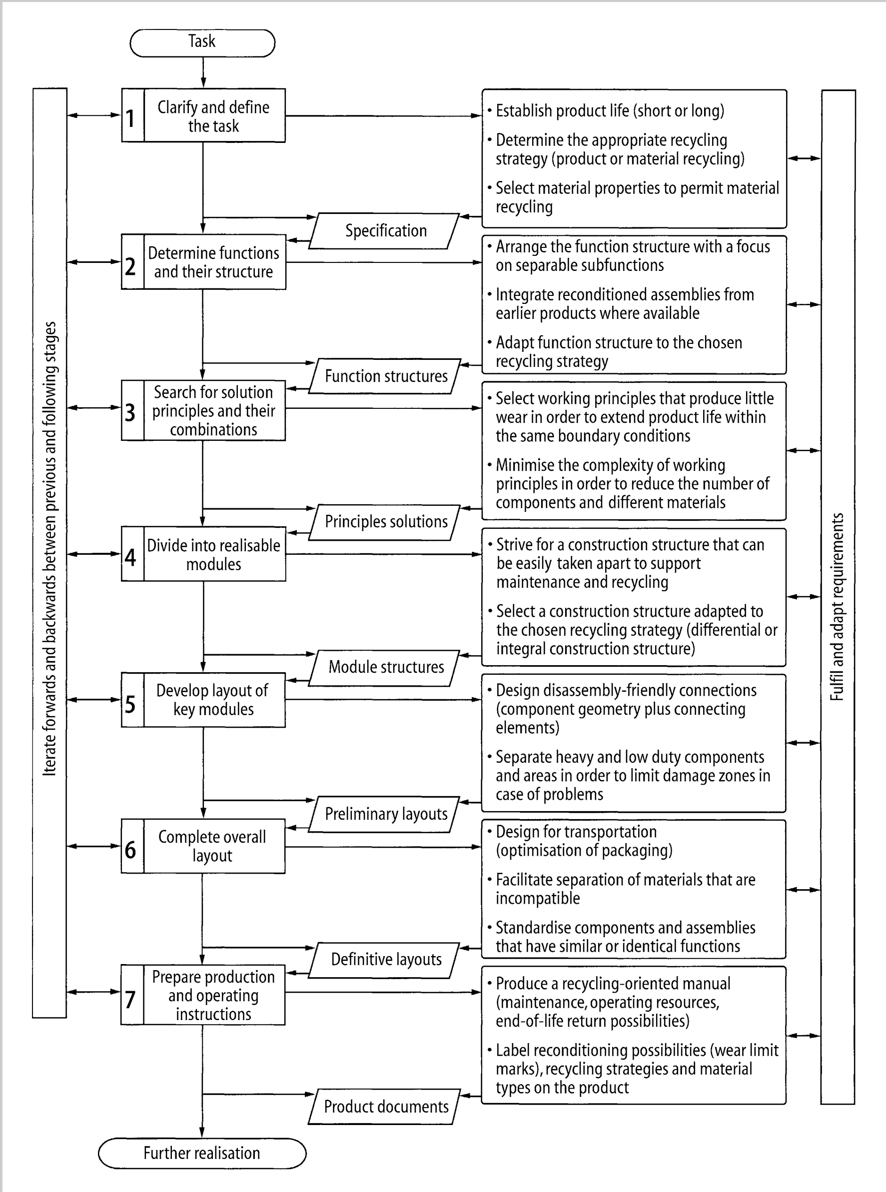


Figure 7.129. Recycling-related tasks allocated to the phases of the design process in VDI 2221 [7.270, 7.293, 7.323]

Interfaces suitable for preprocessing. Interfaces that support high-quality and economic preprocessing should be easily accessed and disassembled, and located near the outer edges of the product. Figure 7.131 shows types of connections that can be easily disassembled. Composite constructions usually require a higher recycling effort [7.119] and should, where possible, be avoided.

	Important synthetic design materials	Additive												
		PE	PVC	PS	PC	PP	PA	POM	SAN	ABS	PBTP	PETP	PMMA	
Basic material	PE	●	○	○	○	●	○	○	○	○	○	○	○	
	PVC	○	●	○	○	○	○	○	●	◐	○	○	●	
	PS	○	○	●	○	○	○	○	○	○	○	○	○	● Compatible
	PC	○	◔	○	●	○	○	○	●	●	●	●	●	◐ Limited compatibility
	PP	◔	○	○	○	●	○	○	○	○	○	○	○	
	PA	○	○	◔	○	○	●	○	○	○	◔	◔	○	◔ Compatible in small quantities
	POM	○	○	○	○	○	○	●	○	○	◔	○	○	
	SAN	○	●	○	●	○	○	○	●	●	○	○	●	
	ABS	○	◐	○	●	○	○	◔	○	●	◔	◔	●	○ Not compatible
	PBTP	○	○	○	●	○	◔	○	○	◔	●	○	○	
	PETP	○	○	◔	●	○	◔	○	○	◔	○	●	○	
	PMMA	○	●	◔	●	○	○	◔	●	●	○	○	●	

Figure 7.130. Compatibility table for plastics [7.146, 7.302]

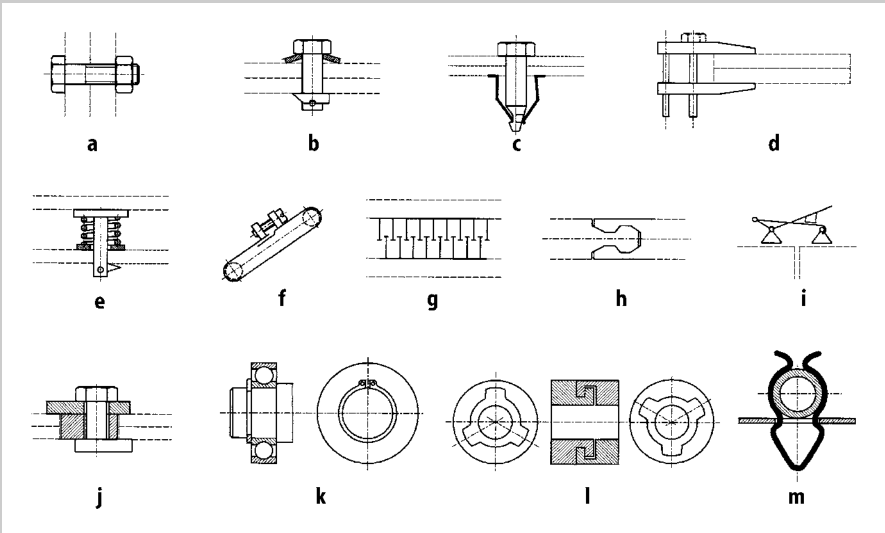


Figure 7.131. Disassembly-friendly connections [7.197, 7.244]. **a** Bolt, **b** quarter-turn fastener, **c** push-turn fastener, **d** clamp, **e** push-push fastener, **f** jubilee clip, **g** velcro, **h** form-fit fastener, **i** lever clamp, **j** eccentric fastener, **k** circlip, **l** bayonet, **m** spring clip

Economical disassembly. Simple tools, automatic processes and untrained personnel are preferred, in particular for disassembly at scrap yards.

High value materials. Valuable and rare materials should be positioned favourably and labelled to facilitate separation.

Dangerous materials. Materials, liquids and gases that can be dangerous to humans and the environment during preprocessing or direct reprocessing should always be easy to separate or remove.

Embodiment Guidelines for Reconditioning

The following guidelines should be applied:

- Ensure easy and damage-free disassembly (see [7.134,7.160,7.194,7.270] for further disassembly guidelines and Figure 7.132 for concepts that ease disassembly—also compare with Section 7.5.9).
- Ensure that all reusable parts can be cleaned easily and without damage.
- Facilitate testing and sorting through appropriate embodiment.
- Ease the reworking of parts or the deposition of material by providing additional material and facilities for locating, clamping and measuring.
- Ease reassembly by using existing tools from one-off and small batch production.

To reduce the number of new parts that are needed, the following measures are useful:

- Limit wear to special-purpose, easily adjustable or exchangeable parts (see Sections 7.4.2 and 7.5.5).
- Make it easy to identify the state of wear of a part and to decide whether it can be reused.
- Ease material deposition on areas of wear by selecting appropriate base materials.
- Minimise corrosion through embodiment and protective measures to increase the reusability of parts (see Section 7.5.4).
- Select connections that function throughout the product life yet can be easily undone, do not slacken after repeated disconnecting, and are not subject to corrosion bonding [7.245, 7.286].

Labelling of Recycling Possibilities

The recycling possibilities and the required recycling procedures for assemblies and modules should be labelled in line with the proposed recycling strategy and the embodiment developed to fulfil that strategy. This allows easy and safe adoption of the required recycling processes and measures. Figure 7.133 provides an example of the labelling of plastic parts.

4. Examples of Design for Recycling

Recycling of Plain Pedestal Bearings (Used Material Recycling)

Plain pedestal bearings (see Figure 9.25) are so common in machines that it is economic to consider recycling. The first possibility is to recycle by reconditioning

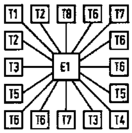
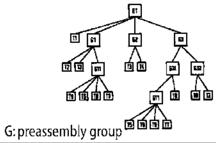
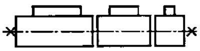
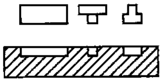
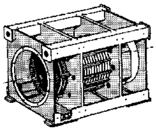
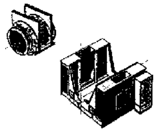
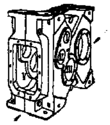
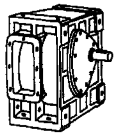
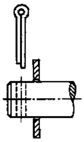
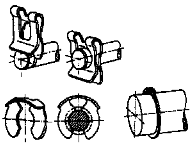
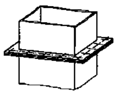
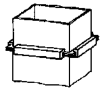
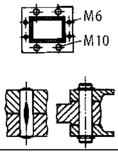
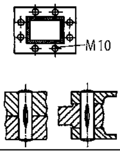
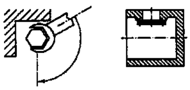
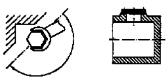
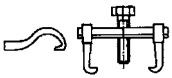
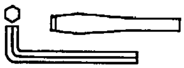
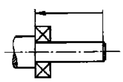
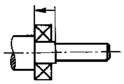
Guidelines	Wrong	Right
Disassembly-friendly construction structures		
Arrange in assemblies consisting of components and materials that are compatible regarding reprocessing.		
Assign the base component of an assembly to a scrap material group suitable for reprocessing.		
Avoid composite construction structures with materials that are incompatible regarding reprocessing where they cannot be separated.		
Reduce the number of interfaces.		
Disassembly-friendly interfaces		
Use connecting and locking elements that can easily be disassembled or destroyed, even after a long operating life.		
Reduce the number of connecting elements.		
Use identical connecting elements.		
Ensure good accessibility for disassembly tools.		
Use simple standard tools.		
Avoid long disassembly paths.		

Figure 7.132. Embodiment guidelines for ease of disassembly [7.197, 7.244]

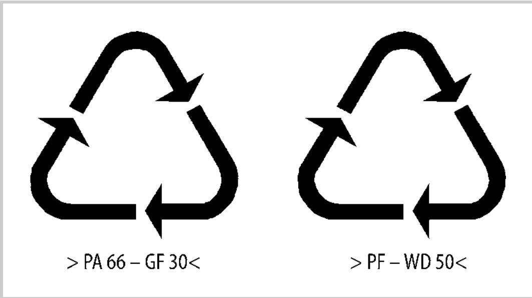


Figure 7.133. Example of labels used for plastic components according to DIN ISO 11469, DIN 7728 T.1 and DIN ISO 1043

the worn-out parts; that is, to provide new or renewed cast bearing shells, lubrication rings and seals. The second is to exchange the bearing completely. Up until now about 99% of used pedestal bearings have been reprocessed as whole bearings (used material recycling, see Figure 7.127), resulting in low reprocessing quality. The reprocessing quality is determined by the purity of the material after the product has been preprocessed. This quality depends on the material combination in the product and the preprocessing technology used. Commonly available plain pedestal bearings, for example, consist of about 74% cast iron, 22.3% unalloyed steel, 3.5% nonferrous metals and 0.2% nonmetals. The weight percentage of the elements in a bearing, similar to the one in Figure 9.25, are compared to the percentages allowed for the scrap material group "unalloyed steel" in Figure 7.134 [7.186]. This figure shows that the percentages of lead (Pb), which can produce a poisonous gas, and both copper (Cu) and antimony (Sb), which cannot be removed, are too high. Thus the recycling of the bearing as a whole has a negative effect on the reprocessing effort and resulting steel quality. Removal of the copper-containing "lubrication ring" and "cast bearing shells" is not economic prior to preprocessing, for example, by shredding. A redesign of the bearing that takes into account recycling consists of choosing materials for these parts that are compatible with the other alloys in the main scrap material group. The lubrication rings, for example,

% of admissible weight / Scrap material group	Evaporating elements ↑			Slag making elements →									Addition to slag				Non separable elements ↓								Evaporating elements ↑ (Toxic)			
	Li	C	Zn	Ca	Mg	Al	B	V	Ti	Si	Nb	Zr	Mn	Cr	P	S	As	Sb	Co	W	Mo	Ni	Sn	Cu	Cl	Cd	Be	Pb
Unalloyed steel	Any	2,0 (3,0)	4,0	Any	2,0 (3,0)	2,0 (3,0)	0,2 (0,5)	2,0 (3,0)	2,0 (3,0)	1,5 (2,5)	2,0 (3,0)	1,0 (1,5)	3,0 (4,0)	0,6	0,1 (0,3)	0,1 (0,2)	≪ 0,1	≪ 0,1	0,1	≪ 0,05	≪ 0,05	≪ 0,1	≪ 0,03	≪ 0,3	•	≪ 0,1	≪ 0,1	≪ 0,1
Plain pedestal bearing ERNLB 18-180	% of weight (more than 1%)																											
% of weight	-	2.12	0.012	-	-	0.08	-	-	-	1.16	-	-	0.17	-	0.16	0.16	0.015	0.4	-	-	-	0.01	0.15	0.8	-	0.02	-	2.07

Figure 7.134. Comparison of the weight percentages of the elements in a plain pedestal bearing with the percentages allowed for the scrap material group "unalloyed steel" (Renk-Wülfel)

could be made out of an aluminium alloy with a low copper content (for example $AlMg_3$), and the bearing shell from grey cast iron, with or without a plastic coating.

Recycling of White Goods (Used Material and Product Recycling)

White goods such as washing machines, dishwashers, refrigerators, etc., are valuable for recycling because they are produced in large numbers and contain valuable materials. Figure 7.135 shows the weight percentages of the main materials in a dishwasher. There are numerous nonferrous metals and nonmetals, and a particularly high percentage of high alloy steels. Preprocessing the product as a whole by, for example, shredding is not economic because the high alloy steels cannot be reprocessed separately. In addition, the nonferrous metals complicate the reprocessing process, or at least increase the reprocessing effort. A product structure more suitable for reprocessing would comprise main assemblies that are easy to

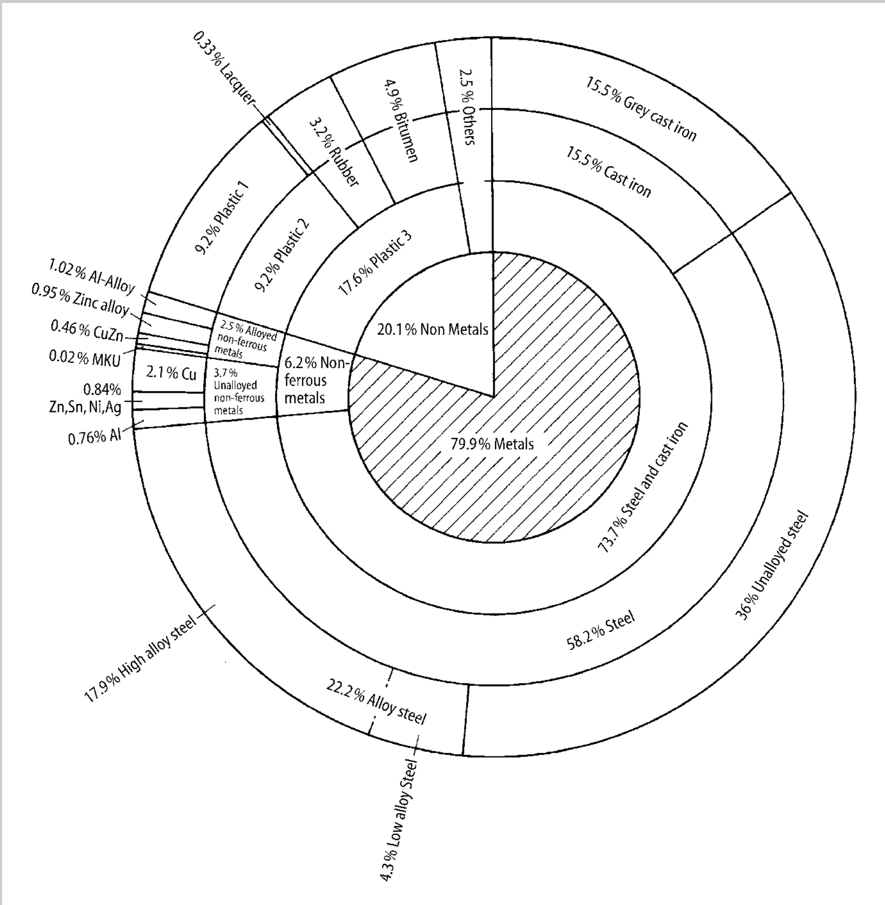


Figure 7.135. Material weight percentages of an AEG dishwasher from 1979/80, after [7.186]

separate or disassemble so that they can be preprocessed separately by, for example, shredding, cutting or compacting. This might also enable the reuse of individual components or even the whole product (product recycling).

Figure 7.136 shows an embodiment variant for the dishwasher. In this embodiment, the base *1* contains all of the accessories including a circulating pump *2*, a water distribution pump *3*, a washing detergent pump *4*, and the electronics *5*. This base assembly has been designed so that no connecting elements are required for the components. They are simply kept in place by the lower part of the casing *6*. The casing and the base can be opened and closed by means of the hinge *7*. The maximum angle achieved after tilting the casing is large enough to assemble all the components and to remove them for recycling (preprocessing or reconditioning).

Another example of white goods is shown in Figure 7.137. Different variants of a washing machine were produced by varying the construction structure of the housing and the location of the functional components. A Use-Value Analysis showed that variant B is the best because it has a lower number of parts and a lower number of reassembly interfaces, which is not only beneficial for recycling but also for maintenance.

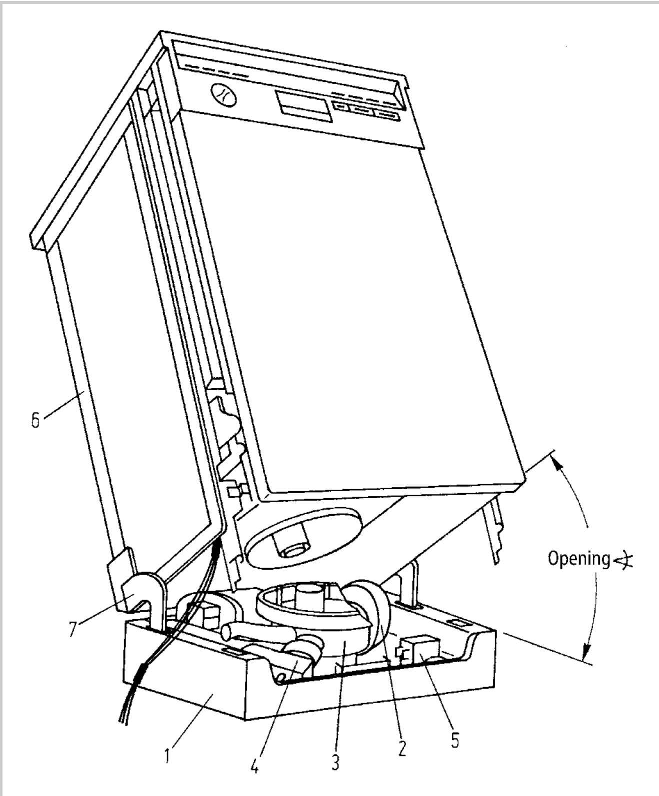


Figure 7.136. Dishwasher designed for recycling (Bosch-Siemens)

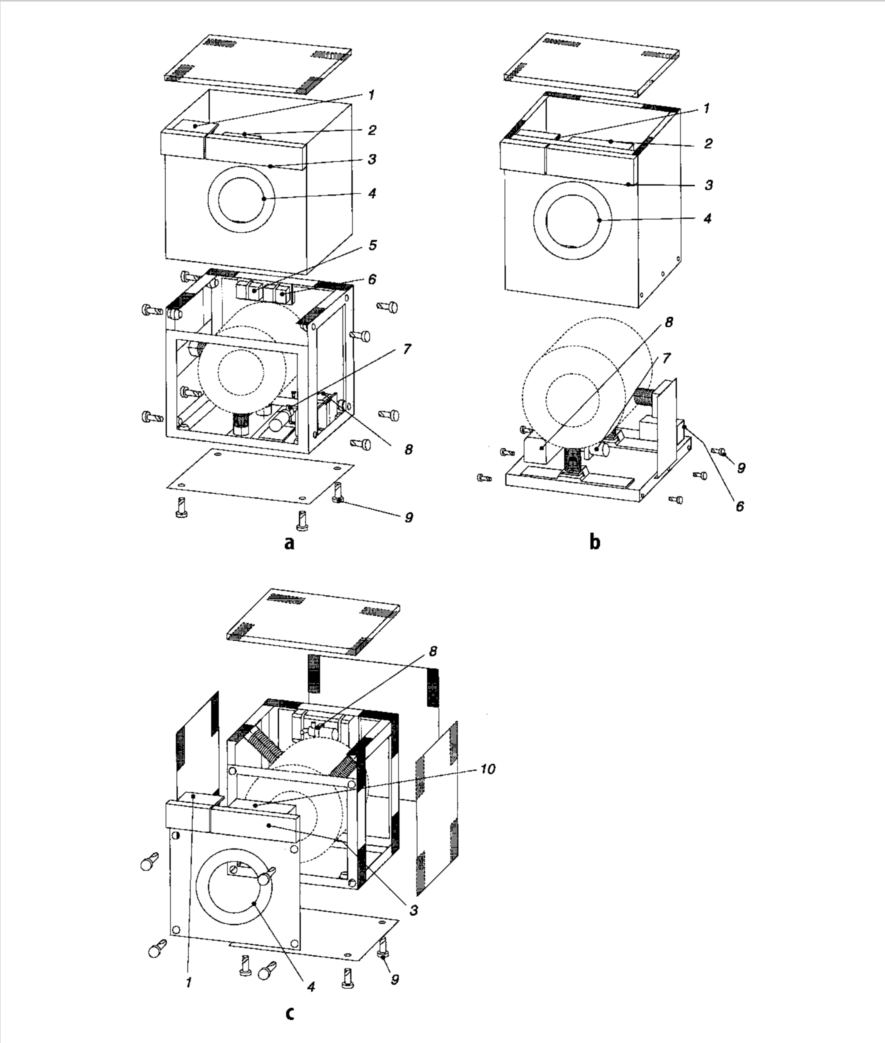


Figure 7.137. Construction structure variants of a washing machine (after Löser, TU Berlin). *1* Dispenser; *2* programme control; *3* display; *4* door; *5* socket and fuses; *6* power electronics; *7* detergent pump; *8* heater; *9* ¼-turn fastener; *10* central electrical unit

Disassembly-Friendly Drive Assembly

Figure 7.138 shows the drive assembly of a hammer drill in which the locating bearing of the motor shaft is not retained axially by the usual circlip, which in principle is easy to disassemble, but by a U-shaped clip that can be pulled out to separate the drive assembly from the motor. The reason for this solution is that a circlip would not be accessible in this assembly.

When developing products that are easier to recycle, particular care must be taken to ensure that they are not more expensive than traditional solutions as far as production and assembly are concerned.

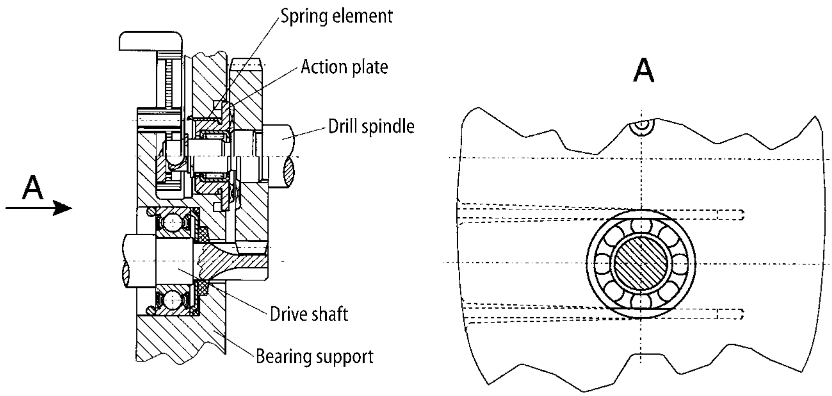


Figure 7.138. Disassembly-friendly gearbox of a hammer drill (Bosch) [7.118]

5. Evaluating Recycling Potential

When developing new products, it is necessary to evaluate solution variants against their potential for recycling [7.160,7.222]. Recycling criteria can simply be included in the evaluation procedure discussed in Section 3.3.2.

Figure 7.139 lists evaluation criteria, which are divided into those related to product recycling and those related to material recycling. To determine the overall rating, the recycling rating can be combined with the technical/economic rating of a product. S-diagrams and value profiles can be used, in particular those that show the distance of a solution variant from an imaginary ideal solution (see Section 3.2.2) [7.11,7.118], to illustrate the individual rating and the overall rating.

Such evaluation procedures can be extended into a product impact assessment [7.118, 7.263, 7.264, 7.324].

7.5.12 Design for Minimum Risk

Despite provisions against faults and disturbing factors (see Chapter 10), designers will still be left with gaps in their store of information and with evaluation uncertainties: for technical and economic reasons, it is not always possible to cover everything with theoretical or experimental analyses. Sometimes all designers can hope to do is to set limits. Thus, despite the most careful approach, some doubt may remain as to whether the chosen solution invariably fulfils the functions laid down in the requirements list or whether the economic assumptions are still justified in a rapidly changing market situation. In short, a certain risk remains.

One might be tempted to always design in such a way that the permitted limits are not exceeded, and to obviate any impairment of the function or early damage by designing a technical system to operate below its potential capacity. Experienced designers know that with this approach they rapidly encounter another risk: the cho-

Product recycling criteria	Material recycling criteria
Function-oriented product structure	**Ease of disassembly:**
Modular structure	Number of disassembly operations
Low complexity	Number of different disassembly operations
Parallel disassembly (flat disassembly tree)	Number of disassembly directions
Easy disassembly (cf. material recycling)	Number of connecting elements
Damage-free disassembly	Number of different connecting elements
Cleaning possibilities	Accessibility
Testing possibilities	Disassembly automation possibilities
Identification possibilities	Low energy for unlocking and separating
Sorting possibilities	Equipment expenditure
Reworking possibilities	Number of required disassembly tools
Reassembly possibilities	
Upgrading possibilities	**Ease of separation:**
Wear detection	Number of required separation process steps and expenditure
Use of standard components	Number of required special treatments and expenditure
Process automation possibilities	Material identification possibilities
	Number of materials to be separated
	Number of materials that cannot be recycled
	Reprocessing opportunities (Degree of utilisation):
	Reuse possibilities for identical functions
	Reuse possibilities for different functions
	Required reprocessing processes
	Degree of recycling
	Degree of quality reduction
	Material upgrading possibilities
	Degree of contamination

Figure 7.139. Evaluation criteria for product and material recycling, after [7.118, 7.197]

sen solution becomes too large, too heavy or too expensive and can no longer compete in the market. The lower technical risk is offset by the greater economic risk.

1. Coping with Risks

Faced with this situation, designers must ask themselves what countermeasures they can take; provided, of course, that the solution was carefully chosen in the first place and that the appropriate guidelines were scrupulously followed. The essential approach is that designers must, on the basis of the analysis of faults, disturbing factors and weak spots, provide a substitute solution to counter the possibility that the original solution might not cover all uncertainties.

In the systematic search for solutions, several solution variants should have been elaborated and analysed. In that case, the advantages and disadvantages of individual solutions will have been discussed and compared, which may have led to a new and improved solution. As a result, designers will be familiar with the range of possible solutions; they will have been able to rank them and also to take stock of the economic constraints.

In principle, the cheapest solution will have been selected, provided that it has sufficient technical merit. Although it may be more risky, it will afford the greatest economic leeway. The chances of marketing the resulting product, and hence of judging the validity of the solution, are greater than those of marketing a costlier product, which might jeopardise the entire development or, because of its "riskless" design, be unable to provide information about performance limits. While they are well advised to adopt this strategy, designers should assiduously avoid risky developments that might lead to damage, breakdowns and a great deal of unnecessary irritation.

If risks cannot be eliminated by theoretical analyses or experiments in good time or with justifiable outlay, designers may be forced to opt for a cheaper and riskier solution, but they should always keep a more costly, less risky alternative in reserve.

To that end, the less cost-effective solution proposals elaborated in the conceptual and embodiment phases should be developed into a second or third solution reserved for critical design areas, and ready for immediate use if needed. Provision for such development should be built into the chosen solution. If the latter does not meet all expectations, it can then be modified, if necessary step-by-step, without any great outlay of money and time.

This systematic approach not only helps to reduce economic risks for a tolerable outlay, but also to introduce innovations one-at-a-time, and to provide a detailed analysis of their performance, so that further developments can be made with minimum risk and at minimum cost. This approach must, of course, be coupled with a systematic follow-up of the practical experiences gained through it.

Through *design for minimum risk*, designers thus try to balance the technical against the economic hazards and so provide the manufacturer with valuable experience and the user with a reliable product.

2. Examples of Design for Minimum Risk

Example 1

A study of possible improvements in the performance of a stuffing box showed that, to increase the sealing pressure and the surface speed, the resulting frictional heat on the shaft had to be removed rapidly in order to keep the temperature in the sealing areas below the limit for the material used in the seals. To that end, it was suggested that the packing rings be mounted on the shaft so as to rotate with it and rub against the housing rather than the shaft. The heat generated by friction could then be extracted through the thin wall (see Figure 7.140a). Theoretical and experimental studies showed that a marked improvement could be obtained if forced convection cooling replaced natural convection cooling (see Figures 7.140b and 7.141).

This raised the difficult question of whether natural convection cooling would nevertheless meet the required operational conditions and, if not, whether the more elaborate and more costly alternative with its additional cooling circuit would be accepted by the customers.

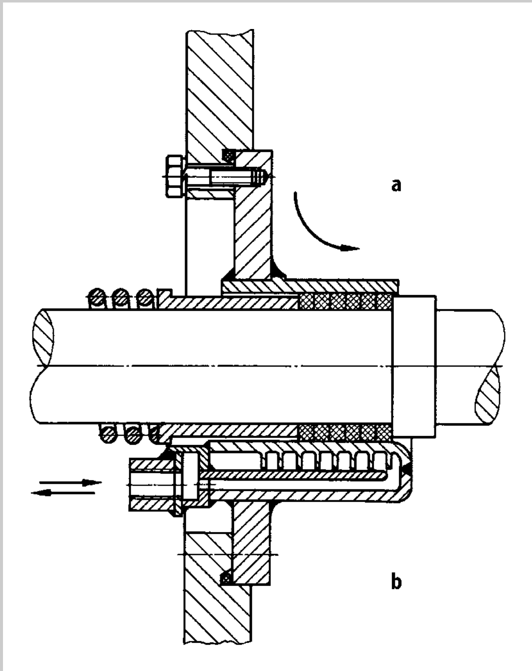


Figure 7.140. Cooled stuffing box in which the packing revolves with the shaft. Appropriate design of the shaft and press ring ensures the internal connection of the packing rings; a very short heat path facilitates good heat extraction. **a** Heat extraction by natural convection currents in the surrounding medium, dependent on the prevailing air flow; **b** heat extraction by forced convection due to separate cooling air flow, ensuring higher flow velocities and increased heat extraction

The minimum risk decision—that is, to construct the housing in such a way that either cooling system could easily be used—helped the designers gain experience for only a small increase in cost.

Example 2

In the development of a series of high-pressure steam valves operating at temperatures of more than 500°C, the question arose as to whether the customary method of nitriding the valve spindles and bushes should be retained despite the fact that the nitrided surface expands with temperature (thereby reducing the radial clearance), or whether very much more expensive stellite hard facing would have to be substituted. When the problem first arose, there was a lack of adequate information about the long-term behaviour of such layers at high temperatures. The minimum risk solution adopted was to select wall thicknesses and dimensions of valve spindles and bushes such that, if necessary and without changing the other components, stellite-treated parts could be substituted for the others. As it turned out, the operating temperature range was considerably lower than had been anticipated, so that nitriding provided a satisfactory solution and also helped to

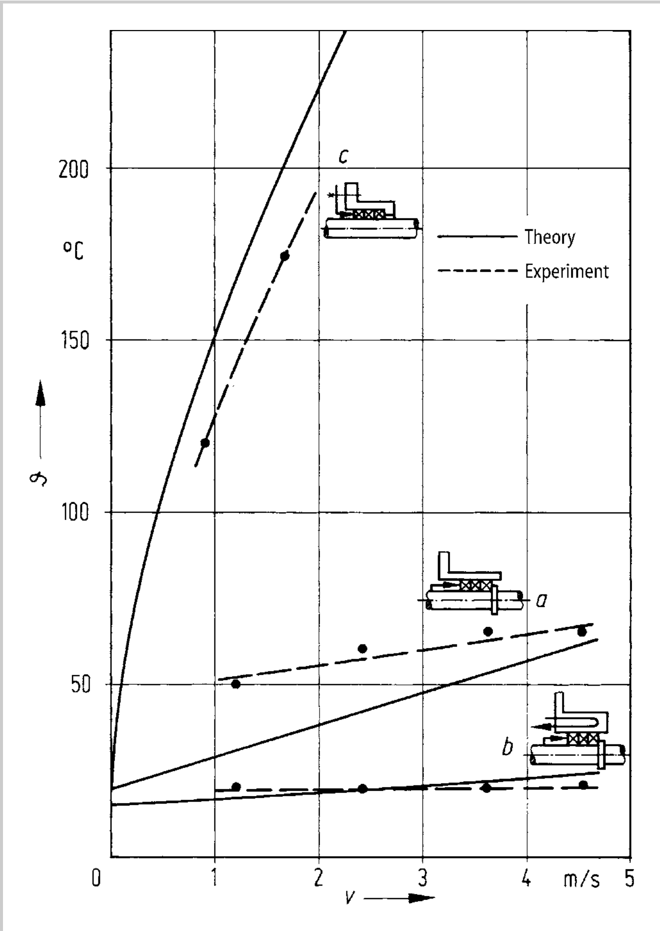


Figure 7.141. Theoretical and experimental temperatures at the seal, plotted against the peripheral speed on the shaft: **a** layout as in Figure 7.140a; **b** layout as in Figure 7.140b; **c** conventional stuffing box with packing attached to the housing

identify the operational limits. Once these limits were known, the more expensive solution could be reserved for more demanding conditions.

Example 3

Reliable design calculations for large machine parts, particularly in one-off production, depend on the analytical methods and the postulated constraints. It is not always possible to predict all characteristics with the necessary degree of accuracy. This applies, for instance, to the determination of the critical whirling speeds of shafts. Often it is impossible to predict the precise flexibility of the bearings and foundations. However, the difference between higher critical whirling speeds in high-speed installations is small within the range of flexibilities normally encountered. In the situation depicted in Figure 7.142, minimum risk design can once again be applied advantageously because the spacing of the bearings, which has a major influence on the critical speed, can be adjusted (see Figure 7.143). Interposed spring laminations (see Figure 7.144), moreover, allow alteration of the effective flexibilities. Both measures, taken together or separately, will produce the

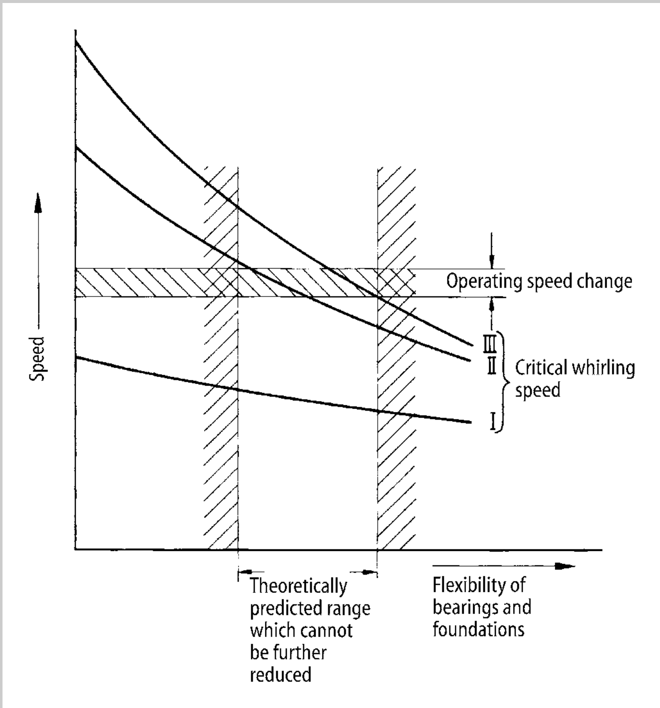


Figure 7.142. Critical whirling speeds (qualitative) for a shaft plotted against the flexibility of bearings and foundations

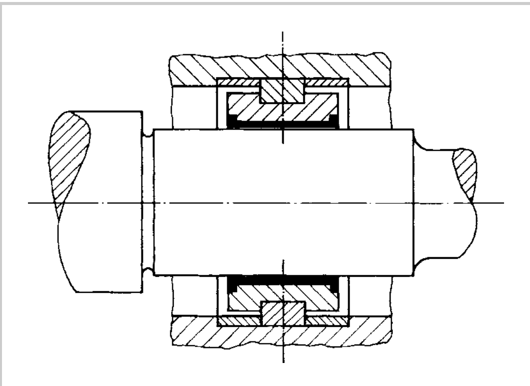


Figure 7.143. Support that allows the distances between the bearings to be varied through the selection of different spacers

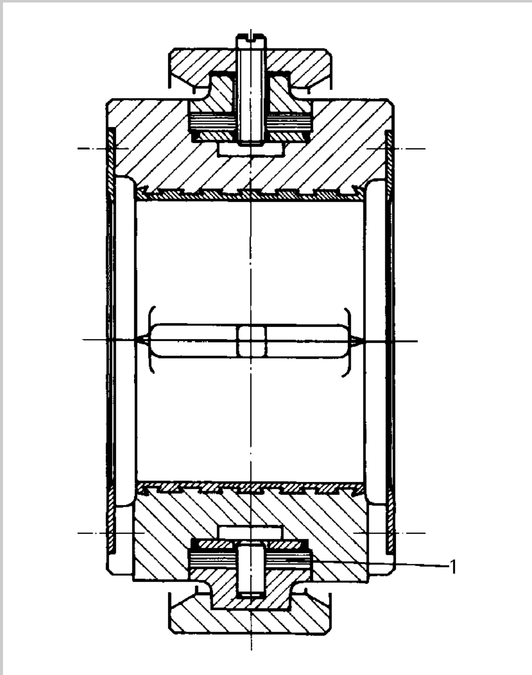


Figure 7.144. Plain bearings with laminated springs 1 that allow the flexibility to be adjusted (laminated springs also have good clamping properties, thus narrowing the critical range)

required effect so that the second or third critical whirling speeds can be eliminated from the operating speed range of the machine.

Example 4

Among the many suggestions put forward for a device to wind a strip into a double-layered ring, two seemed particularly promising (see Figures 7.145a and b).

The solution shown in Figure 7.145a is the simpler and cheaper but also the riskier of the two, because it is not certain whether the inner rotating mandrel *1* alone is invariably able—despite the increased friction produced by the knurling and the pressure of the springs *2*—to move the strip *3* forward.

The solution shown in Figure 7.145b is less risky, because the pressure rollers attached to the ends of the springs and the feed roller *5*, which moreover can be power-driven, make the advance of the strip more certain. This solution, however, is the more costly of the two, and also more susceptible to wear because of the greater number of small moving parts.

The minimum risk solution is to adopt the one shown in Figure 7.145a, but with a feed-in roller as in Figure 7.145b, and arranged in such a way that, if necessary, it can be driven without alteration of the other parts, see Figure 7.145c. This additional feed-in roller proved essential when the machine was tested, and was readily available.

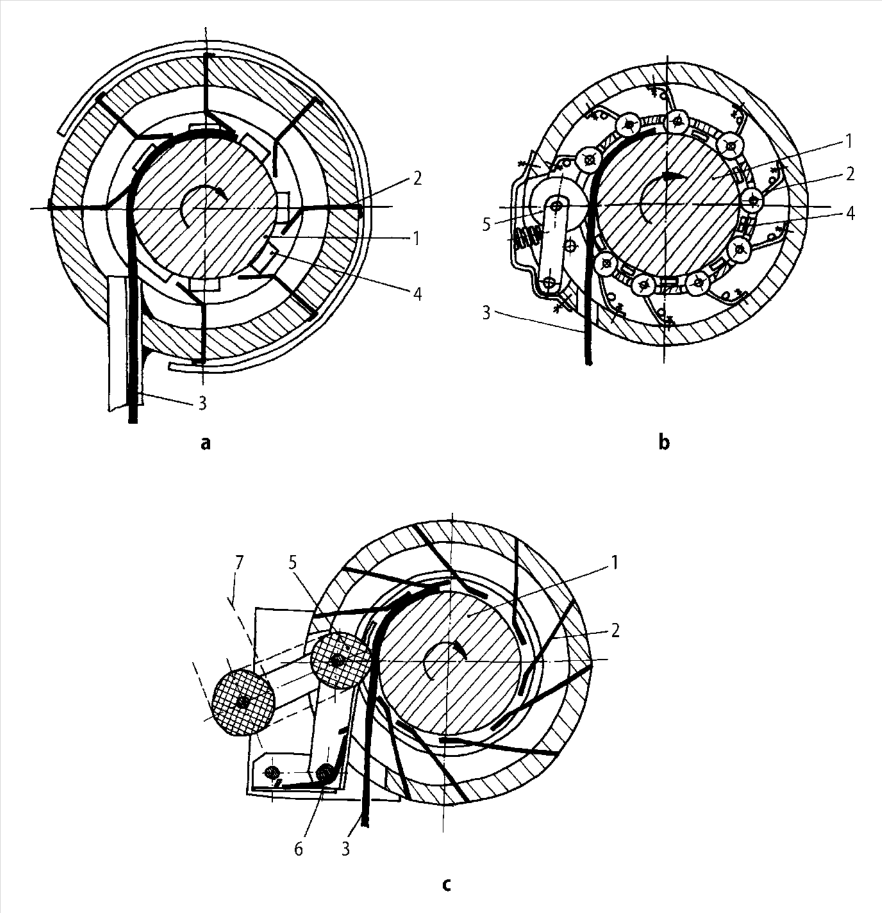


Figure 7.145. **a** Proposed winding device: *1* rotating mandrel, *2* pressure springs, *3* strip to be wound, *4* parts of the ejection mechanism; **b** proposed winding device: *1* rotating mandrel, *2* springs with pressure rollers, *3* strip to be wound, *4* parts of the ejection mechanism, *5* feed-in roller loaded by spring and possibly driven; **c** Chosen solution: *1* rotating mandrel, *2* pressure springs, *3* strip to be wound, *5* feed-in roller tensioned by spring *6* and driven by belt *7*

Example 5

In complex ventilation systems it is often very difficult to precalculate the airflow and pressure losses precisely. An embodiment with minimum risk for ventilators might have blades that can be adjusted before they are welded to the disc. When enough experience has been gained, it is possible to substitute a nonadjustable and cheaper cast construction.

All of these examples are intended to show that designers should meet risks not by simply considering the first potential problem but by also considering the second and third, which can often be done at relatively minor cost. Experience has shown that the application of emergency measures to correct unforeseen faults is many times more costly and time consuming.

7.5.13 Design to Standards

1. Objectives of Standardisation

If we examine the systematic approach outlined in this book in the light of the minimisation of effort, we are bound to ask to what extent can generally applicable function carriers be determined and documented so that designers can have ready access to tested solutions; that is, to known elements and assemblies. This has also been raised in connection with standardisation which, according to Kienzle [7.149], can be defined as follows: "Standardisation lays down the definitive solution of a repetitive technical or organisational problem with the best technical means available at the time. It is therefore a form of technical and economic optimisation limited by the time factor." Further definitions can be found in [7.34, 7.85].

Standardisation considered as the unification and determination of solutions, for instance in the form of national and international standards (BSI, DIN, ISO), of company standards, or of generally applicable design catalogues, and also of data sheets, is becoming of increasing importance in systematic design. Here, the fact that the objectives of standardisation are to limit the range of possible solutions in no way conflicts with the systematic search for a multiplicity of solutions, because standardisation is largely confined to the determination of individual elements, subsolutions, materials, computation and testing procedures, etc., while the search for a multiplicity of solutions and their optimisation is based on the combination or synthesis of known elements and data. Standardisation is therefore not simply an important complement to but a prerequisite of the systematic approach, in which various elements are combined as so many building blocks.

Traditionally, the speed of research and development allowed standards to be formulated only after the relevant knowledge had been verified and proven in practice. Today, the pace of change, for example in information technology, means that regulations and standards regarding new technologies increasingly have to depend on less well-tested knowledge. This situation is has arisen due to the need to remain competitive in a global market and to influence the direction of further developments. Development and standardisation therefore go hand-in-hand [7.98, 7.128, 7.226], resulting in the increasing publication of pre-standards.

In what follows, we shall be examining the possibilities of, the need for, and the limits of standards in the design process. In addition, the reader is referred to the literature [7.34, 7.85, 7.89–7.93, 7.129, 7.150, 7.216, 7.220, 7.227].

2. Types of Standard

The following discussion of *types of standard* is meant to:

- encourage designers to make wide use of standards
- invite them to suggest new standards or, at the very least, to influence the development of standardisation

- remind them of the crux of standardisation, namely, the systematic arrangement of facts with a view to their unification and optimisation in the light of functional considerations.

In terms of their *origin*, we distinguish between:

- national standards of the BSI (British Standards Institution) or the DIN (Deutsche Institut für Normung: the German Standards Institution)
- European (EN) standards of the CEN (Comité Européen de Normalisation) and CENELEC (Comité Européen de Normalisation Electrotechnique)
- recommendations and standards of the IEC (International Electrotechnical Commission)
- recommendations and standards of the ISO (International Organisation for Standardisation).

In terms of their *content*, we distinguish, for instance, between communication standards, classification standards, type standards, planning standards, dimensional standards, material standards, quality standards, procedural standards, operational standards, service standards, test standards, delivery standards and safety standards.

In terms of their *scope*, we distinguish between basic standards (general and interdisciplinary standards) and special standards (standards used in specialist fields).

Besides the national and international standards we have mentioned, designers can also refer to the rules and regulations published by professional engineering institutions, e.g. VDI, ASME, IMechE. These are important as they pave the way for further standardisation after initial trials.

Designers can also refer to a variety of *internal company standards* and regulations [7.86–7.88]. These can be classified as follows:

- compilations of representative standards; that is, a *selection* of general standards that are applicable to the special requirements of a particular company, such as stock lists and comparisons of old with new standards (synoptic standards)
- catalogues, lists and data sheets on *bought-out parts*, including their storage and also data on the acquisition (ordering/supplying) of raw materials, semi-finished materials, fuels, etc.
- catalogues or lists of *in-house parts*, for instance machine elements, repeat parts, standard solutions, assemblies, etc.
- information sheets used for *technical and economic optimisation*, for instance those on production capacity, production methods, cost comparisons (see Section 11.3.2).
- rules and regulations for the *calculation and embodiment design* of machine elements, assemblies, machines and plant, if necessary with a selection of sizes and/or types

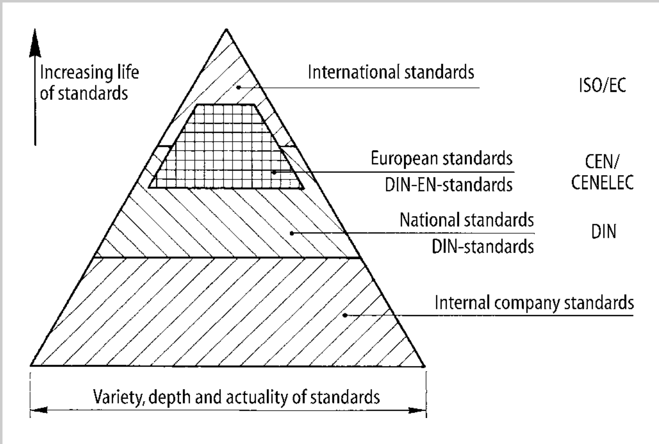


Figure 7.146. Relation between company, national, European and international standards based on DIN

- information sheets on *storage and transport* resources
- regulations concerning *quality control*, for example inspection and testing procedures
- rules and guidelines for the *preparation and processing of information*, for instance of drawings, parts lists, numbering systems and electronic data processing
- rules laying down *organisational and working procedures*, for instance the updating of parts lists and drawings.

The relation between company, national, European and international standards is shown in Figure 7.146. Company standards are developed or selected for specific products or processes and adapted to the actual situation. This implies that their depth and actuality is high. National and international standards require a longer period of development, but are more generally applicable. The variety and depth of these standards is generally lower. It is more difficult to adapt them to changes and therefore their dissemination and effects are more important.

3. Preparing Standards

Searching for and using standards, regulations and other information during the design process requires considerable effort. There are various ways in which information is made available: folders with standards, BSI or DIN handbooks and guides, microfiche and increasingly computer databases. These databases are now being integrated into CAD systems, providing designers not only with the textual information contained in standards, but also with the geometry of the components [7.6, 7.124, 7.238, 7.239].

4. Using Standards

Although there are no absolutely binding standards in the legal sense at the time of writing, national and international standards are widely treated as regulations, adherence to which is of great advantage in the case of legal disputes. This is particularly true of safety standards [7.23, 7.57, 7.115, 7.254, 7.303].

In addition, all company standards should be considered binding within their sphere of application, not least for economic reasons.

The *sphere of application* of a given standard is largely set by Kienzle's definition (see above). A standard can only be valid and binding if it does not conflict with technical, economic, safety or even aesthetic demands. Even in the case of such conflicts, however, designers should guard against rejecting or replacing the relevant standards out of hand, without assessing the possible consequences. Moreover, they should never make such assessments by themselves, but should always consult the standards organisation and the head of department.

In what follows, a number of recommendations and hints for the correct use of standards are listed.

First of all, we recommend adherence to national basic standards since other standards are based on these and the preferred sizes laid down in them help to determine the dimensions of all components. If these basic standards are ignored, then unpredictable long-term consequences (for instance, in the provision of spare parts) and grave technical and economic risks may ensue.

The use of standards should be examined against the checklist provided in Section 7.2, as illustrated by the following examples.

Layout

The basic and special standards—especially constructional, dimensional, material and safety—must be fully taken into account. Testing and inspection procedures also influence the embodiment.

Safety

Established component, work and environmental safety standards and regulations must be rigorously observed. Safety standards must always be given precedence over rationalisation procedures and economics.

Production

Here, the observance of production standards is particularly important and that of factory regulations is binding. This necessitates the continual updating of the relevant standards. Designers should only deviate from production standards after a broad assessment of all the industrial and relevant market (purchase and sales) aspects.

Quality Control

Test standards and inspection rules are essential features of quality control.

Maintenance

Standard symbols (for instance, circuit diagrams) should be used and service standards should be provided.

Recycling

For reuse and reprocessing, test, material, quality, dimensional, production and communication standards are particularly important.

The above recommendations on the use of standards are by no means exhaustive—the work of designers is much too varied and complex for that, and the range of general and company standards much wider than we have been able to cover in our summary. By working their way down the checklist, designers can tell how much a particular standard fits the various headings quickly and accurately.

5. Developing Standards

Since designers bear much of the responsibility for the development, production and utilisation of products, they should play a leading role in the revision of existing standards, and the development of new ones.

To make a useful contribution to the development of standards, they must first determine whether the revision of an existing standard or the development of a new standard is technically and economically justified. There is rarely a clear-cut answer to this question. In particular, completely reliable assessments of the economic consequences are seldom possible because of the complex effects of in-house costs and market influences, and, in any case, they would involve considerable research.

The following principles of developing general standards, and particularly company standards, can be set down [7.7, 7.30, 7.33, 7.271]. Whether something should be standardised depends on several prerequisites, that is, the envisaged standard must:

- document the state of the art of the technology
- be accepted by the majority of experts in the field
- ensure the complete interchangeability of parts, for example if a standardised product is modified in such a way that it can no longer be freely interchanged even with respect to a single feature, its designation (identification number) must be altered
- only be used if it is economical and useful, that is, there must be a need
- only be altered for technical, not purely formal, reasons
- always support a simple, clear and safe solution
- not contain any provisions that conflict with the law, e.g. with monopoly restrictions or safety regulations

- not include solutions that are protected, e.g. by patents or copyright
- not formulate design and production details
- not concern topics that are developing rapidly
- not hinder technical progress
- not allow subjectiveness or interpretation
- not standardise fashion and taste
- not endanger the safety of humans and the environment
- not serve a single individual; that is, the people affected must be consulted during the development and no standardisation should take place when important groups are opposed.

Moreover, the following aspects should be considered:

- Standards must be unambiguous, framed in clear terms and easily understood [7.35].
- Standard dimensions must, as far as possible, agree with preferred number series.
- All standards must be based on SI units [7.93].
- The layout of a standard should support its use and application. In particular, the use of computer-based information systems should be facilitated [7.124, 7.238, 7.239].

The development of a standard should generally include the following steps:

- A standard is proposed.
- The proposal is discussed in a working committee which develops a draft standard.
- This draft is circulated to all interested and affected parties and modified.
- After the draft has been accepted, a pre-standard can be issued for evaluation purposes.
- The final standard is issued.

Because a standard can be regarded as an artificial system, its preparation should also follow the steps of systematic design (see Chapters 4, 6 and 7). This ensures the optimisation of a standard's content and layout, and facilitates its careful development, which can be subsequently verified.

The evaluation criteria set out in Figure 7.147, once again arranged in accordance with the checklist, can be of great help in the assessment of existing or newly proposed standards if they are used in conjunction with the usual evaluation procedure. Not all the evaluation criteria we have mentioned apply to the assessment of individual standards. Thus, the evaluation of a drawing standard is influenced by its clarity; by the improvement in communication; by the simplification of the design activity and the overall execution of the order it provides; by the degree to which it is generally accepted; and also by the costs its development entails. Before

Headings	Examples
Function	Lack of ambiguity ensured.
Working principle	Market position of the product favourably influenced.
Layout	Material and energy expenditure reduced. Complexity of the product reduced. Design work improved and simplified. Use of replacement parts facilitated.
Safety	Safety increased.
Ergonomics	Clarity of instructions improved. Psychological and aesthetic conditions improved.
Production	Materials handling, storekeeping, production and quality control facilitated. Precision and reproducibility ensured. Execution of the orders simplified. Planning improved. Production capacity increased.
Quality control	Inspection and testing simplified. Quality improved.
Assembly	Assembly facilitated.
Transport	Transport and packing simplified.
Operation	Operation clarified.
Maintenance	Replacement of parts improved. Spare parts service and maitenance facilitated.
Recycling	Recycling facilitated.
Costs	Costs of, and/or time spent on, design, work preparation, materials handling, production, assembly and quality control reduced. Test costs reduced. Calculations simplified. Electronic data processing introduced.

Figure 7.147. Evaluation criteria for the assessment of standards

they make an evaluation, standards engineers or designers should therefore grade the importance of the various evaluation criteria and discard those that may not apply. In much the same way as with the recommendations in Section 3.3.2, there must be an adequate value rating to justify the development of a standard.

7.6 Evaluating Embodiment Designs

In Section 3.3.2 we discussed the subject of design evaluation. The basic procedures outlined there apply equally well to the conceptual and to the subsequent phases. As embodiment progresses, the evaluation will, of course, rest on more and more concrete objectives and properties.

In the embodiment phase, the technical properties must be evaluated in terms of the *technical rating* R_t and the economic properties separately with the help of the calculated production costs in terms of the *economic rating* R_e. The two ratings can then be compared in a diagram (see Figure 3.35).

The *prerequisites* for this approach are the following:

- All the embodiment designs have the same degree of concreteness; that is, the same information content (for instance, rough designs must only be compared with rough designs). In many cases it suffices, while keeping the overall perspective in mind, to evaluate only those aspects that show marked differences from one another. Once that has been done, their relationship to the whole, of course, must be examined; for example the relationship between part costs and total costs.

- The manufacturing costs (materials, labour and overheads) can be determined (see Chapter 11). If a particular solution introduces subsidiary costs, such as operating costs, and demands special investment, then—depending on the point of view (the producer's or the user's)—these factors must be allowed for, if necessary by amortisation. In addition, optimisation can help to achieve a minimisation of production and operating costs.

If the calculation of manufacturing costs is omitted, then the economic rating can only be evaluated qualitatively, as it was in the conceptual phase. In the embodiment phase, however, costs should, in principle, be determined more concretely (see Chapter 11).

As we mentioned in Section 3.3.2, the first step is to establish the *evaluation criteria*. They are derived from:

- the requirements list:
  - desirable improvement on minimum demands (how far exceeded)
  - wishes (fulfilled, not fulfilled, how well fulfilled)
- the technical properties (to what extent present and fulfilled).

The comprehensiveness of the evaluation criteria can be tested against the headings of the checklist (see Figure 7.148), which is specially adapted to the level of embodiment attained.

At least one significant evaluation criterion must be considered for each heading, although sometimes more will be needed. A heading may only be ignored if the corresponding properties are absent from, or identical in, all the variants. This approach avoids subjective over-valuation of individual properties. It must be followed by the procedural steps outlined in Section 3.3.2. The economic feasibility should be established by this stage at the latest.

In the embodiment phase, the search for weak spots, errors and disturbing influences, along with their elimination, is essential, in particular when evaluating the final layout.

7.7 Example of Embodiment Design

The *conceptual design phase* involves a process that focuses mainly on functions and working structures and results in principle solutions (concepts).

In the *embodiment design phase*, the emphasis is on determining the construction structures of the individual assemblies and components. In VDI 2223 and in Chapter 4 (Figure 4.3) and Chapter 7 (Figure 7.1) of this book, a systematic approach is proposed that has been tested in practice. The variations in approach

Headings	Examples
Function, Working principle	Fulfilment in accordance with the selected working principle: efficiency, risk, susceptibility to disturbances
Layout design	Space requirements, weight, arrangement, fits, scope for modifications
Form design	Material utilisation, durability, deformation, strength, operating life, wear, shock resistance, stability, resonance
Safety	Direct safety methods, industrial safety, protection of the environment
Ergonomics	Human–machine relationship, workload, handling, aesthetics
Production	Risk-free methods, setting-up time, heat treatment, surface treatment, tolerances
Quality control	Quality standards, testing possibilities
Assembly	Unambiguous, easy, comfortable, adjustable, upgradable
Transport	Internal and external transportation, means of despatch, packing
Operation	Handling, operational behaviour, corrosion properties, consumption of resources
Maintenance	Servicing, checking, repair and exchange
Recycling	Disassembly, reuse potential, reprocessing potential
Costs	Evaluated separately (economic rating)
Schedules	Production schedule and completion date

Figure 7.148. Checklist for evaluating embodiment designs

and methods needed to deal with different tasks and problems are greater in embodiment design than in conceptual design. Embodiment design, characterised by a further elaboration of the selected principle solution, requires a more flexible approach, extensive knowledge of the relevant domain and greater experience.

Explaining embodiment design using examples for different tasks would require too much space. It would also be misleading because such examples might suggest that the specific approach described is the only correct one. The example used in the rest of this chapter is based on the principle solution discussed in Chapter 6. Its only purpose is to show how the main embodiment steps of Figure 7.1 are executed and linked together.

The embodiment task is the concretisation of the principle solution for the impulse-loading test rig for shaft–hub connections (see Section 6.6.2). That section described the clarification of the task and the setting up of the requirements list (see Figure 6.43); the identification of the essential problems through abstraction (see Table 6.2); the establishment of function structures (see Figures 6.44 and 6.45); the search for working principles (see Figure 6.46); the combination of working principles into working structures (see Figure 6.47); the selection of suitable working structures (see Figure 6.48); their concretisation into principle solution variants (see Figures 6.49 to 6.52); and the evaluation of these solution variants (see Figures 6.55 and 6.56). We now continue with the embodiment design of this example following the steps shown in Figure 7.1.

Steps 1 and 2: Identifying Embodiment-Determining Requirements and Clarifying Spatial Constraints

The following items from the requirements list were identified as determining the embodiment features:

- Determining layout:
  test connection held in position
  loading applied to stationary shaft in one direction only
  hubside load take-off variable
  torque input variable
  no special foundation.
- Determining dimensions:
  diameter of shaft to be tested $\leq$100 mm
  adjustable torque $T \leq 15\,000$ Nm (maintained for at least 3 s)
  adjustable torque increase $\mathrm{d}T/\mathrm{d}t = 1.25 \times 10^3$ Nm/s
  power consumption ≤ 5 kW.
- Determining material:
  shaft and hub: 45C.
- Other requirements:
  production of the test rig in own workshops
  bought-out and standard parts wherever possible
  easy to disassemble.

The requirements list did not contain specific spatial constraints.

Step 3: Identifying Embodiment-Determining Main Function Carriers

The basis for this step was function structure variant No 4 (see Figure 6.45) and the principle solution variant V_2 (see Figure 6.47). Table 7.6 lists the *main function carriers* used in the selected solution variant to fulfil the various subfunctions, along with their main characteristics. The function carriers that determined the embodiment are:

- the test specimen
- the lever between the cylindrical cam and the shaft of the test specimen
- the cylindrical cam.

The other main function carriers are:

- the electric motor
- the flywheel

Table 7.6. Main function carriers

Functions	Function carriers	Characteristics
Transform energy; increase energy component	Electric motor	Power P_M Speed n_M Run-up time t_M
Store energy	Flywheel	Moment of inertia J_F Speed n_F Torque transmitted T_{CL}
Release energy	Clutch	Torque transmitted T_{CL} Maximum speed n_{CL} Response time t_{CL}
Increase energy component	Gearbox	Power P_G Maximum output torque T_G at output speed n_G Gear ratio R_G
Control magnitude and time	Cylindrical cam	Power P_{CAM} Torque T_{CAM} Speed n_{CAM} Diameter D_{CAM} Cam angle α_{CAM} Rise h_{CAM}
Transform energy into torque	Lever	Length l_L Stiffness s_L
Load test specimen	Test specimen	Torque T Rate of torque increase $\mathrm{d}T/\mathrm{d}t$
Take up forces and torque	Frame	

- the clutch
- the gearbox
- the frame.

Step 4: Developing Preliminary Layouts and Form Designs for the Main Function Carriers

Figure 7.149 shows a preliminary layout drawing for the three embodiment-determining function carriers.

The embodiment of the test specimen in line with DIN 6885 and of the transmission lever, modelled and analysed as a cantilever, were relatively straightforward. The development and embodiment of the cylindrical cam, however, required a more detailed kinematic and dynamic analysis based on specific items in the requirements list.

A more precise analysis showed that the initial estimates undertaken in the conceptual phase of the cylindrical cam's performance were insufficient to proceed directly to embodiment. The following analysis therefore had to be carried out before determining the main dimensions.

Figure 7.150 shows that:

Torque on the shaft: $T = s_L \cdot h_{CAM} \cdot l_L$

Torque increase: $dT/dt = \pi \cdot D_{CAM} \cdot n_{CAM} \cdot \tan\alpha_{CAM} \cdot s_L \cdot l_L$

Hold time: $t_L = \dfrac{U_{CAM}}{2\pi \cdot D_{CAM} \cdot n_{CAM}} = \dfrac{1}{2 \cdot n_{CAM}}$

The equation for the torque increase is only valid if the lever movement is parallel to the cam track. In order to minimise friction, a roller follower was required (see Figure 7.151), so the actual torque increase was lower than calculated and also varies. We therefore used the average increase in our calculations (see Figure 7.152).

If, in line with the requirements list, the average torque increase dT/dt is used, then the calculation of dT/dt should not involve the full circumferential speed v_X, but instead the effective circumferential speed v_X^*, thus:

$$v_X^* = K \cdot v_X$$

The correction K depends on:

- the cam angle α_{CAM}
- the diameter of the roller follower d
- the rise of the cylindrical cam h_{CAM}.

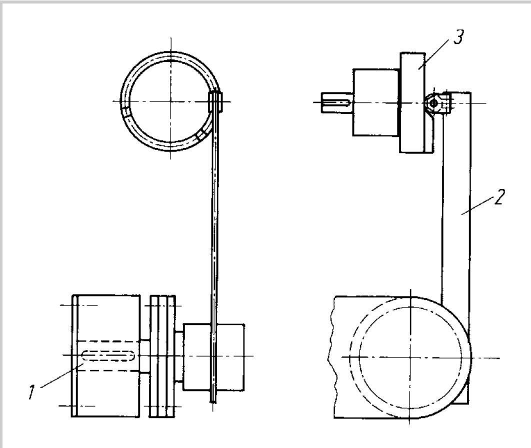


Figure 7.149. Main function carriers that determine the layout: *1* test connection; *2* transmission lever; *3* cylindrical cam

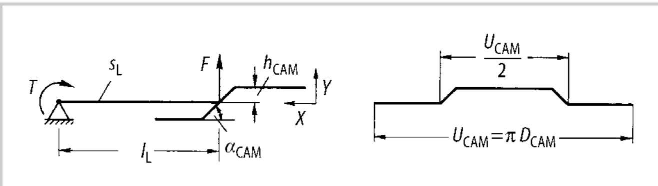


Figure 7.150. Geometric constraints for the cylindrical cam and lever. s_L is the stiffness of the lever

The correction K was derived from Figure 7.153:

$$x = \frac{h_{\mathrm{CAM}}}{\tan \alpha_{\mathrm{CAM}}}$$

$$x = d/2 \cdot \left(\sin \alpha_{\mathrm{CAM}} - \frac{1 - \cos \alpha_{\mathrm{CAM}}}{\tan \alpha_{\mathrm{CAM}}}\right)$$

$$K = \frac{v_{\mathrm{X}}^{*}}{v_{\mathrm{X}}} = \frac{x}{x + \Delta x}$$

The formula is only valid when $d/2 \cdot (1 - \cos \alpha_{\mathrm{CAM}}) \leq h_{\mathrm{CAM}}$, for example:

$$K = \frac{\frac{h_{\mathrm{CAM}}}{\tan \alpha_{\mathrm{CAM}}}}{\frac{h_{\mathrm{CAM}}}{\tan \alpha_{\mathrm{CAM}}} + d/2 \cdot \left(\sin \alpha_{\mathrm{CAM}} - \frac{1-\cos \alpha_{\mathrm{CAM}}}{\tan \alpha_{\mathrm{CAM}}}\right)}$$

To obtain a value for K, the following estimates were made:

- cam angle $\alpha_{\mathrm{CAM}} = 10 \ldots 45°$

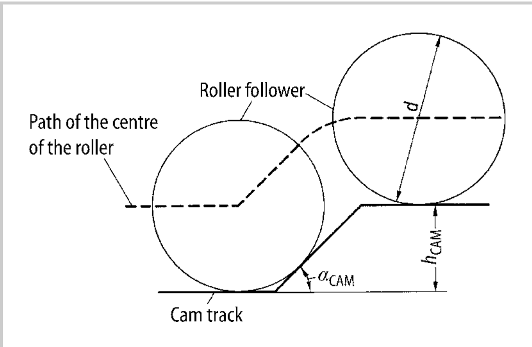


Figure 7.151. Cam path and lever movement

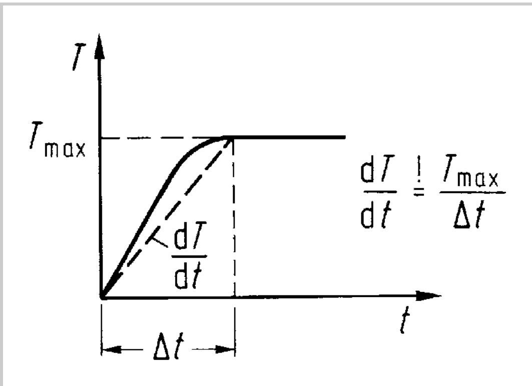


Figure 7.152. Torque increase

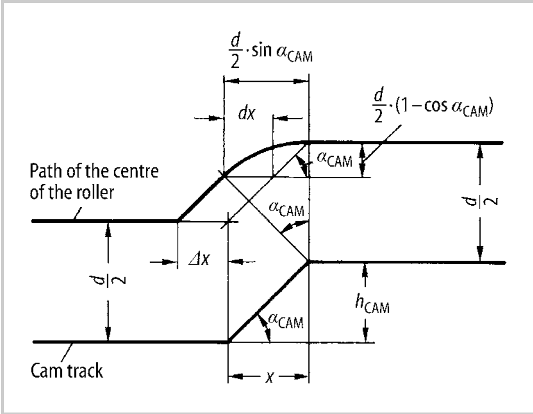


Figure 7.153. Derivation of correction K

- the diameter of the roller follower $d = 60$ mm
- the rise of the cylindrical cam $h_{CAM} = 7.5$ mm and 30 mm, respectively.

Table 7.7 contains the values of K obtained from the above formula.

After converting the cylindrical cam speed n_{CAM} and using the calculated correction value K, the formula including torque increase $\mathrm{d}T/\mathrm{d}t$ became:

$$n_{CAM} = \frac{\frac{\mathrm{d}T}{\mathrm{d}t}}{K \cdot \pi \cdot D_{CAM} \cdot \tan \alpha_{CAM} \cdot s_L \cdot l_L}$$

The speed controller range C

$$C = \frac{n_{CAM_{max}}}{n_{CAM_{min}}}$$

was determined as follows.

If the diameter of the cylindrical cam D_{CAM}, the stiffness s_L and the length l_L of the lever are considered constant for this solution concept, the above formula can be used to calculate the extremes of the speed n_{CAM} in relation to the other parameters $\mathrm{d}T/\mathrm{d}t$, K and α_{CAM} (see Table 7.8).

B is a constant that includes units and the other constants (π, D_{CAM}, s_L, l_L).

Table 7.7. Reference values for K corrections

h_{CAM} mm	α_{CAM}	45°	40°	30°	20°	10°
7.5	K	0.41	0.45	0.62	0.79	0.94
30.0	K	0.71	0.76	0.87	0.94	0.98

Table 7.8. Determination of $n_{CAM_{min}}$ and $n_{CAM_{max}}$

	dT/dt	α_{CAM}	K	n_{CAM}
Minimum	20	10	0.98	116 · B
Maximum	125	45	0.41	305 · B

The speed control range C therefore became:

$$C = \frac{305 \cdot B}{116 \cdot B} = 2.6$$

This meant that:

- The function "control magnitude and time" could not be fulfilled by the cylindrical cam alone.
- The function structure had to change if we wished to maintain the principles underpinning the concept.
- The cylindrical cam had to have an adjustable drive with a speed control range of approximately $C = 2.6$.

Figure 7.154 shows the adapted function structure variants (see Figure 6.45). The subfunction "adjust speed" was added. This could, for example, be realised by a continuously adjustable drive motor. Several variants were possible (4/1 to 4/3).

The quantitative developments of the cylindrical cam based on these formulae resulted in the following values for the main characteristics: spring stiffness of the lever $s_L = 700$ N/mm; lever length $l_L = 850$ mm; cylinder diameter $D_{CAM} = 300$ mm; cam angle $\alpha_{CAM} = 10 \ldots 45°$; constant $B = 0.107\ \text{min}^{-1}$ (see Table 7.8); speed range for the required rate of torque increase ($dT/dt_{min} = 20 \times 10^3$ Nm/s, $dT/dt_{max} = 125 \times 10^3$ Nm/s), $n_{CAM} = 12.4 \ldots 32.6\ \text{min}^{-1}$ for a control range $C = 2.6$.

The requirements for the adjustable torque increase dT/dt could therefore be realised with the selected values.

This was not the case for the required hold time for the maximum torque. This value was $t_L = 0.5 \cdot n_{CAM} = 2.4 \ldots 0.92$ s, which was lower than the required value of 3 s. After a discussion with the client, the requirement was reduced to $t_L \geq 1$ s, which could be realised by using slightly more than half of the circumference of the cylindrical cam.

Before a scale layout for the main function carriers that determine the embodiment could be drawn, the following issues had to be resolved:

- What spatial layout of the test specimen and the cylindrical cam should be used?
- To what extent should auxiliary function carriers be considered?

It was decided that the test specimen should be positioned horizontally, and as a consequence the cylindrical cam should rotate about a vertical axis for the following reasons:

- Easy exchange of test specimen and cylindrical cam (design for assembly).

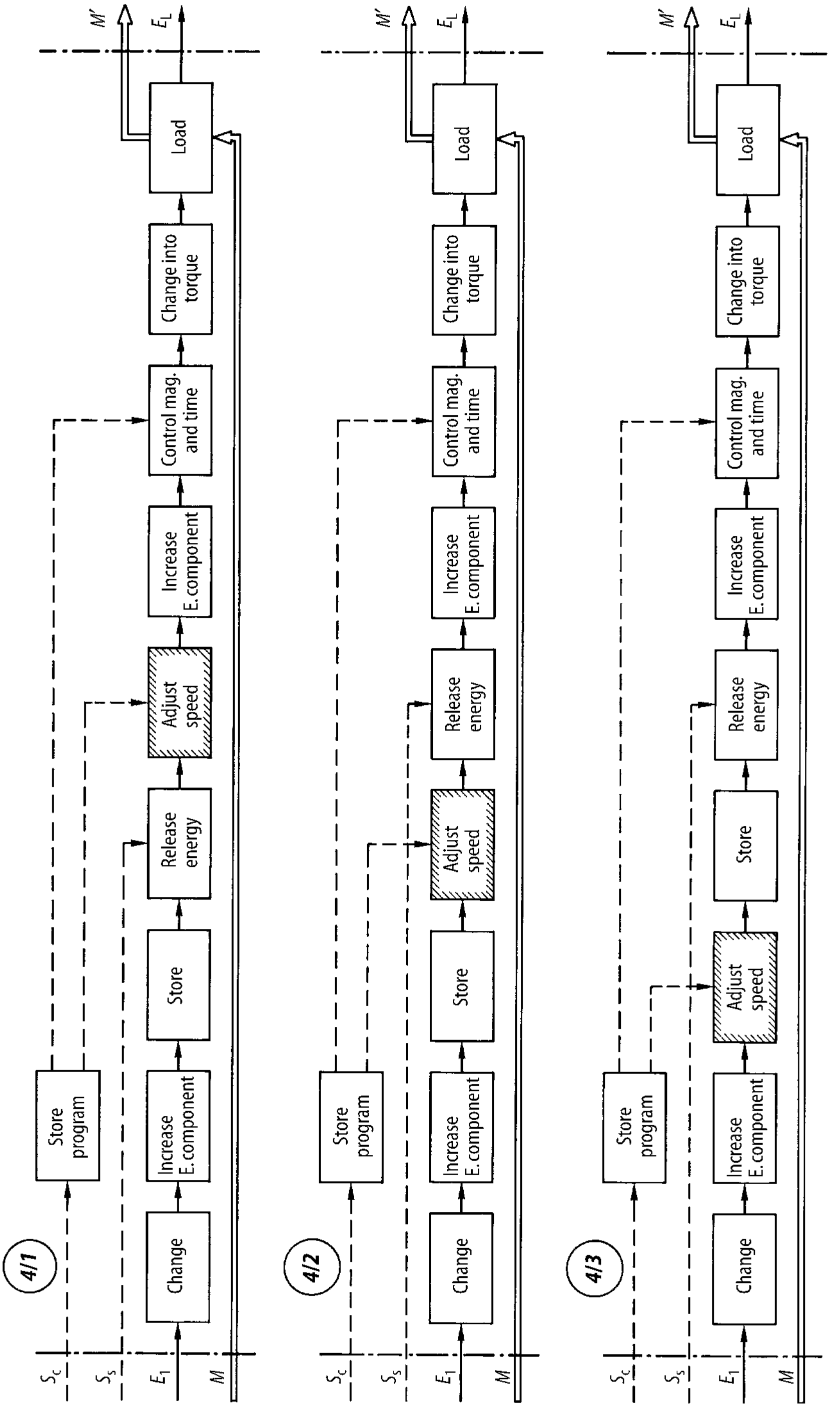


Figure 7.154. Function structure variants for function structure 4, after Figure 6.45

- Easy access to the test specimen for measurements (design for ergonomics).
- Smooth transmission of the clamping forces of the test specimen into the foundation (short and direct force transmission paths).
- Easy resetting of the test rig for different types of specimen, in particular larger specimens (design for minimum risk).

The need for auxiliary function carriers was then assessed and the space requirements determined on the basis of experience. It was found that:

- A separate bearing was needed for the cylindrical cam because of the axial force F_A and the tangential force F_T:

$$F_A = F_T = \frac{T_{max}}{l_L} = 17.6 \text{ kN}$$

- The outer diameter of the bolted joint between test specimen and lever had to be about 400 mm to provide a torsionally stiff connection.

The analysis showed that the auxiliary function carriers had only a marginal influence on the dimensions of the embodiment.

Figure 7.155a shows a preliminary layout based on function structure variant *4/1*, where the speed control is achieved by means of an adjustable mechanism that is located behind the clutch in terms of the energy flow. Figure 7.155b shows

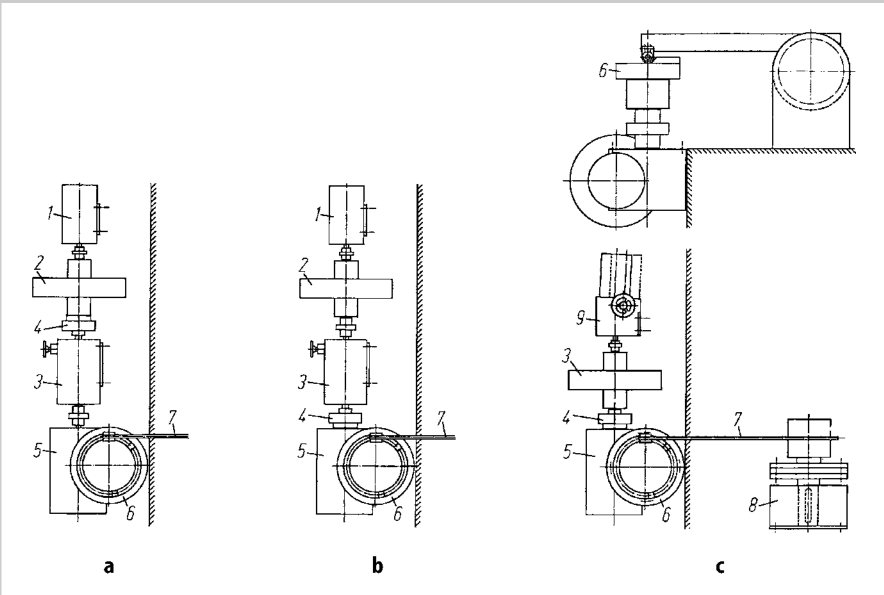


Figure 7.155. Layout of main function carriers: **a** for function structure variant *4/1*; **b** for function structure variant *4/2*; **c** for function structure variant *4/3*; *1* motor, *2* flywheel, 3 adjustable gear, *4* clutch, *5* worm gear (angular), 6 cylindrical cam, *7* transmission lever, *8* test connection, *9* adjustable geared motor

a preliminary layout based on function structure variant *4/2*, where the adjustable mechanism is located before the clutch. Variant *4/3* (see Figure 7.155c) employs an adjustable geared motor.

Step 5: Selecting Suitable Preliminary Layouts

Variant *4/3* was selected for further detailing because it took up less space due to the adjustable geared motor (function integration).

Step 6: Developing Preliminary Layouts and Form Designs for the Remaining Main Function Carriers

The preliminary layouts and form designs for the remaining main function carriers were based on the following requirements identified in step 4:

- motor drive speed for cylindrical cam

$$n_{\mathrm{CAM}} = 12.4\ldots32.6\ \mathrm{min}^{-1}$$

- speed control range

$$C = 2.6$$

- driving torque of cylindrical cam

$$T_{\mathrm{CAM}} = F_{\mathrm{T}} \cdot D_{\mathrm{CAM}}/2 \text{ and } F_{\mathrm{T}} = F_{\mathrm{A}} = T/l_{\mathrm{L}} \text{ gives } T_{\mathrm{CAM}} = 2650\ \mathrm{Nm}$$

- driving power of cylindrical cam

$$P_{\mathrm{CAM}} = T_{\mathrm{CAM}} \cdot \omega_{\mathrm{CAM}}, \text{ thus } P_{\mathrm{CAM}} = 9\ \mathrm{kW}$$

For reasons of safety, the maximum flywheel speed n_{F} (and therefore also that of the motor n_{M}) was chosen to be:

$$n_{\mathrm{F}} = 1000\ \mathrm{min}^{-1}$$

This required a transmission ratio of:

$$i = 80.7\ldots30.7$$

For the other main function carriers, the characteristics were estimated as follows:

- Transferred torque of the coupling based on the driving torque of the cylindrical cam $T_{\mathrm{CAM}} = 2\,650$ Nm and the actual transmission ratio i between the cylindrical cam and clutch

$$T_{\mathrm{CL}} = T_{\mathrm{CAM}}/i$$

- Moment of inertia of the flywheel from the actual torque T_F taken up by the flywheel, the impact time Δt, the flywheel speed n_F and the allowable drop in speed $\Delta n = 5\%$

$$J_F = \frac{T_F \cdot \Delta t}{2 \cdot \pi \cdot n_{CAM} \cdot \Delta n}$$

- The power of the electric motor P_M after calculating the required acceleration torque T_A from the moment of inertia J_F of the flywheel, the motor speed n_M, the run-up time $t_M = 10$ s and the maximum acceleration torque of the motor $T_{A_{max}}$ (from manufacturer's data)

$$T_A = \frac{J_F \cdot 2 \cdot \pi \cdot n_M}{t_M} < T_{A_{max}}$$

Table 7.9 lists the calculated values for the main characteristics. Apart from the flywheel, the main function carriers could all be selected from catalogues and bought directly from suppliers.

The following characteristics were chosen for the flywheel:

- speed $n_F = 1010\ \text{min}^{-1}$
- moment of inertia $J_F = 1.9\ \text{kg m}^2$.

Because losses such as those from friction had not been taken into account, the final value of J_F was chosen to be substantially larger than this.

To save weight, the flywheel was made from a hollow cylinder:

- Outer diameter $D_o = 480$ mm
- Inner diameter $D_i = 410$ mm
- Width $W = 100$ mm
- Mass $m = 38$ kg.

The final preliminary layout drawing was then produced on the basis of the main function carriers shown in Figure 7.155c and by adding the frame.

Table 7.9. Calculated values for the characteristics of the main function carriers of variant *4/3*

Functions	Function carriers	Calculated values
Change energy Increase E-component Adjust speed	Electric motor with mechanical adjustment-variant 4/3	Power $P_M = 1.1$ kW Speed $n_M = 380 \dots 1000\ \text{min}^{-1}$ Speed control range $C = 2.6$
Store energy	Flywheel	Moment of inertia $J_F = 1.4\ \text{kg m}^2$ Speed $n_F = 380 \dots 1000\ \text{min}^{-1}$
Release energy	Electromagnetic clutch	Transferred torque $T_{CL} = 86$ Nm
Increase E-component	Gear	Power $P_G = 9$ kW Nominal torque $T_G = 2650$ Nm at speed $n_G = 32\ \text{min}^{-1}$ Transmission ratio $i_G = 40.7$

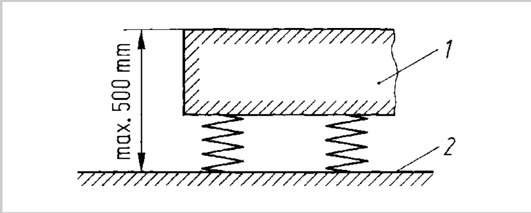


Figure 7.156. Final spatial constraints: *1*, base plate for fixing the test machine; *2*, foundation

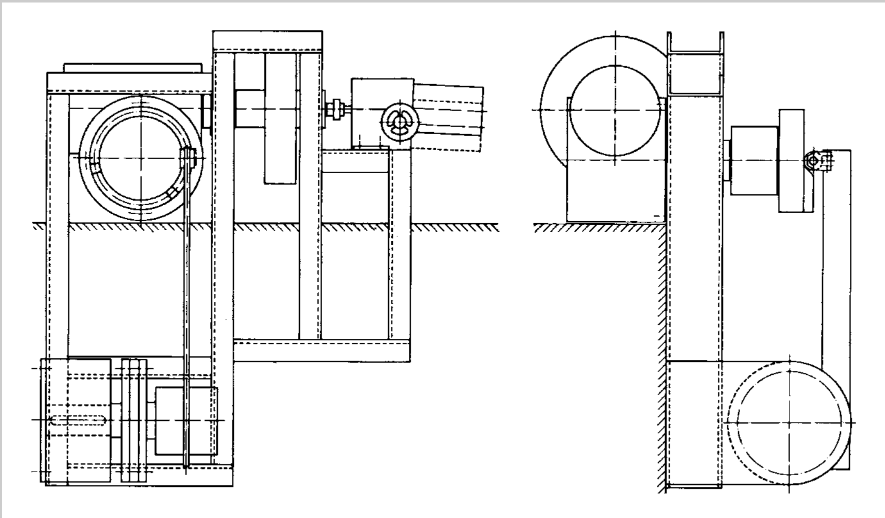


Figure 7.157. Preliminary layout drawing for the main function carriers

Because the combined height of the lever bearing and the test specimen was much smaller than the combined height of the cylindrical cam and the entire drive system, the spatial constraints for the test rig shown in Figure 7.156 were selected after a discussion with the client.

Steel channel sections were used for the frame for the following reasons:

- large second moment of area for a small cross-sectional area
- no round corners
- three flat reference surfaces available
- cheap.

Figure 7.157 shows the completed preliminary layout drawing for the main function carriers.

Step 7: Searching for Solutions for Auxiliary Functions

The production of a detailed layout drawing involved the following steps:

- Searching for and selecting auxiliary function carriers.

- Detailing the embodiment of the main function carriers based on the auxiliary function carriers.
- Detailing the embodiment of the auxiliary function carriers.

These steps were much more interrelated than those for the preliminary layout drawing. They influenced each other because they dealt with more concrete aspects which often required a repetition of previous steps on a higher information level.

The auxiliary function carriers were divided into three groups:

- Carriers that connect the main function carriers together.
- Carriers that support those main function carriers that move relative to the frame.
- Carriers that permanently connect main function carriers to the frame.

The auxiliary function carriers that connected the main function carriers together were:

- A bolted joint between the lever and test specimens; a form-fit membrane to avoid additional bending moments and to ensure easy assembly.
- A torsionally stiff connection between the worm gear pair and the cylindrical cam. This connection can be of two types (see Figure 7.158):
  - a worm gear pair with hollow shaft—cylindrical cam.
  - a worm gear pair—torsionally stiff connection—cylindrical cam.

  The following arguments favour the torsionally stiff connection:
  - separate assembly of worm gear pair and cylindrical cam possible (design for assembly).
  - no interruption of the frame caused by a high shaft position (simple embodiment).
  - easy centering of worm gear pair and cylindrical cam (design for production).
- Torsionally flexible connection between the flywheel and the electric motor.

The auxiliary function carriers used to support those main function carriers that move relative to the frame were:

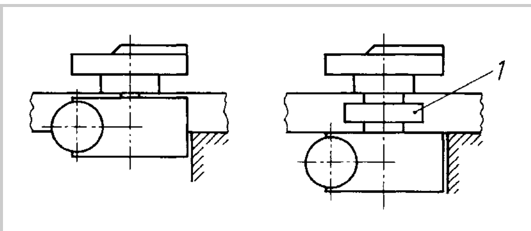


Figure 7.158. Connections between the worm gear pair and the cylindrical cam: *1*, coupling

- *Flywheel support*. The requirements were: simple production (i.e. no accurate balancing needed); direct safety techniques to withstand the dynamic forces (safe-life principle); and suspend from the frame. The use of bought-out parts (bearing housing with roller bearings) was not possible because these bearing housings are usually cast and are more suitable for standing rather than suspended applications. Because the flywheel was to be produced in-house, the magnitudes of the dynamic forces were relatively uncertain and so its support needed to be specially designed.
- *Support for the cylindrical cam and lever*. Commercially available rolling element bearings were selected.

The auxiliary function carriers used to permanently connect main function carriers to the frame were:

- Simple half-finished products (welded sheet steel), to which the main function carriers were bolted.
- A special solution for connecting the test specimen to the lever (i.e. the frame). The requirements were: easy to assemble but separable connection; movable in the axial direction; free of play; and no tight tolerances. A Ringfeder connection was chosen.

Step 8: Detailing the Main Function Carriers Taking into Account the Auxiliary Function Carriers

The main function carriers had to be adapted so as to match the solutions selected for the auxiliary function carriers. This resulted in the following:

- electric motor: bought-out part
- flywheel: see Figure 7.159
- clutch: bought-out part
- gearbox: bought-out part
- cylindrical cam: see Figure 7.160
- lever: see preliminary layout drawing in Figure 7.161
- test specimen: see preliminary layout drawing in Figure 7.161
- frame: modified to suit the geometry of the selected motor.

Step 9: Detailing the Auxiliary Function Carriers and Completing the Preliminary Layout

The flywheel support bearing is taken as an example, using the guidelines for embodiment design shown in Figure 7.3.

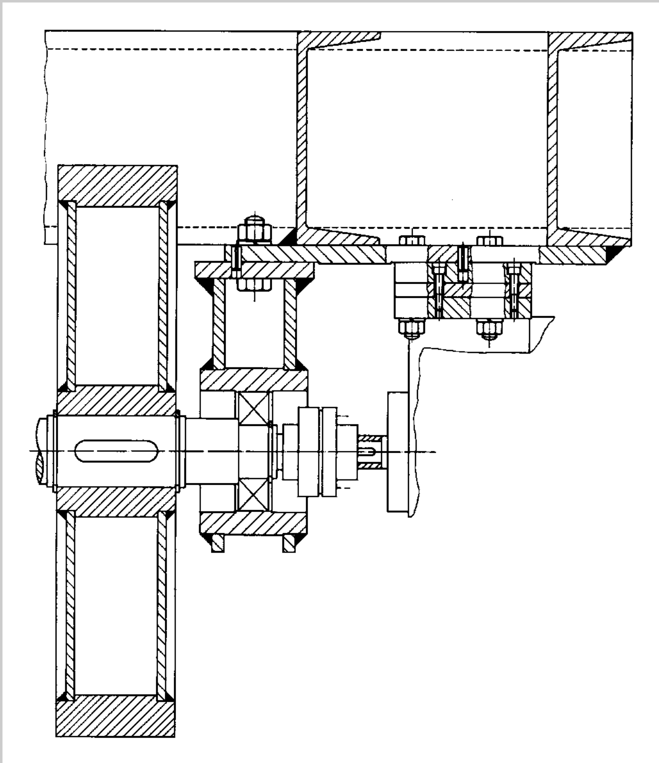


Figure 7.159. Detailed layout of the flywheel and the flywheel shaft bearing

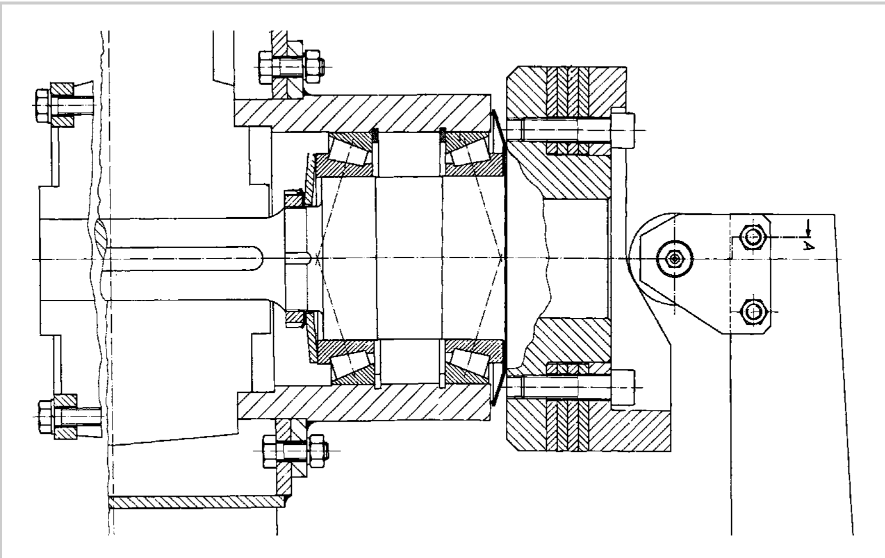


Figure 7.160. Detailed layout of the bearing arrangement for the cylindrical cam

Layout

The bearing forces were estimated as follows:

$$F_B = F_{dyn} + F_{stat}$$

with the weight being:

$$F_{stat} = m \cdot g = 400\ \text{N}$$

and the dynamic force being:

$$F_{dyn} = m \cdot e \cdot 4 \cdot \pi^2 \cdot n_F^2$$

With a mass $m = 40$ kg; speed $n_F = 1\,750\ \text{min}^{-1}$ (= max motor speed); eccentricity of flywheel $e = 0.6$ mm (based on: dimensional and shape accuracy of flywheel = 0.3 mm; play in flywheel shaft and bearings = 0.2 mm; and unbalanced mass distribution = 0.1 mm), the bearing force is:

$$F_B = 1130\ \text{N}$$

This implies that even when additional gyroscopic forces occur, the bearing (dynamic capacity 65 000 N) and all the other parts that are in the force transmission path have adequate dimensions.

Resonance

The embodiment of the bearing and frame was made very rigid so that resonance excited by the flywheel (maximum 30 Hz) was unlikely.

Production

The embodiment allowed easy production because the flywheel support bearing did not require tight tolerances for the frame.

Assembly

The support for the flywheel could be assembled easily due to:

- the application of a simple bottom-up approach
- the easy accessibility to the connecting screws
- the simple adjustment of the clutch using a spacer after accurate location of the flywheel bearing support using dowel pins (possible without the flywheel).

Maintenance

Maintenance-free bearings were used.

Figure 7.161 shows the preliminary layout drawing of the test rig resulting from the embodiment steps discussed above.

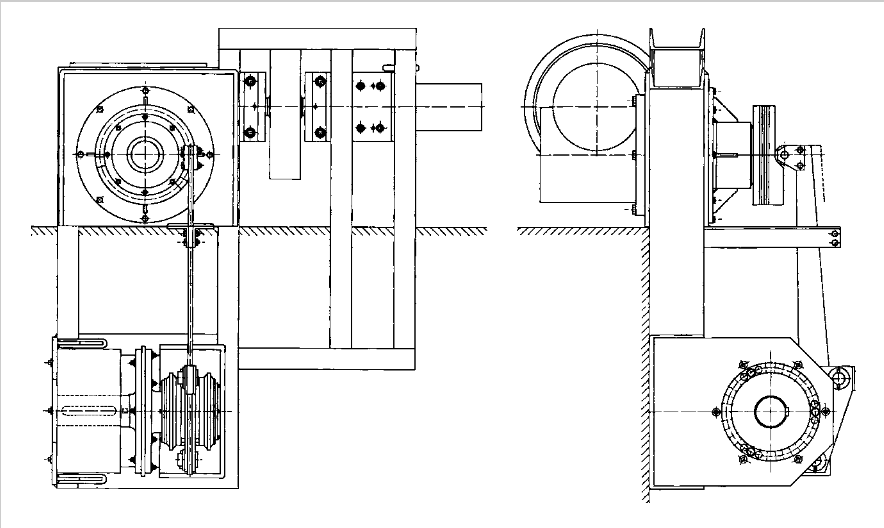


Figure 7.161. Preliminary layout drawing

Step 10: Evaluating Using Technical and Economic Criteria

Because only one final embodiment was developed, no selection was involved, only an assessment of the final embodiment based on criteria derived from the requirements list. The objective was to identify and eliminate weak spots.

The procedure involved the following steps in accordance with Section 3.3.2:

- identifying evaluation criteria
- assessing whether the parameters meet the evaluation criteria
- determining the overall rating
- searching for weak spots
- eliminating weak spots, if required.

For the evaluation we used 11 of the 13 criteria that were used to evaluate the concepts, see Figure 7.162. The use of weightings was not considered to be necessary.

The expected and calculated parameters of the test rig were evaluated against an ideal solution using a value range of 0–4, in line with VDI 2225. A more detailed evaluation did not seem worthwhile. The result is shown in Figure 7.162.

Only the technical rating was used in the calculation of the overall rating because there were no data for a formal assessment of the economic rating:

$$R = 29/44 = 0.66$$

This rating is rather low, so a search for weak spots seemed necessary. First, those parameters that had the lowest values were identified. A proposal was then made to improve those parameters that received only one or two marks:

Evaluation criteria			Parameters		Variant 4/3			Variant 4/3 impr.		
No.		Wt		Unit	Magn	Value	Weighted value	Magn	Value	Weighted value
1	Good reproducibility		Disturbing factors	–	low	4				
2				–						
3				–						
4	Tolerance of overloading		Overload reserve	%	10	3				
5	High level of safety		Danger of injury	–	average	2		see text	4	
6	Few possible operator errors		Possibilities of operator errors	–	high	1		see text	3	
7	Small number of components		No. of components	–	low	3				
8	Low complexity of components		Complexity of components	–	low	3				
9	Many standard and bought-out parts		Proportion of standards and bought-out comp.	–	high	4				
10	Simple assembly		Simplicity of assembly	–	high	3				
11	Easy change of load profile		Change of load profile	–	bad	1		see text	2	
12	Quick exchange of test connections		Estimated time needed to exchange test con.	–	average	2		see text	2	
13	Good accessibilitiy of measuring system		Accessibility of measuring system	–	good	3				
		$\Sigma W_i = 1.0$				$OV_1 = 29$ $R_1 = 0.66$			$OV_2 = 34$ $R_2 = 0.77$	

Figure 7.162. Evaluation chart for embodiment based on Figures 7.161, 6.54 and 6.55

- Few possible operator errors.

  Weak spot: motor speed: (1) the speed could be set at a value higher than necessary for the maximum rate of torque increase; and (2) the run-up of the motor should only take place slowly because of the heat generated.

  Remedy: the allowed range for run-up and operation can be marked on the speed indicator of the motor. The machine can be shut down automatically if the speed becomes too high.

- Easy to change the load profile.

  Weak spot: exchange of the cylindrical cam was not possible because of the clamping pressure of the lever on the cam.

  Remedy: provide a means to lift the lever.

- High level of safety.

  Weak spot: rotating cylindrical cam was not protected.

  Remedy: provide protective cover.

- Quick exchange of test specimens (test connections).

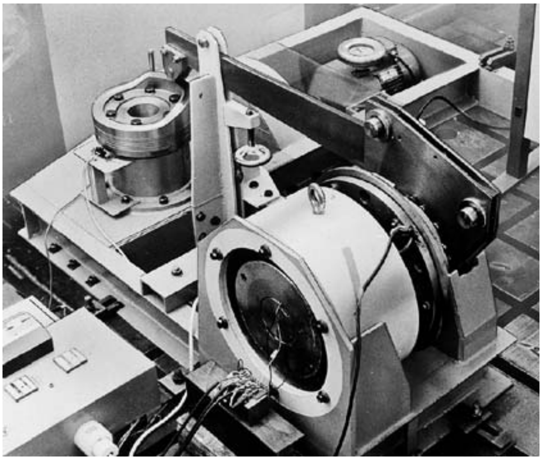


Figure 7.163. Final impulse-loading test rig, after [7.188]

Weak spot: slow because of the number of screws in the Ringfeder connection.

Remedy: no economic alternative possible.

The improved variant has been added to the evaluation chart (see Figure 7.162).

The remaining working steps used to *define the overall layout* proposed in Figure 7.1 are not discussed here. They were not very complex in the case of this test rig because it was a one-off product for a research institute and did not need a high degree of optimisation. The *detail design* of the test rig (following the working steps in Section 7.8) is also not discussed. It only involved conventional drawing and detail design steps.

Figure 7.163 shows the final impulse loading test rig. It fulfilled the main expectations and confirmed the effectiveness of a systematic approach [7.122].

7.8 Detail Design

Detail design is that part of the design process which completes the embodiment of technical products with final instructions about the shapes, forms, dimensions and surface properties of all individual components, the definitive selection of materials, and a final scrutiny of the production methods, operating procedures and costs.

Another—and perhaps the most important—aspect of the detail design phase is the elaboration of production documents, including detailed component drawings, assembly drawings, and appropriate parts lists. These activities are increasingly undertaken using CAD software. This allows the direct use of product data for production planning and the control of CNC machine tools.

Depending on the type of product and production schedule (one-off, small batch, mass production), the design department must also provide the production department with assembly instructions, transport documentation and quality

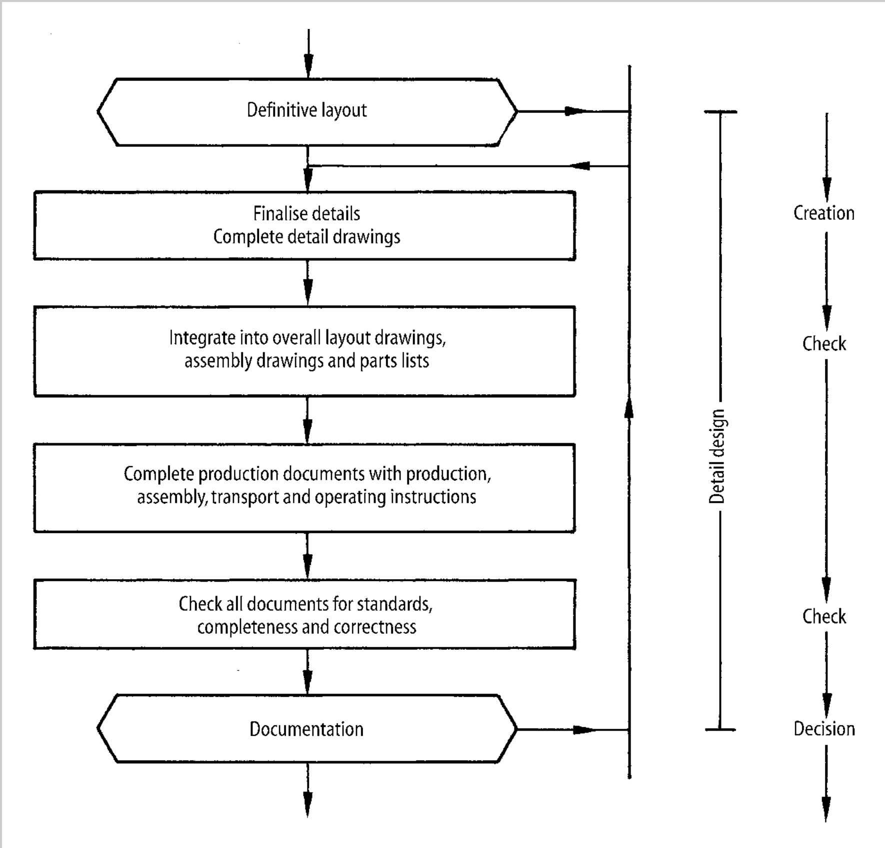


Figure 7.164. Steps of detail design

control measures (see Chapter 10), and the user with operating, maintenance and repair manuals. The documents drawn up at this stage are the basis for executing orders and for production scheduling, that is, for operations planning and control. In practice, the respective contributions of the design and production departments in this area may not be distinct.

The detail design phase involves the following steps (see Figure 7.164).

Finalise the definitive layout, comprising the detailed drawing of components, and the detailed optimisation of shapes, materials, surfaces, tolerances and fits. To that end, designers should refer to the guidelines given in Section 7.5. Optimisation aims at maximum utilisation of the most suitable materials (uniform strength), at cost-effectiveness and at ease of production, with due attention being paid to standards (including the use of standard parts and company repeat parts).

Integrate individual components into assemblies and into the overall product (fully documented with the help of drawings, parts lists and numbering systems). This is strongly influenced by production scheduling, delivery dates, and assembly and transport considerations.

Complete production documents with production, assembly, transport and operating instructions.

Check all documents, especially detail drawings and parts lists, for:

- observance of general and in-house standards
- accuracy of dimensions and tolerances
- other essential production data
- ease of acquisition, for instance, the availability of standard parts.

Whether such checks are made by the design department itself or by a separate standards department will depend largely on the organisational structure of the company concerned, and it plays a subordinate role in the actual execution of the task. The steps of the embodiment and detail design phases overlap in the same way as the steps of the conceptual and embodiment phases often do. Long lead-time parts, such as those involving forging and casting, should be dealt with first and their detail designs and production instructions are often completed before the definitive layout has been finalised. This overlapping of two design phases is particularly common in one-off production and in heavy engineering.

Detail design is very domain- and product-dependent and designers should refer to the many technical handbooks, suppliers catalogues and standards that deal with the detail design and selection of machine elements.

Corners must never be cut during the detail design phase, which has a critical effect on the technical functions, on the production processes and on the elimination of production errors. Detail design has a major influence on production costs and product quality, and hence the success of a product in the market.