Engineering: The nature of problems

About this free course

This free course is an adapted extract from the Open University course T207 *Engineering: mechanics, materials, design* http://www3.open.ac.uk/study/undergraduate/course/t207.htm.

This version of the content may include video, images and interactive content that may not be optimised for your device.

You can experience this free course as it was originally designed on OpenLearn, the home of free learning from The Open University:
http://www.open.edu/openlearn/science-maths-technology/engineering-and-technology/engineering/engineering-the-nature-problems/content-section-0.

There you'll also be able to track your progress via your activity record, which you can use to demonstrate your learning.

The Open University Walton Hall, Milton Keynes MK7 6AA

Head of Intellectual Property, The Open University

The Open University

Hobbs the Printers Limited, Brunel Road, Totton, Hampshire SO40 3WX

Contents

Introduction

The optimistic approach to a problem is to view it as a challenge and an opportunity – a chance to make progress. In this course, the nature of problems is explored by looking at the way they are used as a stimulus for finding solutions. It is presumed from the start that you want to be involved in the process of finding solutions and that you are not expecting simply to be given the answers.

One example that is investigated in this course concerns how to devise lighter bicycle frames, and the way to assess the merits of alternative materials from which to make them. There is no single way to move from a problem like this to possible solutions. In fact there are often several ways to set about finding several solutions, but there are a few general factors that are important to the search.

First it is important to appreciate the needs from which a problem arises. For the bicycle frame it's not just a lighter material that is required, but rather it is one that can be deployed to bear specific loads imposed on a fully functional frame.

Next it is valuable to understand the challenge well enough to be able to specify the nature of solutions, perhaps using the formal languages of engineering, mathematics, science and problem solving. For example, it is unwise to take part in a discussion on 'the best materials for bike frames' without a technical appreciation of both the job a frame has to do and the relevant attributes of the candidate materials. Establishing what you don't yet know usually starts by recognising how effectively you can tell someone else where the challenges arise. You must be able to communicate with a wide range of people, sometimes 'calling a spade a spade', and at other times describing precisely what the word 'spade' actually means.

In passing from a problem towards possible solutions it is essential to be able to evaluate and quantify the technical aspects. Another general factor in the search for solutions is the use of algebra and numbers to compare options and to inform choices. Some calculations are simple evaluations that can be done directly with or without an electronic calculator. Others need a line or two of algebraic analysis. Yet others are too tedious or too complicated to tackle without a computer-based approach using spreadsheets or more sophisticated software.

In the end, the best motivation for learning comes from simply requiring the knowledge in order to make progress.

This OpenLearn course provides a sample of level 2 study in Engineering.

Learning Outcomes

After studying this course, you should be able to:

- view solutions as belonging to particular categories, broadly classified as: innovation by context; innovation by practice; routine
- see how external factors affect engineering projects, and appreciate the range of engineering involved in meeting the basic needs of our society
- recognise and apply a range of problem-solving techniques from each stage of the engineering design cycle, to include the following: physical modelling; mathematical modelling; iteration; use of reference data; refining an engineering specification
- identify when models are likely to be useful and when they are no longer valid
- recognise and distinguish between the following technical terms: differential equation; simultaneous equation; boundary condition; constraint; finite element analysis (FEA); mathematical model; physical model; prototype; demonstrator; anthropometric; ergonomic; product specification; functional specification.

1 Problems and innovation

1.1 Solving problems

It could be said that our species is defined by its irresistible urge to solve problems – it's what makes us human. Strange, then, that the word 'problem' has such negative overtones. I think that the root of this paradox is that the word is used both when we identify a need – the first link in the problem-solving chain – and when we undertake the process of meeting that need. It is the identification of the need and the realisation that it is real and must be met that creates the anxiety and the negative feelings ('Houston, we have a problem …'). The process of finding a solution is the exhilarating part that makes us thirst for more.

I think of my love of skiing. Sometimes I get to the top of a mountain and look down at the precipitous slope I must now descend to get back to safety and a good hot meal, and I am gripped by fear, perhaps even to the extent of wondering whether I'll survive this time. What provides the pleasure, apart from the thrill of speed, is using my skill (such as it is) to meet my need to be safe again.

'The Engineer: Skier on the Technological Piste' is perhaps too bizarre a title for a course, so welcome instead to T207_1 *The engineer as a problem-solver: the nature of problems*.

T207_1 is taken from an Open University course entitled '*The Engineer as Problem Solver*'. The fact that we have prepared a course with this title shows that we think there is something useful to say about the process of solving engineering problems. It seems to imply that there is a technique to be learned – a preferred method. To a degree, this is true; experience has taught us that there are certain ways of proceeding which tend to lead to better solutions than others. This course is designed to give you a flovour of the skills and knowledge that will help you to make active and informed decisions when tackling your own engineering problems.

In engineering, solutions to problems come in three categories:

- innovation by context;
- innovation by development;
- routine.

This is going to need some explanation, so here goes.

The categories differ from one another in the extent to which the solution is a step into the unknown, and this is why it may be chosen before the solution itself is known.

Table 1 Old context/new context versus old technology/new technology

	Old technology	New technology
New context	innovation	invention
Old context	routine solution	innovation

As you can see, there are four sectors in the table, defined by the technological newness of the solution, and the newness of the context in which it is to be applied.

The 'customer' for the solution will often have a very definite idea of the sector in which they wish the engineer to operate. For example, a new heart pacemaker will be heavily constrained in many respects: only certain materials will have been approved for the casing, as they have to be biocompatible; certain safety features must be included, such as methods of making sure the casing is hermetically sealed, making the device immune to electromagnetic interference, limiting the power and frequency of the heart stimulation; as well as other limits such as on the minimum lifetime of the battery and the need to provide sufficient warning of its decline to allow it to be replaced in time. The list is much longer than this. The effect is to discourage excessive innovation (by which I mean a significant change to the way something is made or the way it works – or a new type of thing entirely) and the chances are that an innovation by development will be the order of the day.

We define innovation by development as changing the bit that doesn't work, or that could work better, to improve the function of the whole product or design for reasons of cost, performance, ease of manufacture or gaining competitive edge.

This tight constraining of innovation does not preclude entirely the invention of a new type of pacemaker, but it is unlikely that a manufacturer will be asking its engineers to throw away the rule book and dream up something new. If an innovative pacemaker is to appear, it will be because someone has had a sudden inspiration, and the idea is so very good and the potential benefits so great that a manufacturer is prepared to take a large risk and go through years of testing to gain safety approval.

It is usually where safety is a critical factor that we find a tendency to reduce the amount of innovation, so it would be equally easy to find an example from the oil industry, or from the aerospace or military fields. Very conservative purchasers are the other main reason for holding back on innovation. Perhaps surprisingly, the industrial process control market is one such area. Engineers responsible for the design and installation of processing plant usually take a lot of persuading that a new type of flowmeter, say, will be better than the one they know and have been using for many years.

At the other extreme, there are just as many examples where innovation is essential to the success of a project. There's even a well-known mail-order catalogue whose very name includes the word. The market for gadgets and gizmos is huge, and rather prone to the vagaries of fashion. These two characteristics make it a powerful driver for innovation. Last year's temperature-indicating tea cosy with built-in radio, flashlight and satellite navigation system may have sold a million, but it's a little passé now (Figure 1).

Figure 1

Another important instance where we know in advance that an innovation is required is where the existing technology is very mature, and has been incrementally developed as far as it is possible to take it, yet we have identified a need to improve the performance of the product still further. In this case, we can clearly see that innovation will be the only way to get there. An example of this is the bubble-jet printer. In the late 1970s and early 1980s, the vast consumer and small-business market for printers could not stand the cost of laser printers, yet was demanding better quality, less noise and higher speed from the alternative, which was the dot-matrix printer. This works by transferring ink from a typewriter ribbon onto the paper, using an array of electromagnetically actuated pins. These had been improved over many years, and were about as good as they could get. They had reached the limits of how fine a pitch they could be arranged on at a reasonable price but, even so, they were still rather slow and noisy.

What was required was an innovation, and this came in the form of the bubble-jet, which was an entirely new technology based on the ability to etch arrays of very fine holes in a polymer substrate using photolithography. The ink is transferred to the paper by creating pressure pulses behind the appropriate hole, using electrical heating. A single drop of ink is then ejected at high speed and strikes the paper. At a stroke, the new technology provided marked improvements on speed, pitch and noise level, and all without increasing the cost of the product.

1.2 Innovation by context

The word 'innovate' simply means 'make new'. We have chosen in this course to narrow the meaning of this term to be more or less synonymous with 'invention'. I would argue that innovation by context is as much a process as a result. By that, I'm using the term to mean something more like 'creativity'; and it's creativity that lies at the heart of all engineering. More than anything else in our professional lives, we engineers are excited by the prospect of being responsible for the creation of something better than we had before. This does not mean that all engineers are inventors in the sense that the word is normally understood (i.e. taking out patents on some new gadget). Innovation by context is the fruit of the creative process going on in the mind of the engineer when solving a problem, and can be anything from a clever change in the design of a computer program that allows it to run faster or use less memory, to something revolutionary like the jet engine.

For an engineer, creativity is a daily activity. Sometimes, the result is a big enough innovation to call an invention, and to patent, but mostly it's just small but necessary steps in reaching the goal. You can be creative even if your solution is of the type we classify as routine, as we'll show later. There are numerous great examples of innovative solutions, some of which are inventions. How about the first aqueduct, canal, drain, concrete, pilotless aircraft, the building of the Eden Project in Cornwall, the new roof over the British Museum, radar (mapping the skies, speed cameras and predicting weather), the first mobile phone, waterproof fabric, the microwave oven, the compact disk, Dyson's bag-less vacuum cleaner? … the list goes on. It's difficult to pick one example for looking at in more depth, but one of the best and most simple examples of innovation by context has been the transformation of radio into a self-sufficient technology, described in Box 1.

Box 1 Innovation by context – an example

In 1991, inventor Trevor Baylis saw a television programme about AIDS workers in Africa. In poor countries radio broadcasting had always played a part in health education, but in this programme the workers were explaining how batteries were expensive or unavailable and electricity supplies unreliable or simply non-existent. The programme provided Baylis with a problem, and inspired him to find an innovative solution.

Baylis's invention, as you have probably guessed, was the clockwork radio, Figure 2. He wasn't the first person to use springs to generate electricity, but prior to his design the energy had only ever been produced this way for short bursts at a time – here is the *context*. The *innovation* is in applying springs to the provision of low-power electricity for consumer electronics. Baylis invented a mechanism that gave forty minutes of play from just twenty seconds of winding. The winding action coils a spring, attached to a gearbox, which is connected to a dynamo. When the spring is released the gearbox controls the steady discharge of energy to produce electricity, and the radio works. The dynamo provides three volts at between 55 and 60 milliwatts, but the design also incorporates a solar-powered source to extend its performance.

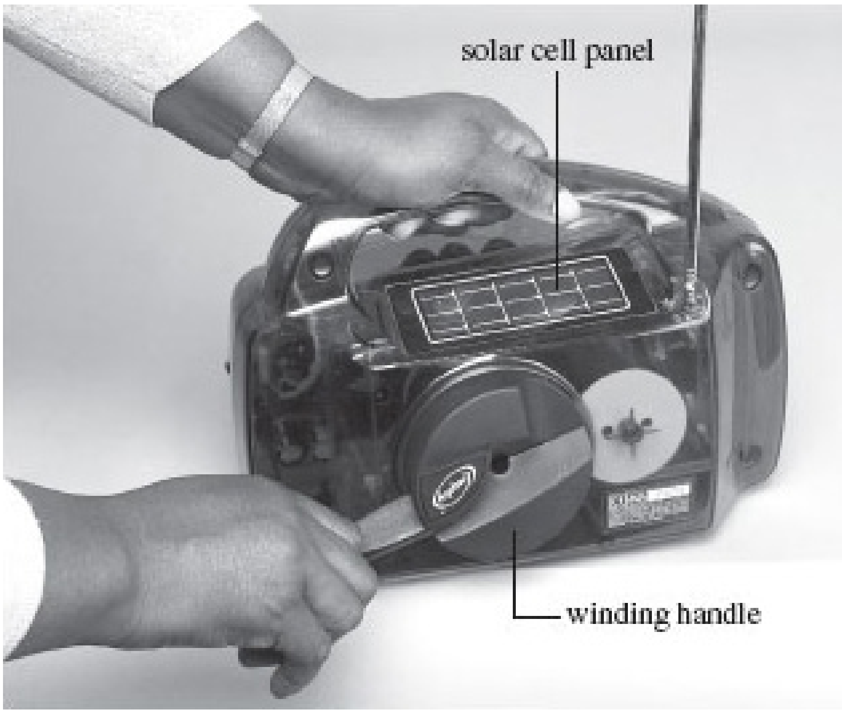

Figure 2 Freeplay wind-up radio is powered using a wind-up handle at the rear and augmented by solar power

Thus power is generated from human input, backed up by free and widely available solar power. Baylis realised that his new technology had huge potential. However, this was only half the task: he also needed to reduce the amount of power consumed by the radio, and this is where the less glamorous and less visible (but at least as important) part of the innovation was done. A team of electrical engineers worked to make improvements in small increments until the power consumption was pared down enough to allow the radio to work for a reasonable amount of time between rewindings of the spring.

Reaction to the radio was initially somewhat sceptical but eventually, in 1994, a prime-time BBC TV programme (*Tomorrow's World*) agreed to feature the idea. Two entrepreneurs who happened to be watching contacted Baylis immediately, going on to form a company with the intention of putting the clockwork radio into production. 'Freeplay', as they were called, raised a government grant for initial development costs and then found investors, to date selling over three million units. The company has the endorsement of heads of state, international aid organisations, royalty, celebrities, the European Union, the United Nations and more, and has gone on to develop other similar products.

The original problem – that of providing a self-sufficient technology so that radios could be widely available in developing countries – has been solved not only in theory but also in practice, the true test of an innovation.

1.3 Innovation by development

Innovation by development is about changing the bit that doesn't work, or that could work better, to improve the function of the whole for reasons of cost, performance, ease of manufacture or competitive edge. You probably noticed in Box 1 'Innovation by context – an example' that Baylis had to incorporate a number of developmental innovations as well. Improvements in materials or production equipment or techniques can present solutions to manufacturing difficulties, and so development becomes incremental not only in a product, but in a chain of production.

It's where the big money lies for companies wanting to keep one step ahead of their competitors without the (in general) higher risks and longer timescales of innovation – 'our powder washes brighter', 'this battery lasts longer' or 'this car is quieter to drive'.

Most technological items in everyday use have been subject to innovation by development. You can see the results in the motor industry, in aeroplanes, trains, mobile phones, computers, fridges, cookers, plastics, household implements … it's more difficult to think of something that hasn't been subject to innovation by development! Box 2 Innovation by development – an example explores a typical product development.

Box 2 Innovation by development – an example

The Black & Decker Workmate portable workbench (Figure 3) has been the DIY and professional craftsman's best friend for thirty years now. At first sight, it is a product that has not changed much at all in that time; but if you look more closely, you will see that it has undergone a considerable amount of innovation by development. However, this has been done over such a long period of time that you need to put the original next to a new one to spot the differences. We can look at some of these changes and try to guess why they were made.

Figure 3 A modern Black & Decker Workmate

The original Workmate had a pair of long leadscrews (Figure 4a), one on each side, to move the jaws of the vice-cum-work surface. The length of these screws was over 30 cm, and they moved the half of the table furthest from the user relative to the other half, which was static. This enabled the bench to offer one of its best selling points: the ability of the vice to accept a large range of workpieces, including wedge-shaped and very wide ones.

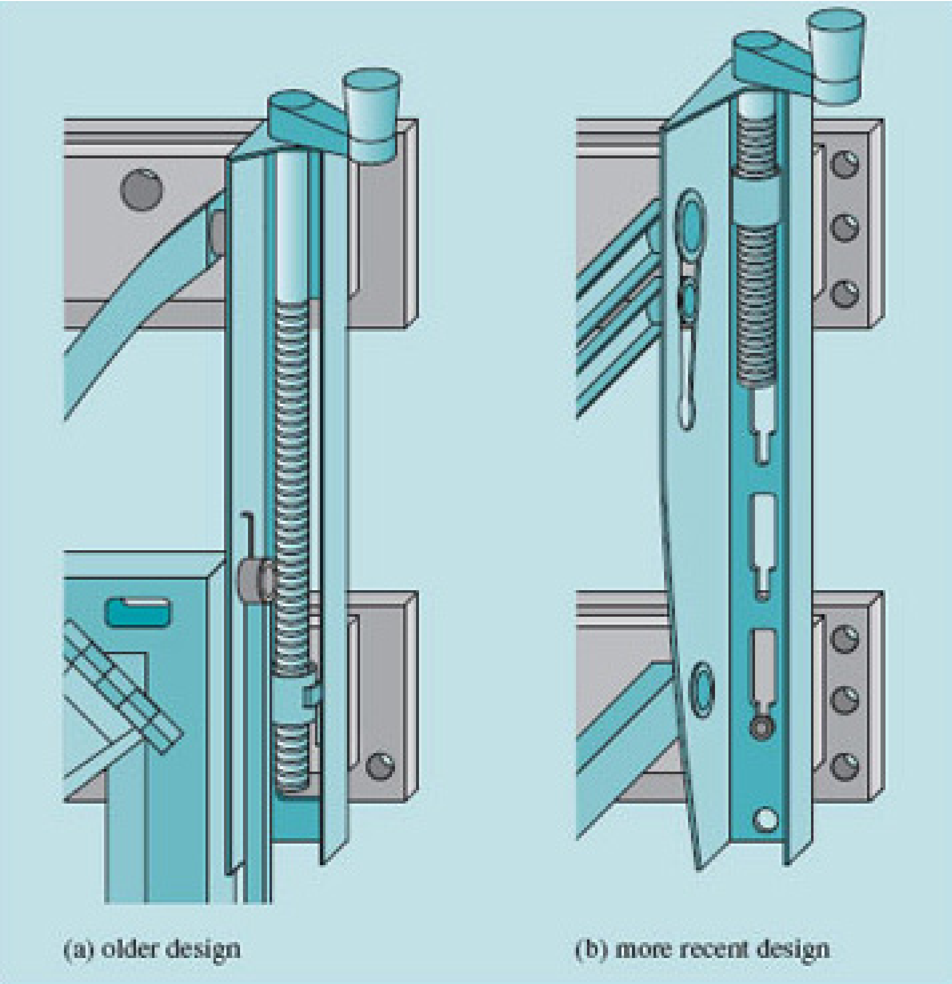

Figure 4 Comparison of old and new designs of Workmate vice, seen from the underside of the work surface

The current design has switched things around, so that it is now the part of the top nearest the user that moves, but the leadscrews are very much shorter and therefore lighter (Figure 4b). This would normally restrict the range of widths of workpiece that can be accepted, but this has been restored by making the other part of the top removable. To accept pieces with a width that is outside the range of the leadscrews, this part of the top is simply unlocked from the framework and moved into one of the alternative locations for it, providing a new range of jaw separations that overlap slightly with the previous position.

Exercise 1

Make a list of the effects of the change to the design of the Workmate's vice mechanism, noting the beneficial and detrimental effects for each one from the manufacturer's and the user's point of view.

Why do you think this change has been introduced?

Answer

- *Cost saving on leadscrews:*

  better margin for manufacturer, lower price for user.
- *Reduction in weight:*

reduced distribution cost for manufacturer, easier to carry for user.

- *Screw does not now pass under the opening of the vice:*

  no direct effect for manufacturer, reduced likelihood of clogging with sawing debris (and therefore excessive friction) for user.
- *Extra operation of removing and replacing the moving jaw required for changes in workpiece width that exceed the leadscrew travel:*

  no direct effect for manufacturer, less convenient for user for small changes in workpiece width that nevertheless require relocation of the moving jaw, though quicker to change jaw separation for large changes in workpiece width.

The effects on both the user and manufacturer of all but the first change are slight. Therefore, the likely reason for the change is the first one – cost saving. This is partly offset by a cost to the manufacturer in making the change – the cost of the design work, changes to drawings, parts lists and order schedules. Note that the cost reduction has a beneficial effect for both the user and the manufacturer.

1.4 Routine solutions

This is the last of our three categories, and possibly the most difficult to define because the approach is not as definite. Routine solutions involve configuration or reconfiguration of existing devices or components, without innovation, because something is broken or needs to be repositioned, or there is simply a better way to do it. If you change the locks in your house or car, you are reconfiguring them; if you tune the car, calibrate the central heating, set the coordinates for your satellite navigation system, change from an overhead lamp to a wall light, or even just change station on your television, you are applying a routine solution to a problem by reconfiguring the bits. As I write, I'm reminded of the ongoing attempt by a group of stalwarts to reconfigure the standard keyboard, originally designed to prevent the letter levers clashing on a manual typewriter, into something more user-friendly for today's computer user.

The biggest examples of challenges requiring routine solutions are, literally, physically big. Things that need configuring are often remote, such as a fibre-optic signal booster in a cable at the bottom of the Atlantic, or, at the other extreme of the planet, a satellite; both of which (as it happens) are critical to intercontinental telecommunication. Box 3 Routine solution – examples looks at some examples.

Box 3 Routine solution – examples

The Hubble space telescope (Figure 5) was conceived in the 1970s. The intention was that it would capture astronomical images, unimpeded by the Earth's atmosphere, and transmit data and images 640 km back to Earth, enabling us to answer some of our most fundamental questions about the universe. It was sent into orbit in April 1990, at a cost of about US$ 2 billion.

However, just weeks into its flight the mission was very nearly lost before it had truly begun, when NASA scientists discovered that the main concave mirror of the telescope had been ground too flat by a depth of 4 micrometres, resulting in images at high magnification that were too fuzzy to be useful.

The operators who control Hubble's flight work in team rotation, driving it 24 hours a day every day of the year, sending an average of over 100 000 instructions a week. The first opportunity to carry out maintenance, install new instruments and correct the error (by giving it 'glasses' in the shape of five pairs of corrective mirrors) came in December 1993, after two years of planning. Engineers operating the telescope trained extensively for the reconfiguration of Hubble. First the telescope had to be set aside from its usual research operations to a 'ready for servicing' condition and capture attitude, then the aperture door was closed and high-gain antennas stowed. Astronauts on board a Space Shuttle made five gruelling space-walks to carry out the installation work. Once this was completed and tested, both Hubble and the Shuttle were configured for battery charging. When charged, everything on the telescope was reactivated and it was released back into orbit. To everyone's very great relief the mission was a success, and Hubble soon began transmitting the great pictures that had been anticipated.

Why do we describe this as 'routine'? Clearly the solution being sought was not expected to be innovative – the commitment to reflective optics was unchangeable. Similarly, the cost of a series of incremental improvements would be prohibitive. What was called for, and what was done, was routine reconfiguration of the bits.

A less glamorous example is found in electronic circuit design. New amplifiers, data acquisition cards and so on are launched every year.

Many are new arrangements of standard components – resistors, capacitors, integrated circuits, etc. The problem solving here has been concerned with choosing component values and characteristics to achieve enhanced performance.

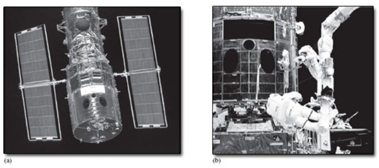

Figure 5 (a) the Hubble orbiting telescope, (b) NASA astronauts on a Shuttle mission to repair and service the telescope in orbit

You should, by now, have a better idea of how to classify solutions to problems, challenges and opportunities. The three groups above overlap. It's possible for a solution to be equally valid in more than one group at a time. It's important to consider the context of whatever you're facing – the invention of the mobile phone was an innovation in terms of electronics, subsequent innovation by development has been largely incremental, and during this development there have been considerable routine design changes. If you're deciding where the solution to a problem belongs, try to narrow it down to its basic elements.

SAQ 1

Group the following tasks as being problems likely to find solutions that are *routine* in nature, that involve *innovation by development* or require *innovation by context*:

Making a lighter ladder

Specifying components for a home-computing workstation

Defining specifications for building services in a new factory, e.g. ambient temperatures in different rooms/areas, air conditioning, waste air extraction, etc.

Designing a taller crane

Replacing lead-based solders with non-lead alternatives

Bridging a wider gap

Setting network and modem parameters for an office PC system

Designing an ejector seat for a helicopter (ouch!).

Answer

Routine:

Specifying components for a home-computing workstation (selecting from among existing components)

Setting network and modem parameters for an office PC system (selecting from existing options)

Defining specifications for building services in a new factory, e.g. ambient temperatures in different rooms/areas, air conditioning, waste air extraction, etc. (selecting from among existing components)

Innovation by development:

Designing a taller crane (extend a shorter crane)

Making a lighter ladder (refine the design to reduce weight)

Bridging a wider gap (extend an old design)

Replacing lead-based solders (devise new alloys)

Innovation by context:

Designing an ejector seat for a helicopter (ejector seats were conceived for fixed-wing aircraft and can't simply be transferred to the pilot's seat in a helicopter).

2 Where does the need arise?

There is a rather obvious question that has to be raised at some point, so we may as well get it over with now: Why do we present ourselves with all these problems? After all, life would be easier without them and we could all go off and do jobs that don't involve them. Do we really need to know everything about the universe? Or to send people into space, at significant cost and human risk? Do we really need to send sound and pictures through space? Do we really need to communicate with people we've never met? Do we really need to educate people about health?

I hope you have at least agreed with the last one, and you can probably see a connection that runs through the points that were used as evidence in the last section. What it illustrates is an order of priority of human needs, ranging from the immediate and essential, to the remote and desirable, and that engineers are active at every level.

Exercise 2

Arrange the following items in order of human physical need, with the most basic requirement at the top:

Communication
Oxygen
Transport
Challenge
Shelter
Entertainment
Water
Energy
Warmth
Education
Food.

Answer

Oxygen
Water
Food
Shelter
Warmth
Transport
Communication
Energy
Education
Challenge
Entertainment

This is only my list, and your own personality will probably dictate how you placed the bottom six. The point is that we can survive no more than a few minutes without oxygen, a day or two without water, and a week or two without food. In extreme environments we can't survive without shelter and/or warmth. As for the rest: well, on this particular scale they can be seen as life's luxuries, although in relatively rich societies we are expecting more and more as our right rather than privilege. Engineers are involved in meeting all these needs at every level and at every depth of complexity.

However you organised the above list, you can see that there is a hierarchy of human requirement where the needs become increasingly refined and complex, and that there are problems, challenges and opportunities for the engineer at every level. All the items in the list could be expanded to consider the engineering involved. Box 4 Meeting the liquid challenge looks at how we meet the fundamental need for a supply of clean water.

Box 4 Meeting the liquid challenges

To all practical intents and purposes, water on Earth is part of a closed system – there is no more or less water on the Earth's surface now than when the first humans were alive. It is approximately 1400 million cubic kilometres of the ultimate recyclable resource, and it is random in its availability. We use it to drink, cook, wash and flush sewers, and without it any one of us would die within a week. Apart from the very air we breathe, it has to be our most basic need.

In temperate zones in the northern hemisphere, we are lucky enough to get a reliable amount of rainfall, which we can store in artificial reservoirs. Water has to be collected from lakes and reservoirs, wells, rivers and underground pools, then treated and transported for domestic and commercial use on a mammoth scale (Figure 6).

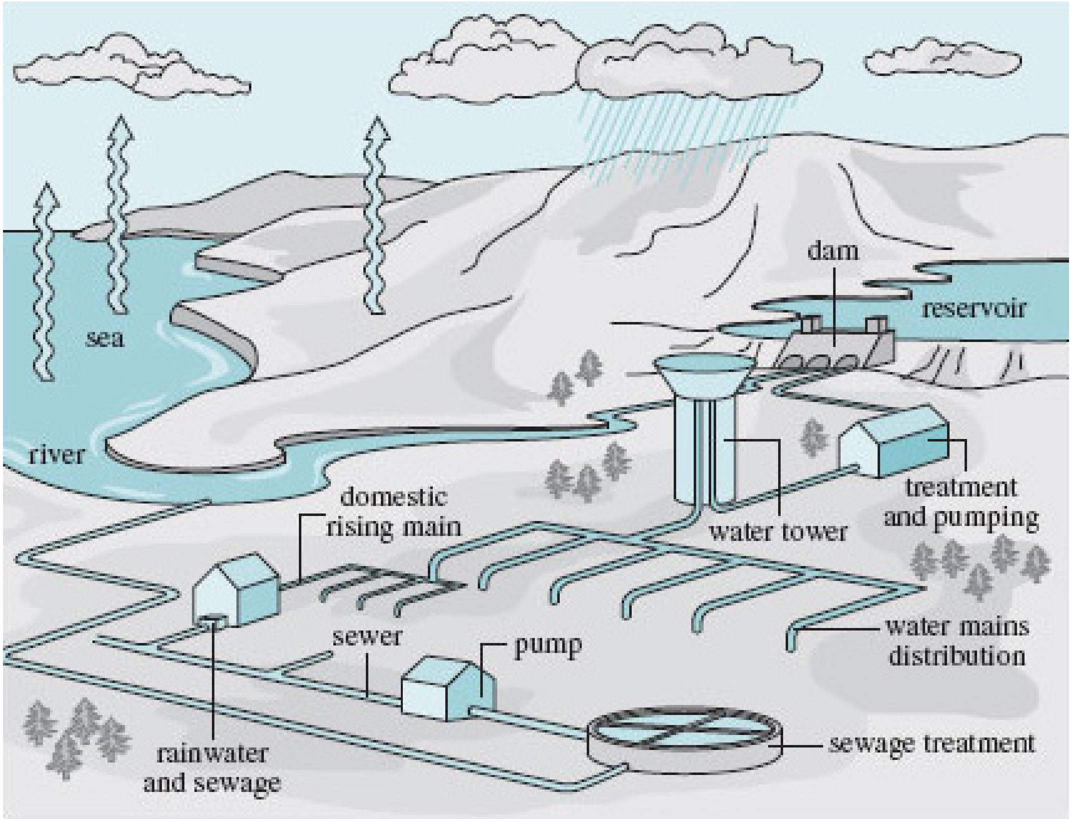

Figure 6 The water cycle

Think about the engineering involved in designing, building and lining reservoirs; designing and laying pipes of the right material and capacity; controlling water flow through the pipes; filtration and purification; delivery to domestic and business premises; removing, storing and treating sewage; managing the logistics of supply and demand; and the financial and technical administration of the water system. We may have cause to grumble about the occasional shortage in supply during long dry summers, but our system is generally robust. Generations of engineers have been responsible for the development of reliable water provision around the world (though there are still places where the challenges remain). If you have ever visited a country where you had to rely on sterilised or boiled water, you will appreciate it all the more.

In many poor countries the history of problems caused by drought or contaminated water is well documented. It is currently estimated that 2.4 billion people worldwide lack access to basic sanitation, and over a billion are drinking unsafe water. The engineering challenges posed in these countries (mainly in Eastern and Southern Africa, and South Asia) are different from those met in most of Europe as the rainfall is less reliable, work is often funded by overstretched charities and, although a long-term infrastructure is needed, there is also an urgent necessity to provide instant clean water. Engineers are working on a small, local scale, sometimes having to show innovation with the materials and resources available and meet needs by practicality at the expense of efficiency. They might have to find water below ground, or find a means of purifying water from a river. Engineers may also find themselves becoming educators – passing on their skills and advice to local communities who can then carry out work for themselves.

Internationally, as in any industry, there are groups of engineers and scientists committed to research and development at the boundaries of our existing knowledge. The most recent high-profile discoveries in the water industry are to do with the desalination of sea water, a huge and largely untapped aqueous resource.

Many sectors of engineering are involved in meeting such a basic need as the supply of fresh water. The three classifications of solution – innovation by context, innovation by development and routine – are all represented many times over, and there are numerous angles of opportunity and challenge. If you consider a similar breakdown for each of the needs you listed in Exercise 2, you begin to get some idea of how and where engineering solutions are required. Here's a summary for this case:

- designing, building and lining reservoirs;
- designing and laying pipes of the right material and capacity;
- controlling water flow through the pipes;
- filtration and purification;
- delivery to domestic and business premises;
- removing, storing and treating sewage;
- managing the logistics of supply and demand;
- the financial and technical administration of the water system.

3 Needs and problems

The last section has established that engineering is about satisfying needs. In fact, with so many needs, it's a wonder that not everyone is an engineer! So, now that we have talked about both needs and problems, the logical progression is to examine the relationship between them.

Take the water example as being a fundamental need. We can state it thus:

> *This village* needs *a supply of clean water.*

When given that statement, we have a natural tendency immediately to start looking for potential solutions – a trough for rainwater, purification for river water, a pump for underground water and so on. We will start asking questions to get a clearer definition of the need – What's the average rainfall? Is the village near a river? Do we know of any existing supplies? What physical resources are available? How much water is needed daily – is ten litres each enough? What do we mean by 'clean'? etc. Seamlessly, the need becomes a problem that requires a solution. The definition of requirements makes it precise.

The problem becomes how to transfer and purify sufficient water from a source, say a river, half a kilometre away.

With this amount of detail we have a problem definition, and all that's left is to find a solution …

SAQ 2

(a) State, in a few words, the need which prompted the development of the Baylis wind-up radio.

(b) Make a list of bullet points that identify the engineering requirements involved in meeting the need for communication, like the list at the end of Box 4 'Meeting the liquid challenge'.

Answer

(a) The need for reliable, affordable access to broadcast health information in remote areas.

(b) adequate radio reception equipment

adequate power provision (clockwork and solar)

adequate manufacture and assembly

shipping and distribution

robust business plan.

4 Looking for solutions

4.1 Advancing knowledge

Over the centuries, engineers have faced and solved a huge number of problems of one sort or another. Each time a problem is solved, knowledge is advanced, something usually gets written down, and so today we have a wealth of experience to draw on. Equally, problem-solving *techniques* have also been developed and evolved through use and refinement, which is rather handy. Not only do we have some idea of existing solutions to similar problems, but we also have an indication of how to go about finding our own solutions.

As we're trying to get a picture of the whole, let's begin by looking at a typical, simple, problem-solution process and then we can break it down into separate elements. Figure 7 is one attempt to map out such a process, from the top down.

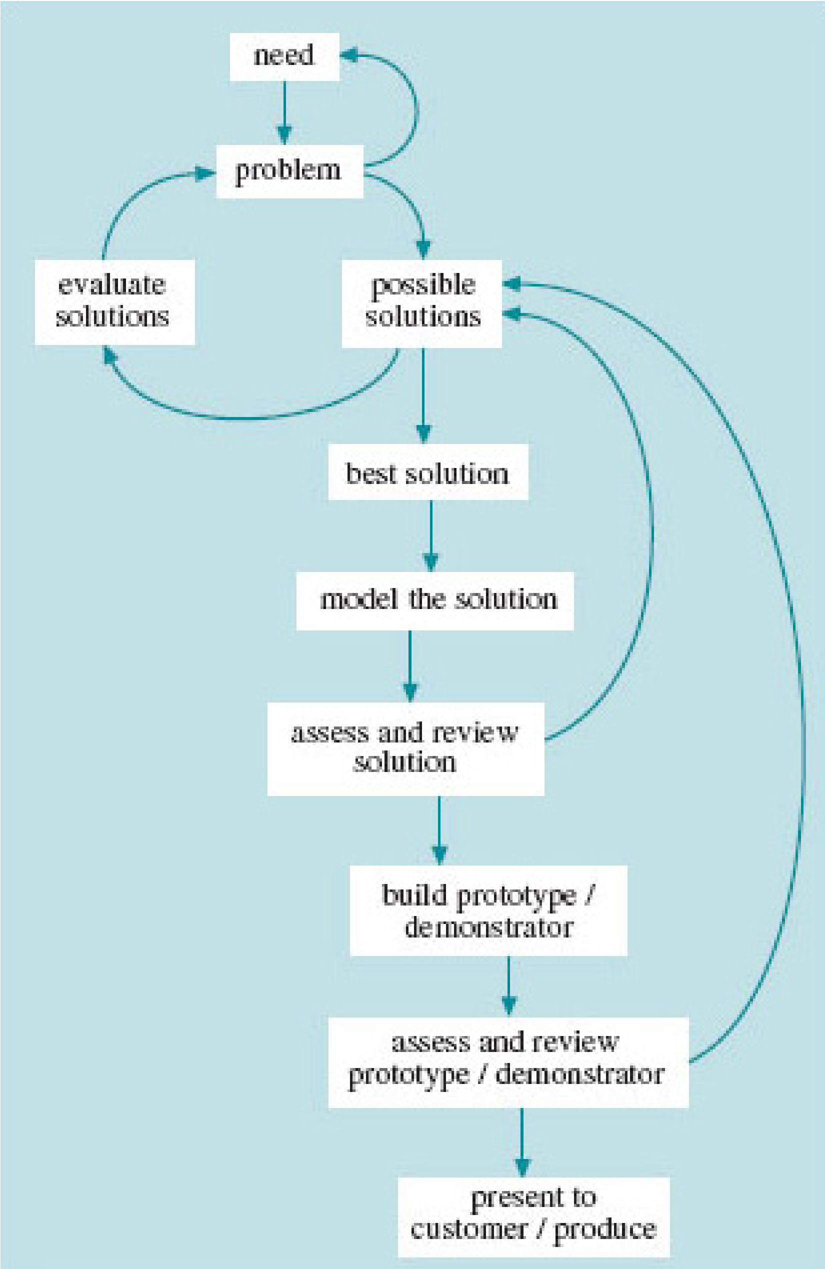

Figure 7 A 'map' or process for solving some problems

I should add, however, that there is no single right way to do this and there are, inevitably, all sorts of diagrams available to illustrate the process of creating solutions to problems. Engineering is a huge field, and procedures are usually shown with a bit more detail than in Figure 7 because they are specific to, say, software design, mechanical, chemical, civil engineering, etc.

4.2 From a need to a problem

So, working from the top down, the process starts with 'need' and 'problem'; see Figure 8.

Although we usually work by identifying a need that converts to a problem, that requires a solution, don't forget the extra arrow at the side, taking this first part of the process full circle. The questions that draw out the problem may also refine needs, or indeed extract further needs that were not stated, acknowledged or recognised at the very beginning.

We've already looked at where these needs come from on a global scale but, unless you are an academic researcher or a totally independent inventor, by the time you reach this stage of design the need is usually coming very directly from a customer. The customer may be your employer, or an external client, who has somehow identified the need to develop a new product or significantly modify, improve or repair an existing one. It's obvious that the better the specification, the less time and money will be wasted in designing or producing a product or solution that doesn't meet the requirement.

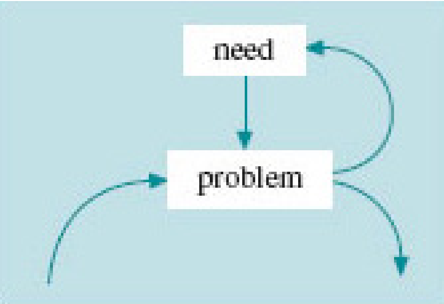

Figure 8 Top part of Figure 7

There are different types of specification – for example, a 'product' specification and a 'functional' specification. If the supplier has more knowledge about the specific product than the customer then a functional specification is appropriate. However, if the supplier is just making a design to order then a full product description must be agreed. The writing of these formal documents that attempt to ensure that the solution matches the requirements has become something of an art. The specification may become a legal contract that binds the engineer to the task, instead of a practical guide to the route and therefore the solution to a problem. While nobody would argue that you should not have some sort of guarantee that you're going to get what you are paying for, it does seem to be a shame that the more control is exerted in this way, the less room there will be for creativity and hence innovation in devising solutions. Used properly, the specification can be arrived at by an open exchange of views and ideas between the two (or more) parties involved, so that the engineering team goes away to look for solutions with a clear record and understanding of both the need and the problem. An example of the process that leads to a specification can be seen in Box 5 From problem to specification.

Box 5 From problem to specification

Between the 1930s and 1980s, millions of industrial and domestic refrigerators and freezers were produced which used chlorofluorocarbon (CFC) gases as the refrigerant and in insulating materials. CFC gases didn't degrade the fridge, were non-flammable, not poisonous in the event of a leak, and seemed to be an ideal replacement for the original refrigerants such as ammonia that were smelly, corrosive, poisonous and not particularly efficient. However, it transpired that CFC gases are damaging to the environment, depleting the ozone layer that protects us from the harmful ultraviolet components of the sun's rays. A need was thus identified – for fridges and freezers that are environmentally friendly (Figure 9).

Figure 9 Piles of discarded fridges and freezers awaiting recycling

It is worth noting that this statement of needs is not at as fundamental a level as the earlier one in the water example. This illustrates the existence of the hierarchy of need. We need the ability to create cooled environments not only for keeping food fresh, but also for countless industrial processes. We are able to state the need in terms of fridges and freezers because there is a long history of market requirement and product development that moves the starting point of our need statement on from 'we need to keep our food fresh', through 'we need to make things cold' to 'we need fridges and freezers'.

There is a general principle here in the formulation of statements of need: the more fundamental the terms in which it is written, the greater the variety of solutions open for consideration, but the greater the possible number of dead ends. There is therefore a balance to be struck between maximising the chances of a really creative solution, and wasting time considering unsuitable ones. So, the statement 'the village needs a supply of clean water' leaves more options open than 'the village needs a water pump and filtration to get clean water from the river'.

To take this statement of the need the next step forward, we have to write a problem specification, which ideally will contain all the information necessary for working out a set of possible solutions.

Our need was stated as 'environmentally friendly fridges and freezers'. We can further refine this to 'we need alternative refrigerant gases to CFCs'. In doing this, we have excluded the possibility of using an alternative technology to the compression/expansion heat pump that is ubiquitous in refrigerators and freezers. There is a very good reason for this: the closed-cycle mechanical heat pump is the most energy-efficient known means of refrigeration, and to go to something that uses more energy could add an unacceptable environmental cost.

We can now state the requirements for the solution: A refrigerant gas with the following properties:

- Not an ozone depleter
- Compatible with conventional heat-pump technology

- Non-corrosive
- Non-hazardous (i.e. non-toxic, non-flammable)
- Not an unacceptable source of some other pollution
- Able to be manufactured in comparable volumes and costs to CFCs.

This is a technical specification. Clearly, it is at what could be called the top level; there are no numbers against any of these requirements. Once these have been added, though, we will have the beginnings of a formal document to bind the engineer to the task. The specification may even specifically exclude certain types of solution (as with the Hubble telescope repair).

SAQ 3

There is a need to reduce the amount of pollution from airborne particulates in cities all over the world. A major factor is exhaust emissions from diesel engines. The search is on for an alternative fuel that doesn't produce the pollution locally.

Write a specification (as a list of bullet points) for an alternative automotive fuel that sets out the problem and is clear about the requirements that must be met in the solution.

Answer

The problem is to find or manufacture an alternative fuel for vehicles, because the existing fuels cause too much pollution in the towns and cities where they are used most intensively. The characteristics of the new fuel should include the following:

- a significantly lower producer of particulate emissions
- not a significantly worse producer of any other pollutants (including CO_2, which would result if it is less energy efficient to produce and use)
- no more toxic in unburnt form than existing fuels
- no more hazardous than diesel
- approximately the same cost to produce and use as existing fuels and possible to produce in similar volume (in terms not of litres, but of vehicle kilometres) to existing fuels
- preferably compatible with existing internal combustion engines, i.e. no solutions that are 'innovation in context' (though not necessarily if a longer-term solution is wanted – vehicle lifetimes are relatively short, so a new technology could be brought in).

4.3 Possible solutions

According to Figure 7, our map of the problem-solving process, once we've defined the problem according to the need the next step is the creative bit – to look for 'possible solutions', Figure 10.

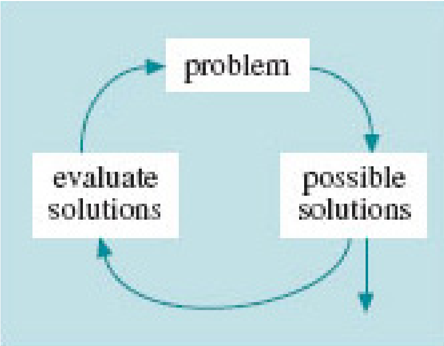

Figure 10 A part of Figure 7

Depending on the need, this may require innovation by context, innovation by development or a routine solution. Contrary to what you might expect, innovation is not the only interesting or challenging option – there may be any number of potential solutions using standard parts, but only one really elegant combination. As good engineers, in an ideal context (remember this point), we are not just looking for a solution; we are looking for the *best* solution. However, although practising engineers will be looking for the best solution, they do not always have the time or resources actually to reach it. Even if they did have as many resources as they wished, it might still not be possible to know what the best solution is, or whether it has been reached. So if the engineer cannot know when 'best' has been reached, we see that compromise is an abiding characteristic of solving engineering problems.

Creative thought has to come without inhibition, influence or bias, and if you consider how difficult that is in the light of the problem in SAQ 3, then it makes sense that we shouldn't just expect it to happen. We can exercise our brain in much the same way as we exercise our bodies, and we can sharpen particular abilities by repeated action of the same or similar process. There are whole courses available in finding creative solutions, working up from simple questions to complex theoretical posers. The more you practise 'thinking outside the box', the better you will become. Foster, in a book called *How to Get Ideas*, has a good summary:

> Think laterally. Think visually. Play "What if?" Look for analogues. Look for things to combine. Ask yourself what assumptions you're making, what rules you're following. Screw up your courage and attack.
>
> Foster (1996) p. 159

This is fine for working alone and stimulating your own creativity with no one around to question your ideas. In most commercial or industrial situations, however, you are more likely to be working as part of a group or team. Here, the so-called brainstorming approach is popular. There need not be a hierarchy within a brainstorming group – you mix contemporaries from different disciplines or representatives of other departments, with the assurance that each member is accorded respect and allowed to express suggestions 'without prejudice'. In a group situation everyone should feel totally at ease, free to put forward any idea that occurs, however lateral, apparently silly or unlikely. Next, the group is at liberty to ask questions and put the idea up for enquiry, but reasons to discard any idea must be rational and valid. Thus the atmosphere, relaxed and receptive, is open to completely new and innovative solutions.

Not all engineering is about innovation. Techniques such as brainstorming, used as above, and creating Spider diagrams of ideas, are a way of bringing ideas together, but not all the ideas will be original.

Box 6 Spider diagrams

A technique widely used for stimulating free thought is the spider diagram (sometimes known as a spray diagram). It works on the principle of removing the hierarchy of importance, implied when items are written in list form. Instead, the title of the problem or need is put at the centre, and as items are thought of, they are placed in more or less random positions on the paper. The connections between related items are then represented by lines, to produce a multi-limbed structure that gives the diagram its name. Usually, the process of drawing the relationships between items stimulates the addition of further items. Once the relationships have been made, it is possible to group several items together in themes, and a tree-like structure emerges. The diagram is now useful as a map of the whole topic, its critical issues, and the relationships between them. Figure 11 is an example.

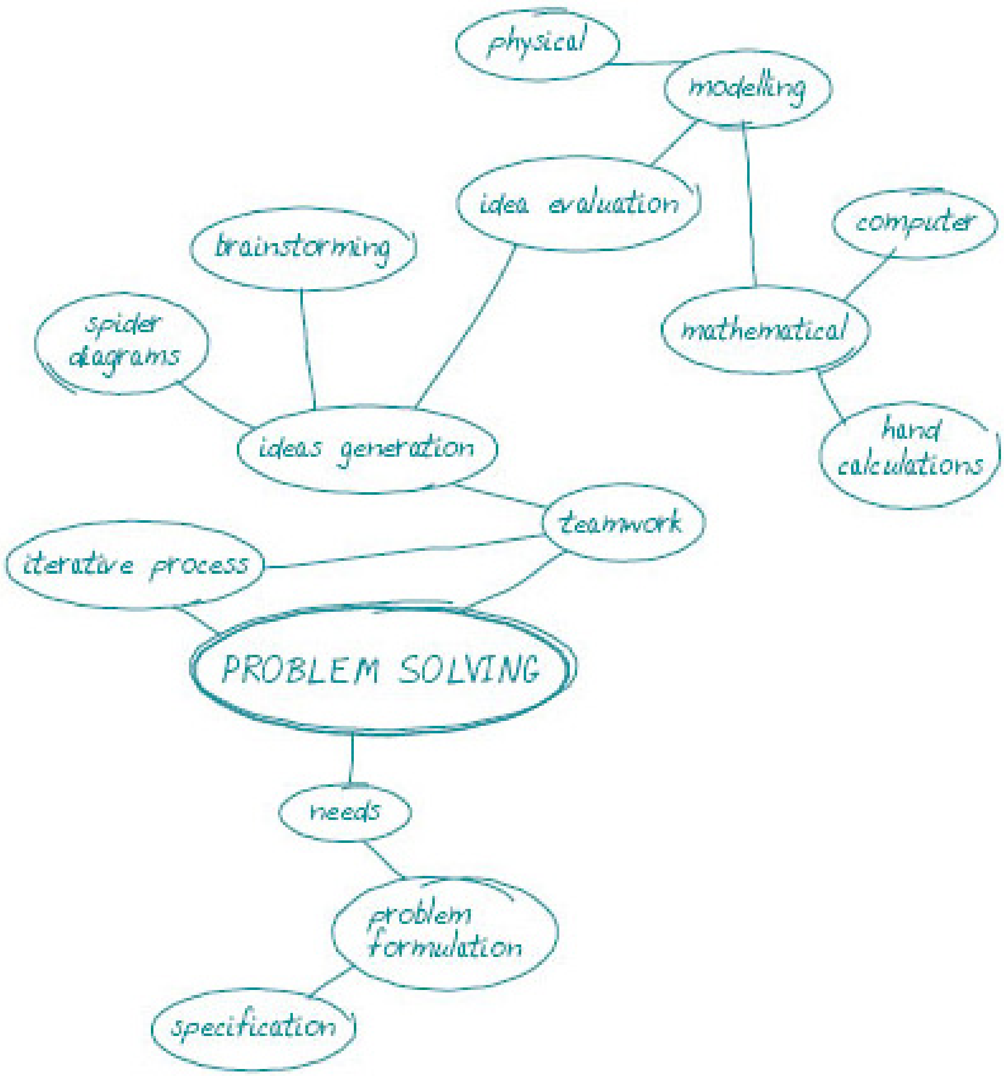

Figure 11 A spider diagram of issues in problem solving

In all but the most 'blue sky' organisations, constraints are present in terms of cost, time, capacity, environment, manufacturing capability – you think of it, it's a constraint. Add to this the limits imposed by our knowledge of the physical world – things like data storage capacities, material, fluid or gas properties, Anthropometrics and ergonomics and so on –

and the problem up for attention has either shrunk significantly or has become more complex.

Box 7 Anthropometrics and ergonomics

An engineer uses *anthropometric* data when designing something that will be operated or used by a person, or rather more specifically, by any unknown person. It represents the weight and measurements of the average man, woman and child, usually presented in centiles (the 50th centile being the median average), and covers everything from basic height to the length of a little finger.

What the average person can *do*, on the other hand, is presented as *ergonomic* data. This is about how hard we can push, what pressure we can exert on a foot pedal, the most comfortable reach, etc. In the same way that we use tables of data for solid, gas or fluid properties in the specification of materials, we use anthropometrics and ergonomics to design for people.

Tables 2 and Tables 3 show the kind of information that is typically gathered and used.

Table 2 Anthropometric data f or adult British population (age 19–65 years)

	Male		**Female**	
Body dimension	Mean/ mm	SD*/ mm	Mean/ mm	SD*/ mm
Stature	1740	70	1610	62
Eye height	1630	69	1505	61
Shoulder height	1425	66	1310	58
Elbow height	1090	52	1005	46
Sitting height	910	36	850	35
Sitting eye height	790	35	740	33
Sitting shoulder height	595	32	555	31
Sitting elbow height	245	31	235	29
Thigh thickness	160	15	155	17
Buttock-to-knee length	595	31	570	30
Buttock-to-popliteal length	495	32	480	30
Knee height	545	32	500	27
Popliteal height	440	29	400	27
Shoulder breadth (bi-deltoid)	465	28	395	24
Hip breadth	360	29	370	38
Chest (bust) depth	250	22	250	27
Shoulder-to-elbow length	365	20	330	70
Elbow-to-fingertip length	475	21	430	19
Forward grip reach (from the back of the shoulder blade)	780	34	705	31

Upper limb length	780	36	705	32

Source: Pheasant (1986)

*SD=standard deviation, representinga statistical departure from themean value.

Table 3 Ergonomic data: maximum finger pushing force in newtons

	Thumb	Index finger	Middle finger	Ring finger	Little finger
Mean force/N	17	11	10	8	5
Range/N	14–20	8–14	8–12	5–10	3–9

Source: Haaland et al. (1963)

SAQ 4

The back of a particular airline seat is 100 mm thick. Explain why it would be unreasonable to install 12 rows of seats in a cabin with a floor length of 7 m. Suggest, with a brief justification, a more reasonable number of rows.

Answer

The space available per seat with 12 rows in 7 m would be

$$\frac{7}{12} = 0.583\text{ m} = 583\text{ mm}$$

Of this, 100 mm must be allowed for the seat back, leaving just 483 mm for passengers' legs. According to Table 2, the mean buttock-to-knee length of an adult male is 595 mm and that of women is 570 mm. Thus, there is insufficient space for an average man or an average woman.

A more reasonable capacity might be based on the mean, male, buttock-to-knee length (595 mm) plus one standard deviation (31 mm), plus the depth of the seat back (100 mm) for each seat. That amounts to 726 mm per seat. The number of rows would then be 7000/726 = 9.6; in practice this would need to be rounded down to 9 rows with 777 mm per seat.

For comparison, typical economy-class seat spacings are 710 to 860 mm, and the UK Civil Aviation Authority's current minimum (under review) is 660 mm.

The search for solutions must involve a thorough appreciation of the problem. It may involve detailed analysis and calculations based on scientific and engineering principles and using technical data. This is mathematical modelling, and it is useful at many stages in the process of identifying solutions; it is specifically addressed later in Section 4.5.

In practice then, finding a solution is usually a delicate balance between finding 'the best design' and getting something into the market-place 'by yesterday'. Earlier on I asked you to remember a point:

> *As good engineers, in an ideal context, we are not just looking for a solution; we are looking for the best solution.*

Now you see why I added the condition about context. I think that any course that hopes to contribute to the formation of professional engineers has a responsibility to make this clear. Over the duration of an engineering qualification, you will learn a little about many of the tools you need to solve problems. You are likely to specialise, and learn more about, say, mechanical, civil or electrical engineering, building services, software, chemical processes, nuclear power and so on. It would be impractical to expect you to study every problem-solving technique tailored to every conceivable context. However, a good course of study will make sure that you are aware of the constraints that we have discussed above, and that you have some practice in bringing together maths, science or technology in ways that create practical, physical solutions. That way, when you call on your skills 'for real', you won't be surprised when your 'best' solution is ditched in favour of the one that has a quarter of the durability but costs half the money and can be made within the organisation. Instead of being depressed about this, understand and use those constraints to shape your next design; if you know that such limits will be imposed, make them your goals and work to achieve them.

Getting back to our problem-solving diagram, you may have noticed that Figure 10 shows another circuit – a loop from possible solutions, to evaluating solutions, and back to the problem. You may also remember that the problem is linked back to a need (Figure 8), and so at this stage any suggestion may take you right back to the start, asking new questions about the need and refining or redefining the problem, quite possibly by going back to the customer. The trick here is *not* to redefine the problem in order to suit your solution, but to be sure that your solution is *meeting the need*.

4.4 Evaluate solutions

If the obvious solution has been identified and everyone is in agreement, then a formal evaluation of solutions is unnecessary, and we would move on to modelling the design. However, if there is dissent then some stricter method of elimination is required, and this is usually achieved through a process of rank-ordering. There is little to be lost and potentially much to be gained by returning to the customer at this point for opinion, clarification or guidance.

4.4.1 Selecting the best candidate

Assuming there is more than one likely looking candidate solution, we need to make a selection now so that we don't waste time taking all the candidates through the next steps, which become progressively more expensive and time-consuming. The rigour and formality of this step is very variable, but in general all schemes boil down to the same process you might use to choose some consumer item, such as a TV set. You would have a list of criteria that are important to you, and you would evaluate each candidate against those criteria. In many cases the list is short enough, or a single criterion of such importance relative to the others, for it to be possible to have it in your head. Usually it is worth writing down that list (which should look like the specification), and assigning a weighting to each criterion according to its relative importance. You then give each candidate solution a score against each criterion. When multiplied by their respective weightings, these scores add up to a figure of merit for each solution. The one with the largest number wins.

A benefit of using a system such as this is that it tells you quantitatively what kind of a 'squad' of substitute solutions you have to draw from (to use a sporting analogy). This is important because it tells you whether or not you really ought to take more than one on to the next step and beyond, until there is a clearer preference. It also tells you whether or not you have only one possible solution that's going to be worth considering. If this is the case, it may be good news if you are confident it will work, or it may prompt you to go back now and try to generate some more ideas.

Let us try this approach with the example of choosing a TV set. First, we ask what are our criteria, and what relative importance do we attach to each of them? This can be set out in a table:

Criterion		Relative importance (weighting 0–10)
1	Large screen	8
2	Good picture quality	10
3	Good sound quality	8
4	Compatible with other audio/video systems	10
5	Attractive cabinet	5
6	Cost	10

Next, we need to score each of our candidate models of TV set against each of the criteria, then multiply the scores by the appropriate 'importance' weighting.

	Model 1		Model 2		Model 3	
Criterion	score (0–10)	× weight	score (0–10)	× weight	score (0–10)	× weight
1	10	× 8	6	× 8	6	× 8
2	6	× 10	5	× 10	10	× 10
3	5	× 8	4	× 8	6	× 8
4	10	× 10	0	× 10	5	× 10
5	3	× 5	10	× 5	8	× 5
6	8	× 10	7	× 10	10	× 10
Sum of weighted scores	375		250		386	

For each model we arrive at the sum of the weighted scores.

On the basis of this, we might eliminate Model 2, but we might need to consider some additional criteria to choose between Model 1 and Model 3.

4.5 Model the best solution

In moving from the 'possible solutions' to the 'best solution' box, Figure 12, we have to assume that a certain amount of evaluation has been done in the previous loop. The solution is still on paper, and probably not much more than a sketch, but something is badly wrong if the best solution to come forward has not been recognised to be at least feasible in the most basic terms of function, cost and implementation. The next step is to model the solution to estimate how well it will perform. Depending on the subject of the problem, this could take many forms.

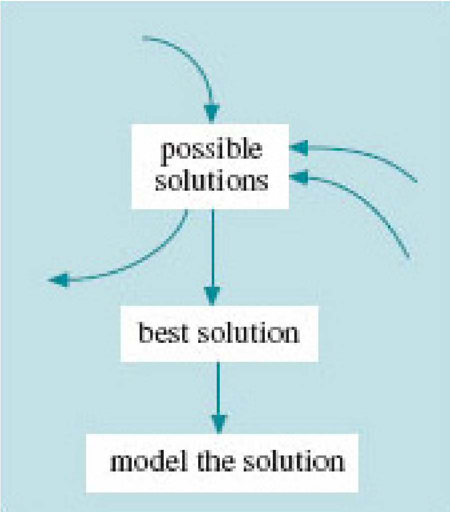

Figure 12 A part of Figure 7

The model itself doesn't have to be physical and sometimes a mathematical model can be used. A pretty thorough knowledge of relevant physical properties of real systems, materials and structures is required if a model is to be of any practical use.

4.5.1 Mathematical models

Computers in the last few decades have, in many cases, made mathematical modelling a lot easier. Models that used to require hours of manual cranking through long equations can now be created on a screen using specialist software. Processes can be recreated – modelled – in the time it takes to press a few buttons.

For example, when designing a pipe network to carry a gas or fluid, such as in the village water supply problem, you might wish to know how the flow would be distributed within the network, i.e. at what rate the fluid would exit from each distribution point. In order to calculate this, you would need figures for: fluid density; the number of pipes in the system; the length and diameter of pipes joining each connector; the friction factor for each pipe (a constant determined by the roughness of the pipe wall); pressure losses at the pipe fittings; volumetric inflow at each connector; and so on. Imagine the calculations necessary to cover all of these, in what is actually a relatively uncomplicated arrangement of pipes.

Engineering has a whole host of branches, covering a huge variety of physical quantities. The following are examples of where computer-based techniques are used.

- Thermal and stress analyses model the distribution of quantities like temperature and stress across a shape or an area, and are useful for spotting potential weak points in a design.
- Computational fluid dynamic studies analyse turbulent flow in gases and liquids, such as in the simulation of weather.
- Universal Modelling Language simulates computer software.
- Circuit simulators model the performance of electronic circuits.

Of course, not all modelling is done on computers. Creating a mathematical model of something can be as simple as drawing a picture, putting in the mandatory dimensions and using the diagram or some simple calculations using algebra to determine the remaining quantities. For instance, look at the Box 8 Mathematical modelling example.

Box 8 Mathematical modelling example

In another aspect of the village water supply problem outlined earlier, part of the design of the network may include some sort of storage tank. Let's assume that the tank is cylindrical and that it has been specified to be capable of holding 1500 litres (1.5 m^3). In order to keep costs down, we want to use the minimum amount of material in constructing the tank and must therefore use mathematics to model the problem, finding the minimum surface area capable of holding the required volume.

The first step is to write down algebraic expressions for the key factors – volume and surface area in terms of the unknown quantities (radius *r* and length *l* of the cylinder) that are to be determined.

Volume:

$$V = \pi r^2 l$$

Surface area (two ends plus the side):

$$\begin{aligned} s &= 2\pi r^2 + 2\pi r l \\ &= 2\pi r(r + l) \end{aligned}$$

Next, combine these expressions by using information from one equation to simplify the other.

Rearranging the equation for volume to express it in terms of length:

$$l = \frac{V}{\pi r^2}$$

We can substitute this expression for *l* into that for the surface area, bringing the problem down to one unknown, *r*, as the volume *V* is specified:

$$s = 2\pi r\left(r + \left(\frac{V}{\pi r^2}\right)\right)$$

Multiplying out and tidying up:

$$s = 2\pi r^2 + \frac{2V}{r}$$

There are two ways to find the minimum area for the required volume. The first is trial and error, for which a calculator or computer spreadsheet is invaluable. Try, for instance, a radius of 1 m and one of 2 m – which gives the smaller area? What value would you choose next? Try it. Often plotting a graph of the results helps you see where to direct your attention.

The second method is based on calculus, an analytical technique that should be familiar to you from your study of courses in mathematics. If it is not, you will have to stick to trial and error. For those who can follow it, the calculus approach is given below; otherwise move on to the text following Figure 13.

Differentiate the expression for the surface area with respect to *r*:

$$\frac{\mathrm{d}s}{\mathrm{d}r} = 4\pi r - \frac{2V}{r^2}$$

For maxima and minima, d*s*/d*r* = 0 (remember that d*s*/d*r* is the gradient of the curve of *s* against *r*), so:

$$4\pi r - \frac{2V}{r^2} = 0$$
$$r^3 = \frac{2V}{4\pi}$$

Substituting the value V = 1.5 m^3:

$$r^3 = \frac{2 \times 1.5\ \mathrm{m}^3}{4\pi}$$
$$r = 0.62\ \mathrm{m}$$

We can use the earlier expression for the volume to find l:

$$l = \frac{V}{\pi r^2} = \frac{1.5\ \mathrm{m}^3}{\pi(0.62)^2}$$
$$l = 1.24\ \mathrm{m}$$

Mathematical modelling tells us that the tank must be 1.24 m high with a radius of 0.62 m – that is, the length l is equal to the diameter (2r). Figure 13 shows how the shape of a cylinder looks as the ratio of r and l varies. It is obvious from our everyday experience that the cylinder which will hold the most water is the one that is neither extremely thin nor extremely short. Our numbers from the calculation look reasonable in this respect; if you looked at the cylinder side-on, it would look square.

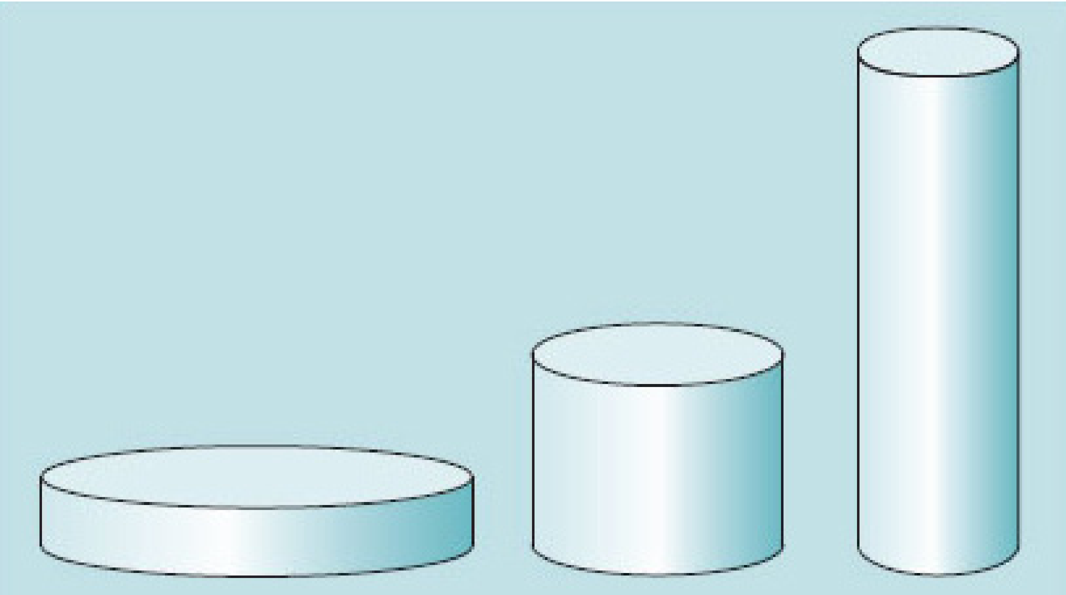

Figure 13 Thin, 'square' and short cylinders with the same volume

Exercise 3

In the mathematical modelling example we arrived at an 'optimum' cylinder which had a square cross-section (diameter equal to its length). But let's challenge this result by asking if any assumptions have led us away from the *best* solution.

A major assumption was made in setting up the problem, just before the mathematical modelling was presented. What was it? Do you think it was a reasonable one?

Can you think of reasons why the cylindrical shape may not be the best solution to this very specific problem?

Hint: think about the context and how this imposes constraints of its own.

Answer

The assumption was made that a cylindrical shape is going to be the best, and all we need to decide is the ratio of length (or height) to diameter. In fact, the shape that encloses the largest volume with the smallest surface is a sphere.

It seems reasonable to assume that, in this context, a spherical shape should not be used, as the problems of forming and joining or welding the sheet are considerable, and the necessary equipment is not likely to be available.

Here is my list of reasons why the cylinder may not be preferred – you may have thought of others:

- the shape of available sheet material
- the ability to cut circles and form the sheet into curves
- the space it has to fit in
- cultural preference for or aversion to a particular shape.

SAQ 5

Making the cylindrical tank from a rectangular sheet of metal will create offcuts of material where the top and bottom circles have been cut out of the sheet. Show that if the most efficient cuboidal shape had been used instead (i.e. a cube), it would use less sheet material than the cylindrical shape. Assume that the width of the sheet material for each case is ideal.

Answer

To get a volume of 1.5 m^3 in a cubic tank requires the side of the cube to be $\sqrt[3]{1.5} = 1.145\ \mathrm{m}$. The cube has six faces, so this can be made from a sheet of material of dimensions 1.145 × (6 × 1.145) = 7.866.

The total area of the sheet is therefore 7.866 m^2.

To make the tank cylindrical, we need the dimensions previously calculated, i.e. a diameter of 1.24 m and a height also of 1.24 m.

If our sheet of metal was 1.24 m wide, we could make the two ends and the sides of the cylinder from a piece that was (2 × 1.24) + (π × 1.24) = 6.376 m long – sketch it if you need to.

This has an area of 1.24 × 6.376 = 7.906 m^2

The cylindrical tank uses 7.906 − 7.866 = 0.040 m^2 more sheet material (but this difference is only about 0.5%).

4.5.2 Physical models

A physical model of an artefact or component is often built on a reduced scale, in size and/or by using materials that are cheaper and easier to manipulate than those intended for production. At this stage, we are not necessarily producing what you might think of as a prototype, but investigating particular aspects of the design. For instance, maybe we would produce a racing-bike frame to a new design but in a cheap material such as balsawood, in order to assess the air flow around it in a wind tunnel.

A good example of this is the way modelling was used in solving the problem of the steering for the Thrust SSC supersonic car (Figure 14), the first land vehicle to travel faster than the speed of sound (Mach 1), covering a kilometre in an amazing 2.9 seconds.

Figure 14 Thrust SSC supersonic car

One of the first questions that needed to be asked, even before aerodynamic issues were considered, was 'can a rear steering mechanism work at speed?'.

Ron Ayers, the engineer in charge of aerodynamics on the car, explains:

> [So] we had to steer with the rear wheels. As soon as we suggested that, all the experts on car dynamics waved their arms in the air and said it was unstable, it could never work, think of forklift trucks and shopping trolleys and other irrelevant comparisons. However, Professor Crolla at Leeds University did a theoretical study for us and said yes, it could work under certain circumstances and gave us those circumstances. There were, however, still plenty of critics who said it wouldn't work.
>
> The rear-wheel steering Mini was built to demonstrate that you can steer a vehicle from the back [Figure 15]. Although it's a lot smaller than Thrust SSC, the wheel configuration and track, etc., is in fact a scaled-down version of the full-sized car. We drove that ancient Mini with this kind of extension out the back at 90 mph up and down the test track at MIRA and proved that it was very controllable – that we really could, with finger-tip control, keep it on the line.
>
> Glynne [Bowsher, the mechanical and structural designer] designed it and built it with his brother-in-law. It's his brother-in-law's Mini, actually. I think they spent £300 and it cost three weeks and two near-divorces.

Figure 15 Mock-up steering system

Figure 15 shows that the model looks nothing like the car that in 1997 made engineering history by smashing the barrier of sound! The Thrust team used mathematical and physical modelling intensively throughout the development of the car. Given the costs involved in trialling the final product, they needed to be as sure as possible that it was going to work. A model is built to *prove* it will work and to collect data or to test some aspect of design.

SAQ 6

Identify which aspect of the Thrust SSC was not being addressed by the physical model using the adapted 'Mini'.

Answer

The most important aspect of the Thrust SSC not addressed by the physical model is the aerodynamics and how this affects the steering at full speed.

We will return to modelling at the end of this free course on the problem-solving process. When we look at it again, we will present two contrasting cases. The first is a familiar product, namely the bicycle, and will show the merits of careful 'hand calculations'. The second, probably less familiar, is an acceleration sensor for triggering vehicle airbag systems. It will illustrate the use of a more complicated, computer-based model.

4.6 Assess and review

Following our problem-solving map, we have reached the stage of 'assess and review solution', Figure 16.

Figure 16

If we've got everything right first time in the preceding stages, then the solution successfully meets the need, with no obvious nasty side effects, and we can pass directly to 'build prototype/demonstrator', pausing briefly for cheers and back-patting all round. On the other hand … this may be the time to recognise that the 'best solution' hasn't worked. This isn't necessarily as disastrous as it sounds; unless the solution was particularly simple, there are likely to be elements that have been found to be successful where others have not. This usually means that you are not entirely back to the drawing board, but will have to revise the bits that haven't worked and replace them with other suggestions that have been on ice from the 'possible designs' stage. If we were prolific with ideas then there may be an orderly queue of replacement solutions, such as alternative materials, configurations of components, substitute objects (in software design) or whatever. However, if there is no obvious alternative then we may end up having to reconsider the original need (Figure 17).

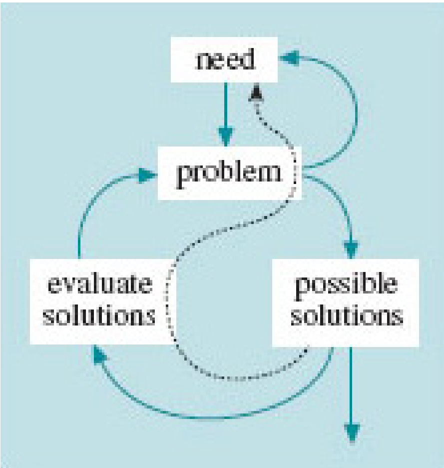

Figure 17 A part of Figure 7

There's no shame in going back to the customer at this point and asking for clarification. It may even be that your solution worked according to the need they expressed, but some incidental aspect has proved unworkable and so the problem needs restating. Returning to the water problem and the need to transport water from the river – perhaps you have discovered that the flow of the river isn't enough to meet the needs of the village. In this case, there is no need to abandon the solution entirely, as it will still be possible to supply water from the river, but a supplementary water source must be identified and harnessed, and this will be the new problem. You put forward possible solutions … and so the cycle begins again.

4.7 Build prototype/demonstrator

The physical models we talked about earlier are prototypes or demonstrators of a sort. However, for the purposes of making a clear distinction in the process, I'm referring here to prototypes or demonstrators as functioning preliminary models of the essential finished product or construction or service, bringing together all the elements of the design that may or may not have been previously physically tested (Figure 18). This is still a model – if appropriate, it may be a full-sized, full-colour replica – but it may also be a scaled-down version where only the vital working parts are fully functioning.

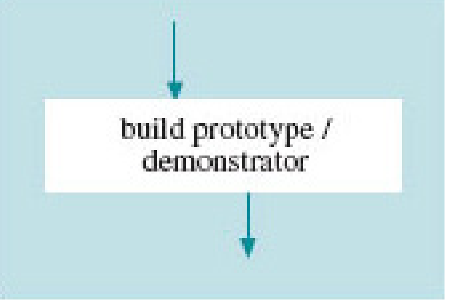

Figure 18 A part of Figure 7

In the village water supply problem, you may produce a model version of the layout of pipes to ensure that they will conform to the landscape, or you may not need to produce a demonstrator at all. The point is, where necessary, to make sure that all the essentials will come together and operate as anticipated.

In software design, for example, the prototype may be a program that combines all the essential elements of code, written by various teams and controlling various parts, but with a dummy user interface. The vital commands will be in place, but asides such as 'glossy' add-ons (for example the electronic games on a mobile phone) are not necessary to the success of the solution, and may be designed separately and added to the overall design at a later stage. Note that to do this, you have to be *very* sure that these extras will not affect any important aspect of the rest of the product. In a well-run project, you will have to prove to the rest of the team or the project leader that this is a valid assumption before being allowed to leave it off the demonstrator.

Put simply, with good planning and thorough preparation and groundwork the demonstrator should work when tested in the conditions in which it will be expected to perform.

4.8 Assess and review again

If you've been following the stages of our problem-solving map, then the chances are you're ahead of me here (Figure 19). Yes, if it works, hurrah; if it doesn't then off we go again, all the way back to 'possible solutions' and selecting the best of the rest. Or maybe even going back to the beginning. No one will be amused by a failure at this stage, when considerable investment in time and money has already been made. We should be looking at fine-tuning only, collecting data to give the marketing people, and finalising decisions about fabrication, manufacture, production quantities (if relevant) and processes. In mathematics, this process of refining a solution to get ever closer to a final value is called iteration.

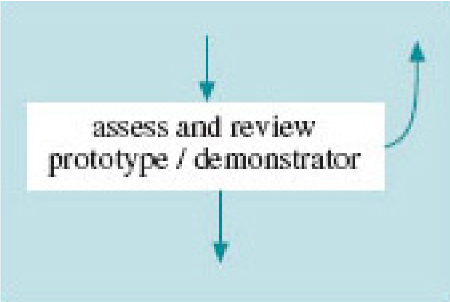

Figure 19 A part of Figure 7

4.9 Final implementation

The line you take here obviously depends on the problem you set out to solve. If you were creating a new product for retail or industry, then the final step of the process would be to put that product into manufacture and watch it go off into the world to begin its life cycle (Figure 20). If the solution were a one-off, such as the village water supply problem we considered at the start, then it would be built and installed.

Figure 20 A part of Figure 7

Not all engineering problems will fall neatly into this pattern, inserting need at the top and extracting a solution at the bottom. Furthermore, not all problems and challenges are to do with designing something from scratch – many real challenges are concerned with a more restrictive need to improve or repair an existing system. Engineers work in different ways, under a variety of conditions and often without the luxury of the time, resources or finance it would take to follow the above process to the letter for every need that was presented. The process will vary according to the nature of the problem, and the experience and understanding of the team engaged to find the solution.

SAQ 7

Explain why the list of criteria used in selecting the best candidate solution will look like the specification. State the reason for assigning weightings to each criterion.

Answer

The specification should be the document that defines everything that is wanted from the solution. Therefore this is the most natural basis from which to generate the list of selection criteria. A full specification will contain a large number of items, including some which are preferences rather than absolute, hard-and-fast necessities. If no weighting is applied, then the ranking of the candidate solutions could be incorrect in relation to the real need, simply because the best solution has a lower score on several relatively unimportant criteria.

In the next two sections I want to show some aspects of the problem-solving process at work.

5 A problem in bicycle design

5.1 The development of the bicycle

Section 4 has looked at how we can follow a logical route or map, from the expression of a need, to arrive at possible solutions to a problem. In Sections 5 and 6 we look in more detail at two quite different examples of engineering problems. Our first example is the historical development of the bicycle frame; the second concerns a vital component of a car's airbag system.

The weight of a bicycle frame is a major burden that the cyclist has to bear. There have certainly been times when I felt I needed a lighter bike, usually when going up hills. So, I want us to begin 'identifying and evaluating solutions' for the problem of how to make a bicycle frame lighter, while retaining its performance. A little historical background is important as we are not the first to address this problem.

The modern bicycle frame is central to a huge international business, dominated by American and Chinese markets. In the USA, current estimates put the market at about 15 million bicycles per annum and, in an increasingly environmentally conscious society, the future looks pretty much guaranteed. In this context, the bike is a mature product in an established market, but at some point in history there were no bicycles and the first design must have been produced in order to meet a need. We can guess that this need was for a mode of transport that was quicker than walking and that didn't need feeding.

In fact, the first vehicle that worked with two wheels in a line (the very basic characteristic that came to define a bicycle) was introduced in France in 1791 with a non-steering front wheel, no pedals and a wooden horse's head! For its time, this 'bicycle' was an innovation, and unbelievably it didn't change much for the next twenty years or so. In 1817 a steerable front wheel was introduced (in Germany) and pedals made their first appearance in 1839 (in Scotland). The chain, the sprocket-driven rear wheel and equal-sized front and back wheels were added in England in the early 1880s and were followed by pneumatic tyres, two- and three-speed hub gears and then derailleur gears before the turn of the century.

Each of these major changes came about through engineers finding solutions through innovation by context, but along the way there were literally hundreds of small innovations by development and numerous routine design improvements. For example, in 1879 Charles E. Pratt wrote in his handbook *The American Bicycler*:

> From 1868 until the present time the patented improvements have been numerous, and the mechanical details of construction have been thoroughly worked out, until the machine has become a marvel of ingenuity and of workmanship; and the modern bicycle has been developed to its present state of perfection in strength, lightness, ease of propulsion, certainty of control, and gracefulness of design and operation.

This is an accolade that would swell any engineer's head, but it implies that any advances since 1879 have been extraneous!

SAQ 8

The pneumatic tyre was a novel use of a sealed tube of pressurised air to provide the running surface of a bicycle wheel. Say whether you view its introduction as an innovation or a routine improvement.

Answer

The pneumatic tyre was an innovation by context. Tyres had not been made in this way before, though the idea of sealing pressurised air into a container was not new (e.g. balloons, footballs, etc.).

An interesting aside to the technological advancement of the bike is in respect of the effect that engineering has on society. The bicycle is generally attributed as being pivotal in the liberation of women in the USA and Europe. Women in England were able to travel independently and in relative safety for the first time, particularly after Queen Victoria made it a socially acceptable practice by riding her tricycle alone in public. A postmistress in the USA, Amelia Bloomer, designed the first trousers for women – bloomers – specifically for riding bikes. The same Amelia Bloomer went on to spearhead the Suffragette fight for women's right to vote in the USA.

Bringing the bike up to date, the maturity of the contemporary model can be seen by comparing Sir Edward Elgar's Sunbeam of 1903 (Figure 21), a relatively sophisticated design in the history of the bike, with a modern all-terrain bicycle (Figure 22). The ancillary equipment has changed but the functional form of the frame structure is still based on a triangulated tubular diamond shape.

Figure 21 Sir Edward Elgar with his 1903 Sunbeam bicycle (courtesy of the Elgar Birthplace Trust)

Figure 22 A modern mountain bike

Today's market for bicycles is highly competitive, and manufacturers strive to appear better by being different. The 'engineering problem' that I identified at the start was to do with making a lighter frame. It is important therefore to separate real structural and performance improvements to the frame from fashionable gimmicks that might even add weight. To facilitate such comparisons it is necessary to understand the fundamentals of the product, and that calls for a bit of modelling.

So, back to basics: what is a bicycle frame?

SAQ 9

Write a brief specification (five or six lines) for a bicycle frame, considering its physical form and what it must be physically capable of achieving.

Answer

In engineering terms the need is for a space frame that enables two wheels to be held in place and supports the forces developed by the rider's mass and his or her efforts. The frame should be rigid enough to keep the rear wheel in line with the chain wheel, which is attached to the pedals. In addition the front forks must also rotate to allow the front wheel to steer at low speeds.

The bicycle can be seen as a variation on a seven-membered truss (Figure 23), a frame structure that is a common element in structural and mechanical engineering.

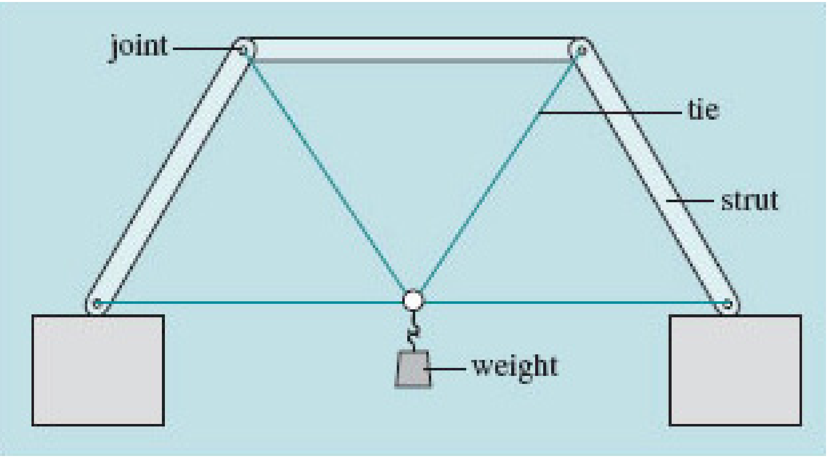

Figure 23 A seven-membered truss constructed from four tie-strings and three struts, and joined by pin joints

If the axles of the wheels replace the side supports, the bike begins to reveal itself, as shown in Figure 24.

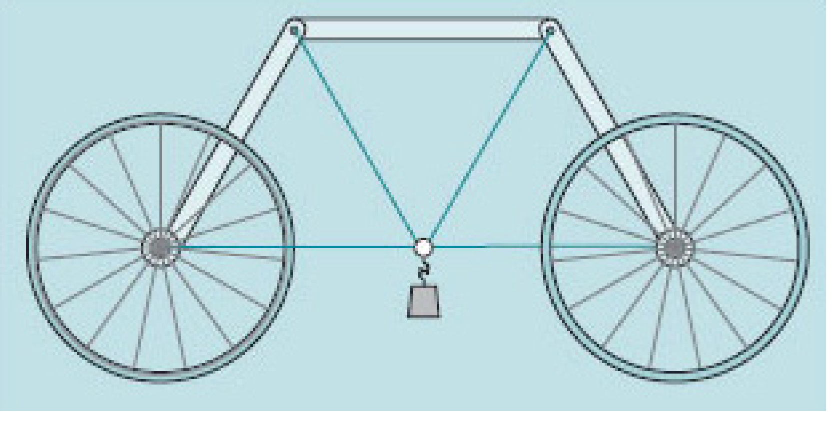

Figure 24: Seven-membered truss supported on wheels

This structure is capable of supporting a central load as shown and will clearly subject the truss-based space frame to the same in-plane forces – compression in the struts and tension in the tie-wires. There are only a few simple adjustments left to turn the basic structure of Figure 24 into a 'functional' bicycle frame configuration, Figure 25.

One of the tie-wires has been removed in order to allow the front wheel to rotate, and the strut attached to the front wheel has to be free to rotate at the pin joint. Add a saddle, chain, pedals and handlebars and away you go.

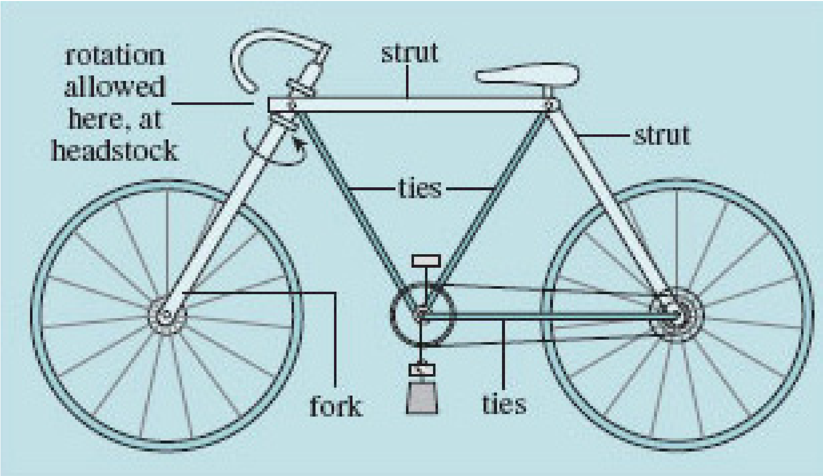

Figure 25 The strut-and-tie bike frame evolves

By making a comparison with a simple structure that can be fully analysed, we have been able to confirm the fundamental elements of the bicycle structure, the bit that carries the major forces. However, there is a difficulty. As we shall see in the next block, Figure 24 represents what is called a statically determinate structure. Removing one component risks turning it into a mechanism: that is, it may change shape in response to forces. Removing the bottom-left tie in the truss of Figure 24 would cause it to collapse unless something were done to rectify or compensate for its absence. If part of the top-left pin joint is fixed and does not allow the attached side-strut to swivel about the joint axis, the strut is turned into a cantilever, which restores stability to the structure but at a price. The original strut, now acting as a cantilever, is no longer subject to just axial compression, but has additional forces that lead to bending about the frame region called the headstock. Such a configuration creates high stresses in the material, in this case reaching a maximum at the top of the cantilever fork close to the headstock. The relatively simple problem of allowing this cantilever to rotate is soon solved by the use of bearings that can support forces whilst rotating. Hence, the 'bicycle' shown schematically in Figure 25 could function, if manufactured from adequate materials for the struts, ties and cantilever.

The next thing to do is to see how we can, in general, evaluate solutions based on different materials. Then we'll reintroduce the specific needs of a bicycle frame.

5.2 Material comparisons

I want to depart from the specific example of the bicycle to make some more general points.

In most simple structural analysis the self-weight of the structure is ignored, as it is considered to be small in comparison with the loads carried. However, as an illustration of engineering practice in the search for efficient structures to employ in product design, it is worth examining how the strength and weight of particular materials compare.

These comparisons are illustrated through the use of modelling. As an example, let's estimate the maximum length of a hanging tie-rod and the maximum height of a column. See Box 9 Long ties and high columns.

Box 9 Long ties and high columns

Figure 26 shows a parallel-sided rod of material of density ρ and yield stress σ_y hanging from a support.

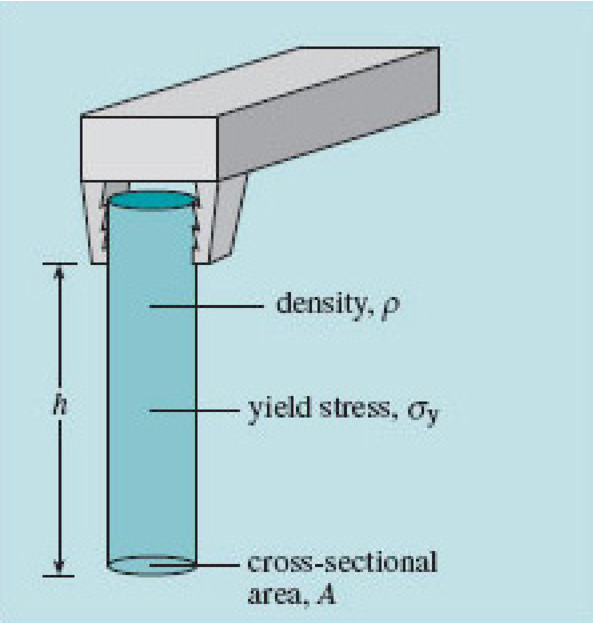

Figure 26 Stress in a rod suspended from a support

The stress in the cross-sectional area *A* increases on each horizontal plane as you go up the rod. It will fail by yielding in the uppermost section when the stress there reaches the yield stress, which is a characteristic of the material. If this occurs when the rod is *h* metres long, we can say that:

Force on the failure plane

$$\begin{aligned} &= \text{mass hanging below it} \times \text{acceleration due to gravity} \\ &= \text{volume} \times \text{density} \times g \\ &= A \times h \times \rho \times g \end{aligned}$$

If the material yields then the force on the plane at failure =

$A \times \sigma_y$.

Hence, equating these values gives:

$$A\sigma_y = Ah\rho g$$

Therefore,

$$\sigma_y = \rho g h$$

Notice that the cross-section term *A* cancels out from the equations.

Example

A mild steel has a density of 7800 kg m^{-3} and yield stress 300 MN m^{-2} What is the maximum length of rod that could be dangled from a high building without yielding under its own weight? Take g = 9.8 m s^{-2}.

Answer

Using the equation $\sigma_y = \rho g h$, and rearranging:

$$h = \frac{\sigma_y}{\rho g}$$
$$h = \frac{300 \times 10^6}{7800 \times 9.8} = 3925 \text{ m}$$

Hence the rod could be almost 4 km long before it would break under its own weight. But could we build the tower from which to hang this extremely long rod?

Example

Consider building a brick tower from which we could suspend a rod for testing. Brick has a crushing strength σ_c of 70 MN m^{-2} and a density of 2000 kg m^{-3}. What is the maximum height of a parallel-sided tower that can be built without crushing the bottom course of bricks? Assume the mortar thickness to be negligible and of higher strength.

Answer

The relevant equation comes from relating the weight of bricks above the bottom layer to the force required to crush it. The result is:

$$\sigma_c = \rho g h$$

The crushing strength and density of the brick are inserted into the formula to give:

$$h = \frac{70 \times 10^6}{2000 \times 9.8}$$
$$= 3571 \text{ m}$$

That's right, over 3.5 km (2 miles) high!

Hence the crushing strength of brick would limit the construction of a tower to test the steel rod. You would have to taper it to be narrower at the top.The limiting equation for both long ties and struts is:

$$\sigma = \rho g h$$

Rearranging this gives a general expression for maximum length in terms of the maximum self-supporting height:

$$h = \left(\frac{1}{g}\right) \times \left(\frac{\sigma}{\rho}\right)$$

Now (1/g) is a constant, and if g stays constant h is proportional to (σ/ρ). This quantity is a measure of a material's ability to support itself in tension or compression (depending on which value is used for the strength). It is called a 'merit index for self-supporting strength. This means long self-supporting ties and long self-supporting struts are more feasible if the merit index (σ/ρ) is big. Not surprisingly, this occurs with high-strength, low-density materials.

5.3 Back to the bicycle

Let's assume that our bicycle frame could still be constructed from ties and struts. If we want to select the material to minimise the weight of a frame for a particular frame strength, we need to devise a merit index as follows.

The mass of the tie-rods and struts needed for the frame is given by:

$$m = \text{volume} \times \text{density}$$

Therefore,

$$m = Ah \times \rho \qquad (1.1)$$

where *h* is the length of a component and *A* is its cross-sectional area. The failure force for tensile yielding, *F*, is given by:

$$F = \sigma_y \times A \qquad (1.2)$$

in which σ_y is a property of any chosen material. Eliminating *A* from Equations (1.1) and (1.2) gives:

$$m = h\rho \times \frac{F}{\sigma_y} \qquad (1.3)$$

Usually we want a materials-based index that gets bigger the better the material. Hence it is better to express our index in terms of (1/*m*), which gets bigger the lighter the frame.

Hence, rearranging Equation (1.3) gives:

$$\left(\frac{1}{m}\right) = \left(\frac{\sigma_y}{\rho}\right) \times \left(\frac{1}{Fh}\right)$$

Now for a tie-rod of particular length *h*, able to resist a particular force *F*, the bigger the value of the material merit index (σ_y/ρ) the lighter the frame could be for the same performance.

From identical considerations, the bigger the value of (σ_c/ρ) the lighter a frame could be made for a specified performance.

Using a comparable approach you can also select a merit index to find light struts and ties that limit elastic strain, giving a particular deflection under a given load.

In these circumstances selecting the material with the highest value of (*E*/*r*), where *E* is the Young's modulus, has the potential to give the lightest frame components that deflect by a particular amount under load.

Table 4 shows absolute values and some merit indices for a range of recently used frame materials.

Table 4 Absolute values and merit indices for frame materials

Characteristic:	σ_y	σ_c	E	ρ	$\frac{\sigma_y}{\rho}$	$\frac{\sigma_c}{\rho}$	$\frac{E}{\rho}$
Units:	**MN m^{-2}**	**MN m^{-2}**	**GN m^{-2}**	**kg m^{-3}**	**kN m kg^{-1}**	**kN m kg^{-1}**	**MN m kg^{-1}**
Material:							
Alloy Cr—Mo steel	700	700	210	7870	88	88	26
Aluminium alloy	350	350	70	2800	125	125	25
Titanium alloy	650	650	105	4500	144	144	23
Carbon-fibre composite	500	200	60	1100	**454**	**181**	**54**
Magnesium-based alloy	300	300	45	1780	176	176	25

You can see the great potential for carbon-fibre composites and the strong competition between the other frame materials, particularly for the deflection-based index E/ρ. However, there are three major limitations that need consideration before we all go out and start manufacturing carbon-fibre bike frames for a living:

- Firstly, because the frame is subjected to much more complex load patterns than axial tension and compression within ties and struts, it turns out that additional merit indices are required.
- Secondly, the figures for carbon-fibre composite are based on idealised production conditions where the optimum amount of carbon fibres can be reliably incorporated into the appropriate matrix, whereas the figures for the metallic alloys are those that can be expected regardless of manufacturing conditions.
- Finally, the techniques needed to manufacture advanced composites in complex three-dimensional shapes with good surface finish are extremely expensive, so the external factor of cost limits the potential market to the successes and failures covered in Box 10 Carbon-fibre composites to win at all costs .

Box 10 Carbon-fibre composites to win at all costs

The development of commercially viable carbon fibres for engineering purposes was only made possible because large quantities, supplied to the sports goods industry, sustained early progress and allowed prices to fall to acceptable levels. The fishing rod and golf club shaft industries have to be thanked for supporting the manufacture of very expensive early production quantities. The high ratios of tensile strength to weight allowed golf club shafts to deflect to higher values without snapping. For some players this increased the distance of their tee shots. The same properties were also very attractive for whippy fishing rods that had higher strengths than equivalent bamboo and glass-fibre predecessors, Figure 27.

Figure 27 The use of carbon-fibre composites gives enhanced performance in competitive sports such as fishing and golf

SAQ 10

Identify the types of need for which carbon-fibre tubes are attractive solutions.

Answer

The need is for lighter tubular structures that have the same, or better, strength and stiffness characteristics as compared with conventional materials.

More recently, carbon-fibre composites have been used to make other simple structures that benefit from the high *E*/ρ ratio for the material. Lightweight products that give minimal deflections include wing sections for aerospace vehicles and racing cars (Figure 28), and 'roach poles' for fishing, which make it possible to use a particular technique to reach further across wide stretches of water than was previously manageable.

Figure 28 Formula 1 racing cars make extensive use of carbon fibres

Very recently there have been some great successes and failures associated with complex three-dimensional carbon-fibre composite products. In cycling, a carbon-fibre composite frame which is very light and very stiff has been found not to be indestructible – at least one has failed in an accident under conditions that a metal frame might have survived.

Exercise 4

Many short stubby struts, or 'chocks', used for supporting dry-docked ships, stored goods, vehicles and the like, need to be made as cheaply as possible. Figure 29 shows an example of a cardboard pallet. Table 5 gives values of crushing strength ρ_c and cost C in euros per cubic metre for some materials.

(a) Derive a merit index that increases as the total cost K of supporting compressive load L with a strut of height h decreases. *Hint*: You will need to introduce the cross-section area A into the total cost and the maximum load – you can then eliminate it from a combination of the two expressions.

(b) Calculate the index for the candidate materials in Table 5, and select the cheapest option.

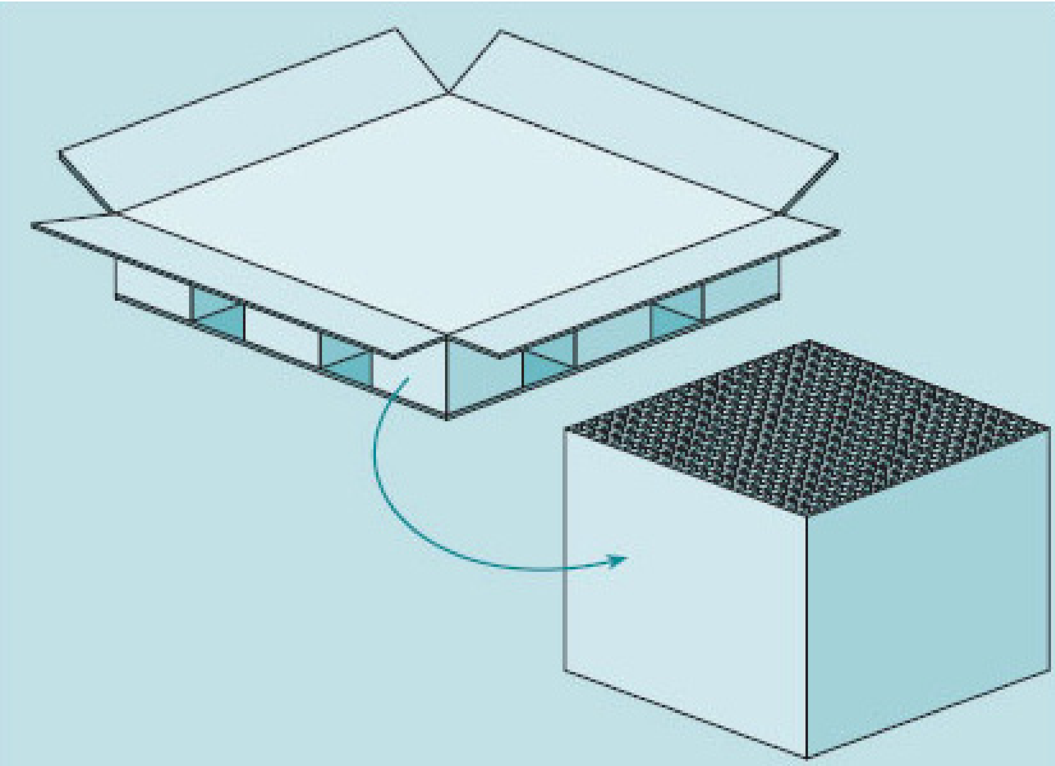

Figure 29 Cardboard pallet supports a load and can be lifted by a fork-lift truck. Inset shows an enlargement of one of the cellular blocks

Table 5 Characteristics of some candidate materials

Material	Crushing strength σ_c/MN m^{-2}	Cost C/€ m^{-3}
Softwood	7	40
Hardwood	15	20
Mild steel	300	1500
Recycled thermoplastic	18	560
Cellular cardboard	0.7	15

Answer

The cross-section *A* needed is given by:

$$L = \sigma_c \times A$$

The total cost *K* is given by:

$$K = C \times h \times A$$

The number 1/*K* will increase as the cost *K* goes down.

Rearranging the equation and substituting for *A* gives:

$$\begin{aligned}\frac{1}{K} &= \frac{1}{C} \times \frac{1}{h} \times \frac{1}{A}\\ &= \frac{1}{C} \times \frac{1}{h} \times \frac{\sigma_c}{L}\\ &= \left(\frac{\sigma_c}{C}\right) \times \left(\frac{1}{hL}\right)\end{aligned}$$

Hence for a fixed load *L* and strut height *h* the required merit index is *σc/C*. Hence, not surprisingly, materials with a high crushing strength to low cost per unit volume are preferred. You should note that this analysis does not place a limit upon the space required by the cross-section of the strut *A*, which can get large for low-density materials. I have calculated the merit index for each of the materials, Table 8.

Table 8 Merit index σc/*C* for the candidate materials

Material	σ_c/C/ MN m €$^{-1}$
Softwood	0.17
Hardwood	0.75
Mild steel	0.20
Recycled thermoplastic	0.03
Cellular cardboard	0.05

So at these prices the preferred chocks are hardwood, followed by mild steel, hence their prevalence in commerce for such tasks. Note that, although the crushing strengths are unlikely to change, the relative prices can change in response to local availability, which can influence the merit index.

Returning to the analysis of the bicycle frame, although the frame shown in Figure 25 could function, its performance would be limited to resisting forces in the vertical plane of the frame. Unfortunately, it is essential that frames resist the out-of-plane forces that are generated when a cyclist leans the bike over for hill climbing, sprints and cornering as shown in Figure 30.

Figure 30 High out-of-plane forces are generated when a bicycle is inclined to the vertical

This is when the maximum stresses are generated, as other forces add to the rider's own body weight. Flexible tie-wires and even thin rods have no resistance to such bending forces and so must be replaced by solid cantilever devices that have bending and torsional resistance, to limit the deflection shown in Figure 31. The pin joints are also eliminated to add to this bending resistance.

Finally, as can be seen in Figure 31, the rear triangle is divided to allow the rear wheel to be centrally located, and the front forks are usually divided in a similar way to produce the familiar bicycle-frame configuration. Such deflections are an indication of the overall frame stiffness.

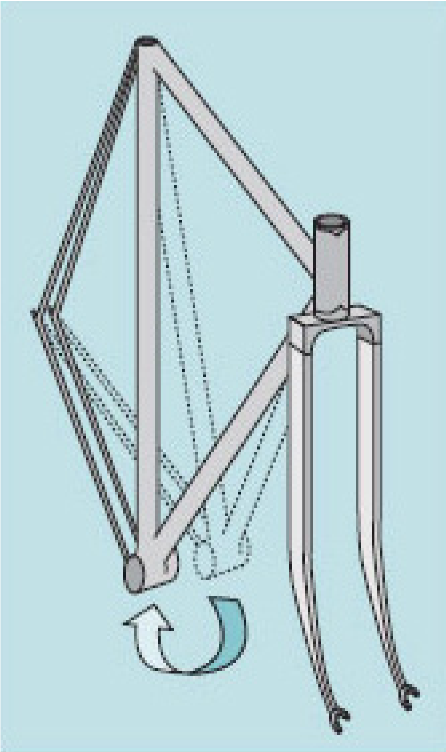

Figure 31 The frame deflection produced by out-of-plane forces (exaggerated)

In this section, although we have revealed the potential for using different materials, we have not found a 'best solution' for a lighter frame. In fact, we have done what often happens in the search for solutions. We have refined the problem and demonstrated our need to know more technical background, especially on the behaviour of loaded structures.

6 A problem with sensors

The problem we will look at in this section concerns the analysis of the design of a component used in cars that are fitted with airbags. The airbag has to be inflated rapidly when an electronic circuit in the system decides that a serious collision is taking place. The crucial component in the electronics is the accelerometer, which therefore has to be extremely reliable. Motor manufacturers have turned to a technology called MEMS (micro-electromechanical systems) for these accelerometers, because it enables large numbers of devices to be made at low cost, but with fantastically high reliability. The sensors are made on silicon chips, using the same manufacturing methods and equipment as electronic chips, the difference being that the results are mechanical structures rather than transistor circuits. Figure 32 shows an example of such a sensor. Notice the scale of the device.

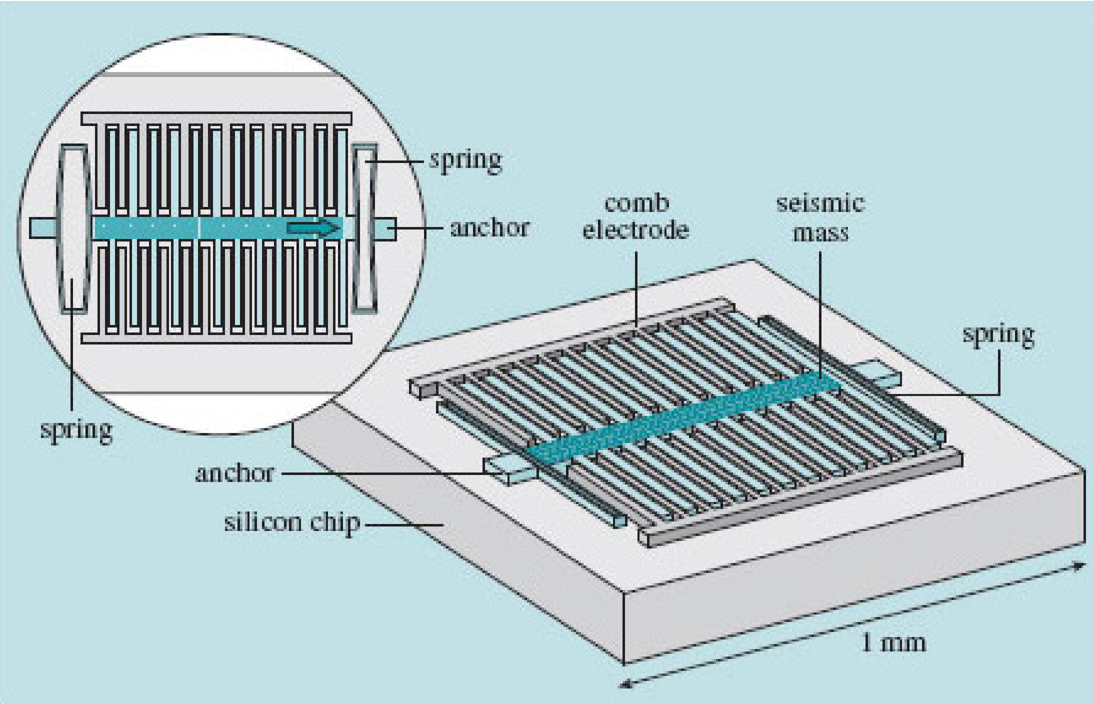

Figure 32 Detail of a typical MEMS accelerometer

Most airbag accelerometers are of the type shown in Figure 32. They consist of a silicon chip, into which the sensor and the sensing structure are fashioned. It is made entirely of silicon and is in two parts: the first is a lump (often called the proof mass or seismic mass) suspended by means of a spring formed at each end; and the second is a pair of fixed sensing electrodes that enable the electronics to detect the movement of the lump relative to the surrounding platform of silicon.

The way it works is like this: when the chip is subjected to an acceleration, the lump moves a little relative to the chip and the fixed structures on it, in the same way as your shopping might fall off the back seat of the car if you brake hard. The amount of movement depends on the size of the acceleration, the stiffness of the springs, and the mass of the lump. When the lump is deflected, the electrical capacitance between it and the sensing structures on the chip changes, and this change is detected by the electronics, which converts it to a value for acceleration.

From the point of view of building prototypes and mock-up devices to test and refine the design, the trouble with MEMS is that the things you make are very small – too small to poke with a finger to see how they're working, and too small to measure directly how much they move when the acceleration is applied. You can't even build a scale model and make the measurements on that instead, because the material properties don't all scale up in the same way. Crucially for the accelerometer, the mass is proportional to the length-scale cubed (because mass is directly related to volume, not length), but the stiffness of the support springs would scale only in proportion to their length. Therefore, you would have to build your scale model out of a different material from the silicon of the real device if you wanted to mimic its behaviour on a magnified scale whilst maintaining the same ratio of mass to stiffness. A material with the right combination of properties probably doesn't exist.

You can go some of the way towards being sure that your design will work just by doing hand calculations. For instance, you would be able to calculate the stiffness of the springy support structure by using the appropriate formula for a beam of that type. This would enable you to estimate how far the mass would move under a given acceleration. Things get much more difficult if you want to predict how much it will bend if subjected to a sideways acceleration, because the manufacturing process demands that it has lots of holes in it. This makes the structure very complicated, and the standard equations for stiffness of uniform beams don't apply.

So, to test different designs of accelerometer, it looks as though you may have no choice but to make some for real. Unfortunately, the set-up costs for small runs of MEMS devices is very high (electronic chips are cheap only because millions of them are made at the same time).

The way out of this is to use finite element analysis (FEA) to check as many aspects of the device's behaviour as possible before spending money on building any.

FEA is most commonly used to find out what happens to a structure when a mechanical load is applied to it, but it has many more applications. It can be used to predict the temperature distribution in a central heating boiler, the blood flow patterns around an artificial heart valve, the acoustics of a loudspeaker, or the magnetic fields in an electric motor. In short, anything where there is an interaction between a field (e.g. temperature, magnetic, electrostatic, acoustic, flow, force) and an object.

FEA solves the difficult differential equations that are involved by breaking the problem up into many smaller, but related problems. A computer is used to solve a huge number of Box 11 Simultaneous equations , and the solution to the whole problem is presented as a visual display in two or three dimensions. Figure 33 shows a computer model of the accelerometer, ready for analysis by the FEA program.

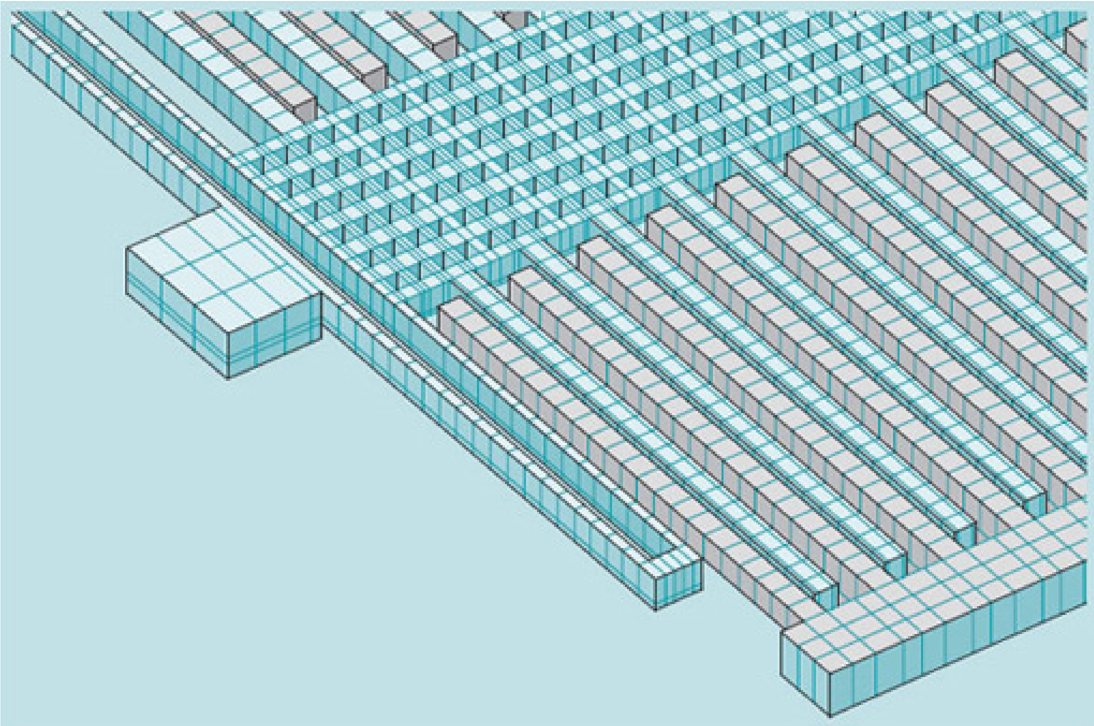

Figure 33 Seismic mass of MEMS accelerometer, showing meshing

Box 11 Simultaneous equations

I'm going to present a simple example of simultaneous equations, to illustrate why they crop up in FEA, and why we need computers to solve them even though the equations themselves may be quite straightforward. First, to remind you about simultaneous equations: they occur whenever you want to find the answer to a question where more than one condition has to be satisfied at the same time.

Cutting a piece of wood

A trivial example will get us started. Suppose you have a piece of wood 2.4 m long that you want to cut into two, with one piece four times the length of the other. This problem is so simple that you would do it in your head without realising that you had been solving simultaneous equations, but bear with me as I go in slow motion through the process, as it illustrates the general principle.

The two conditions that must be satisfied give rise to the two simultaneous equations that describe this problem. In ordinary language they are:

> 'The length of both pieces added together equals 2.4 m.'
>
> 'The length of one piece equals four times the length of the other piece.'

If you call the length of one piece x, and that of the other y, then:

$$x + y = 2.4$$
$$x = 4y$$

By substituting the second equation for x into the first, we can rewrite the first equation like this:

$$5y = 2.4$$

and we quickly find that

$$y = \frac{2.4}{5} = 0.48$$

Now that we know *y*, we can use either of our original equations to find that

$$x = 1.92$$

Our results for *x* and *y* are in metres, of course.

A system of springs

We need to take this example a step further to give an insight into why we get simultaneous equations in FEA. If you reformulated the woodcutting problem in terms of springs and forces, you could say: 'I have two springs, but one is four times as stiff as the other. They are linked end to end, and the pair is anchored at one end (Figure 34). I move the free end by a distance of 2.4 mm; how far has the point where the two springs are linked moved?

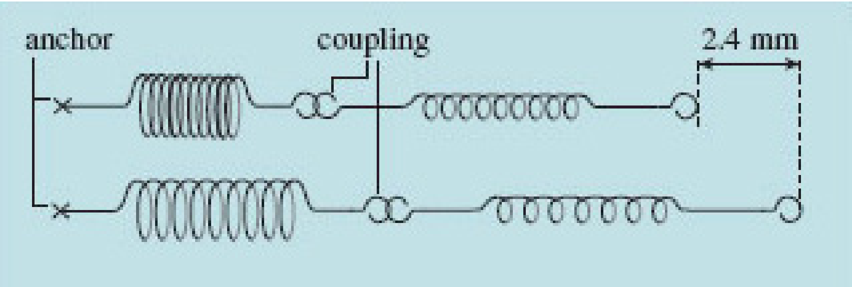

Figure 34 Linked spring system, showing extension

The force that has been applied to move the free end of the pair of springs is transmitted throughout the spring system, so we know that the force acting on each spring is the same. In a spring, the amount it extends and the force pulling it are related by the stiffness, through the expression $F = kx$, where F is the applied force, k is the stiffness, and x the extension.

We have said that the stiffness k_2 of one spring is four times that of the other, k_1. We can write this as:

$$4k_1 = k_2$$

We have just said that the force experienced by each spring is the same, so we can also say that

$$k_1 x = k_2 y$$

where *x* is the extension of one spring, and *y* that of the other. We can get rid of the *k* terms by saying

$$k_1 x = 4k_1 y$$

and so

$$x = 4y$$

This is the first of our simultaneous equations, and it says 'the amount the first spring stretches is four times as much as the second spring'.

The other equation states that the extension of both springs added together is 2.4 mm.

So,

$$x + y = 2.4$$

We needn't go through the maths because it is exactly the same as the wood-cutting example (except that our results will be in millimetres).

The point between the two springs and the end where the force is applied are effectively nodes in a simple finite element mesh. The springs are just representations of the stiffness of the material. What we have just solved is a small finite element problem. In a real problem, the numbers of calculations that need to be made are much bigger, and we may have 10 000 nodes and 30 000 simultaneous equations, which is why we use a computer. The calculations are generally rather more complex than in the spring example, because usually we are dealing with continuous materials, not 'lumps' like the springs, and there may also be non-linear behaviour.

If instead of two springs we had ten, each with a different stiffness, we could fairly comfortably still solve this by hand, to get the new positions of each node. But what if the springs extended in two or three dimensions, as in Figure 35?

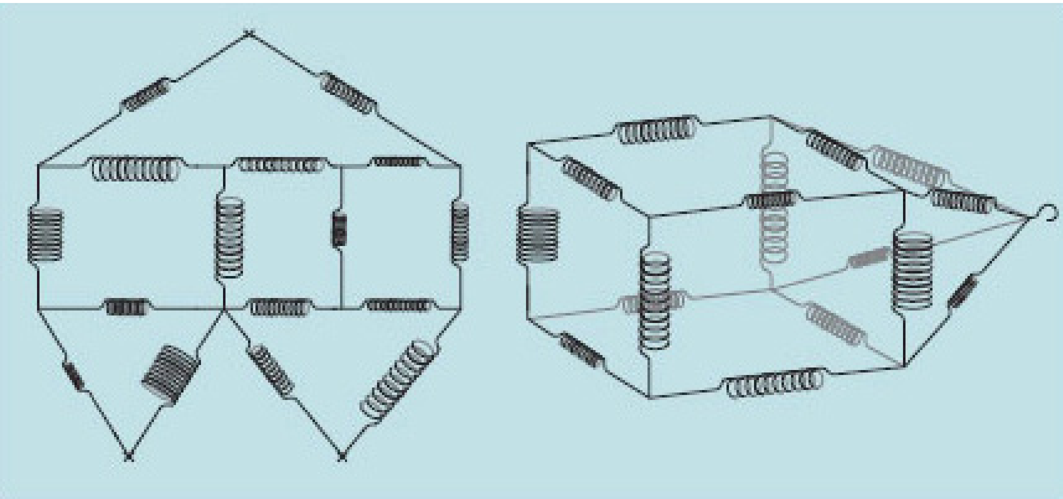

Figure 35 2-D and 3-D arrays of linked springs

The array of springs now resembles a mattress (Figure 36). Imagine a weight were placed on it. You can see how this would now be very awkward to calculate, because the depression of the mattress would cause some of the springs representing the upper and lower fabric skins to change the direction in which they pull, according to how close they are to the weight. The awkwardness comes in the large number of interdependent calculations (simultaneous equations) and that is where a formal approach and the power of a computer are able to come to the rescue.

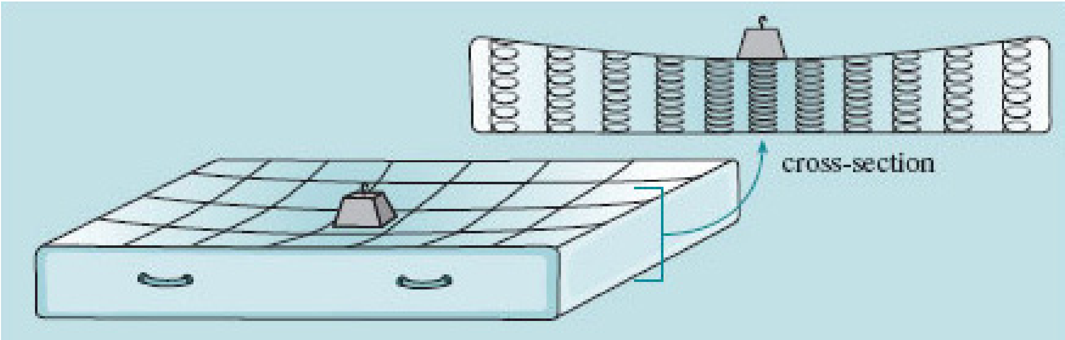

Figure 36 Cross-section of a mattress with a weight on it

Returning to finite element analysis, the structure (in the case of the accelerometer, the mass and springs) is first divided up into a large number of small blocks, or *elements*. The size and shape of these elements can vary. They are made to be much smaller than any features of the structure near them. Usually, their form is tetrahedral or hexahedral (i.e. with four or six faces), but the essential properties they have are that they completely fill the volume of the structure, and that they are connected to their neighbours at their vertices. These elements are usually not regular shapes – they have to be distorted to fit the geometry of the structure being analysed, which could have any shape. Within reason, this distortion does not matter, provided that the elements are properly connected to their neighbours.

If enough elements are used, the continuously varying quantity that is to be determined (in our case the displacement) can be approximated into simpler variations within each element (for example, a linearly varying displacement across the element). According to their position in the structure, each element is assigned material properties. These properties are used to solve, for each element, what is happening within it. Because each *node* of each element is shared with neighbouring elements, the whole assembly of elements is linked, and the solution arrived at for each element is consistent with those arrived at for all its neighbouring elements.

The process of dividing the structure into these discrete elements is called *meshing*. The size of the elements in a mesh needs to be considered carefully: if the elements are too large relative to the structure, the result of the analysis will be inaccurate; if they are too small, the analysis may take far too long for the computer to execute. In modern FEA software packages, much of the routine and arduous work of meshing has been taken away, so that the meshes are generated automatically by the software, once the user has answered some questions such as what type of element they want to use.

But all we have done so far is divide the structure into elements and told the computer what they are made of. The program needs to be set going somehow, and this stage is known to FEA practitioners as 'setting the *boundary conditions*'. The computer needs to be told what other things are known about the problem. This is in a quite literal sense setting out what is happening at the boundaries or edges of the structure. In the case of the accelerometer, you would apply a direction and magnitude for an acceleration. Another boundary condition would be where the mass is attached to the silicon chip, and what sort of attachment this is. Is it fixed in position but free to rotate about one or more axes, or can it slide in one direction?

The deflection of a material, and the stress generated due to an applied load, are simply related to one another by the stiffness of the material. Therefore, solving the problem for deflection also provides the solution for stress. Figure 37 shows the results of running a finite element analysis of a MEMS accelerometer seismic mass subjected to a sideways acceleration. The colour shading shows where the stresses are concentrated.

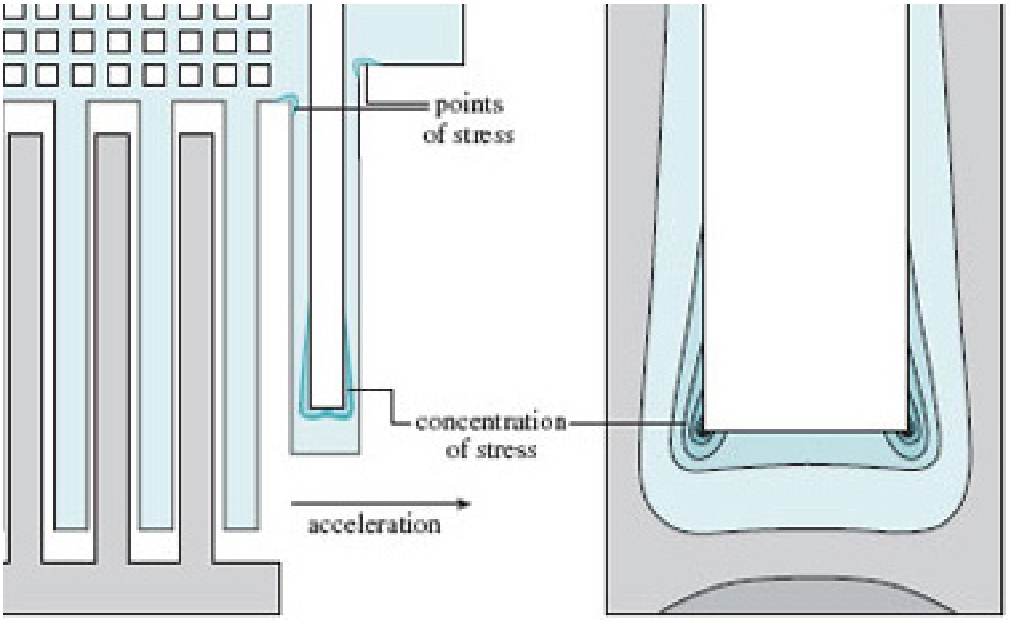

Figure 37 FEA results showing stress distribution on MEMS accelerometer

One important thing to understand about FEA is that there are many opportunities to create a result that looks convincing, but is completely incorrect. Meshing is one such opportunity – usually where the mesh is too coarse to allow accurate calculation. This is most likely to happen near features in the structure, such as holes and corners, as in Figure 38. This is why in a properly meshed structure the element size is smaller around such features than elsewhere.

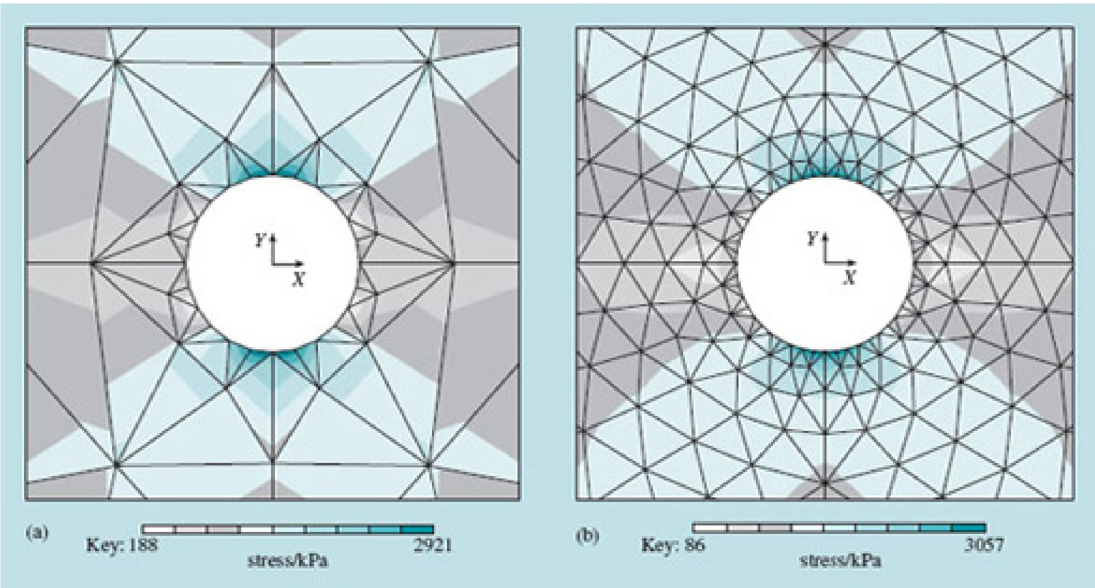

Figure 38 FEA results for stresses around a circular hole in a square plate, using different mesh sizes. Notice the difference in the results: (a) is inadequately meshed

It is good practice when doing FEA to run what is called a meshing analysis. This is where the same structure and boundary conditions are run with three or four different levels of mesh refinement, and the results are compared with one another. You can have reasonable faith in the quality of a mesh when the answers are very similar for two different levels of refinement of the mesh. But it is possible to go too far: if the elements get so small as to approach the size of the microstructure of the material (e.g. grains in a metal), then the implicit assumption in FE modelling – that the material is continuous – breaks down.

Another common source of error is incorrectly specified boundary conditions; for example, the wrong type of constraint at anchor points. In the accelerometer, the anchor points are where the springs are attached to the substrate. These should be defined as constraining motion in all three translations and all three rotations.

All this points to a need for independent verification of the results, and this often involves doing real tests on real structures. Ordinary hand calculation is useful too, because even if it doesn't give you accurate data, it is normally possible to get an estimate within a factor of two or three of the right answer, and this allows you to spot really gross errors in the FEA results. The FEA software companies also provide large numbers of worked examples of standard problems to allow you to check that your model behaves correctly.

To sum up these thoughts on the shortcomings of FE analysis, we can say that what we are working with is only a model, and models by their nature are never exactly like the real thing. The skill of the engineer lies in knowing how far the model can be stretched and how deeply probed, and yet still yield information about the real world.

SAQ 11

Where does FEA fit into the problem-solving map in Figure 7?

Answer

FEA is a sort of mathematical model. It is important in the evaluation of possible solutions. It can also be used to test a demonstrator before building it.

7 Responsible engineering

7.1 The engineer and society

Section 2 outlined some of the needs for engineering. Society relies on engineers to create solutions to the problems involved in meeting those needs.

This is a good time to pause and point out that inevitably, in return for all this fun and power, engineers have a responsibility to society. The people who employ our services, directly or indirectly, have to have an assurance that we are working within certain social, safety and ethical boundaries. Particularly given the increasing trend in the Western world towards litigation, it is in our own best interest to uphold this responsibility.

In considering the responsibilities of engineers, this section also provides an opportunity for putting the whole course in context.

7.2 The professional engineer

It has been suggested that there are four main criteria that identify a profession:

> Custody of a clearly definable and valuable body of knowledge and understanding associated with a long period of training.
>
> A strong unitary organization which ensures that the profession generally speaks with 'one voice'.
>
> Clearly defined and rigorous entry standards, backed up by a requirement to register with the professional association.
>
> An overriding responsibility to maintain the standards of the profession for the public's benefit.
>
> Collins, Ghey and Mills (1989)

It is the role of the professional engineering bodies (Institutes, Institutions and Societies) to ensure that there is a focal point, and to coordinate the profession. A key issue within this role for a professional body is the support of the continuing professional development (CPD) of its members. This is vital for keeping pace and ensuring safety in a world where new technologies are developing daily. As a contribution to that process this course is aimed at extending your knowledge and awareness over a broad range of engineering activities.

7.3 Ethics and safety

A practising engineer makes ethical decisions, with moral and physical implications of varying magnitudes, on a daily basis. Examples of ethical dilemmas are limitless, ranging from the engineer who takes home the odd pen, file or discarded paper 'for the children', to the engineer who signs off a project without checking the details and identifying a simple arithmetic error of magnitude. The implications of either may be negligible – such as where the cost is more than compensated in unpaid overtime, the error merely accidentally increases the factor of safety – or catastrophic, such as when a discarded piece of paper has sensitive industrial information that ends up with a major competitor, or an arithmetic error decreases the factor of safety and a component fails in use at the cost of human life. For the occasions when the ramifications of our decisions are not apparent to anyone else, then ethics are a matter of personal conscience. However, when the ripples of our actions spread out and cause damage or injury then we are legally responsible for the result. Very often, the difference between the two is a matter of luck.

The very nature of engineering implies that safety must be a primary issue. Even the most remote of robots will have some human interface somewhere along the line, and most engineering design, whether industrial or domestic, requires direct contact at one or more levels. Ethics and safety are often closely interwoven – our responsibility for safety in design is as much moral as it is professional – and there are safety practices to be observed at every stage of the design process.

Much of what we know now has been learned from bitter experience but, amazingly, evidence suggests that we are still inclined to become complacent over long periods of technological triumph, leading us to more narrow margins of safety and, ultimately, repeated disaster. Consider what you know of the most publicised engineering disasters over the last century, and how safety was compromised in each case. Often, these great catastrophes are the result of some very minor error, and not the technological billion-to-one misfortune we might hope to believe – see Table 6.

Table 6 Causes of some notable engineering disasters

Date	Disaster	Fundamental cause
2000	Concorde: fire and crash, shortly after take-off (113 dead)	Debris on runway and fuel tank susceptible to damage from same
1986	Chernobyl: meltdown of nuclear reactor core, and large-scale radioactive contamination	Safety procedures ignored, and design flaws
1986	Challenger Space Shuttle: exploded 73 seconds into flight (7 dead)	Design flaw in O-ring seals on the booster engines
1981	Hyatt Regency Hotel: suspended catwalk collapsed over a dance floor (114 dead)	Design change and failure to anticipate overload
1979	Three Mile Island: 51 per cent meltdown of nuclear reactor core	Incorrect procedures
1940	Tacoma Narrows Bridge: bridge collapsed	Unexpected wind-induced vibrations

The study summarised in Table 7 investigated 800 cases (and millions of pounds worth) of structural failure, in which 504 people died and 592 were injured. When engineers were to blame, the study categorised the causes of failure (and hence breaches in safety).

Table 7 Causes of failure, where engineers were to blame*

Insufficient knowledge	68%
Underestimation of influence	16%
Ignorance, carelessness or negligence	14%
Forgetfulness, error	13%

Relying on others without sufficient control	9%
Objectively unknown situation	7%
Imprecise definition of responsibilities	1%
Choice of bad quality	1%
Other	3%

Source: Swiss Federal Institute of Technology

*Note that the percentages add up to more than 100 – some failures were attributed to more than one cause.

You can see that in a whopping 68 per cent of cases, 'insufficient knowledge' on the part of the engineers was a contributing factor. Again, this has to be an ethics issue – can we really accept that all these engineers were so lacking in self-awareness that they truly believed in their own abilities, or were some of them just not brave enough to admit they were out of their depth at the time? The lesson is clear. You don't need to store everything you study in a photographic memory compartment, but it *is* essential to remember that, as a professional engineer, you are accountable for your actions; and this includes recognising when you need to bring in expertise from a colleague or external sources.

7.4 The impact of technology on society

Engineering is apparently driven by the needs of society. The technology that results, in turn, drives other changes in our everyday lives. One of the basic needs identified in Section 2 was for shelter. There are many fine examples of long-surviving structures such as pyramids, aqueducts, bridges, walls, functional buildings, and so on. Remarkably these constructions were completed without the depth of analysis and understanding that is available today (though we don't necessarily know much of the failures). The challenge to be more efficient in terms of space, materials, cost of ownership, etc. gets harder every year. Understanding the properties of static structures is important in creating tomorrow's solutions.

Conclusion

We have seen how a solution falls into one of three categories (innovation by context, innovation by development, and routine solution) according to the need that drives it. Furthermore, the need is shown to be the point of reference that should be kept in sight throughout the process of finding solutions. Unless the need is accurately stated, the ideal solution cannot be obtained – a case of 'garbage in, garbage out'.

We have examined the process of finding a solution step by step, using examples to help us see where and why particular approaches are most appropriate at various stages, such as for instance the best kind of modelling to use. Sometimes it is enough to make some rough calculations by hand in a few minutes, but at other times this is not sufficiently accurate. So at the other extreme, a physical mock-up or computer-aided modelling technique such as finite element analysis may be needed to provide the necessary data.

The bicycle design example enabled us to explore the idea that a solution is always a compromise, but that the best compromise can be found by the use of quantitative tools such as merit indices.

We saw that the solution-finding process generally contains loops, where certain steps are repeated until an acceptable result is obtained. The important point to note about this is that a trip round one of these loops (so long as it's not the loop that leads back from the very end of the process to the very beginning!) is not a failure, but a means of refining the solution.

Finally, we looked at the engineer in the context of wider society and saw that engineering has been central in providing the high quality of life enjoyed by many. The other side of this picture is that a significant proportion of the spectacular disasters we have witnessed involving failure of components have been attributable to poor work on the part of engineers. This lays a heavy responsibility on our shoulders to make sure we know what we're doing.

References

Collins, S., Ghey, J. and Mills, G. (1989) *The Professional Engineer in Society*, London: Jessica Kingsley Publishers.

Foster, J., with Corby, L. (illustrator) (1996) *How to Get Ideas*, San Francisco: Berrett-Koehler Publishers, Inc.

Haaland, J., Wingert, J. and Olson, B. A. (1963) 'Force required to activate switches, maximum finger pushing force, and coefficient of friction of mercury gloves', Honeywell Memorandum.

Pheasant, S. (1986) *Bodyspace: Anthropometry, ergonomics and the design of work*, London: Taylor and Francis; as reproduced by Kothiyal, K. in *Ergonomics, Productivity and Safety*, [accessed 01 August 2002].

Acknowledgements

This free course is adapted from a former Open University course called 'Engineering: mechanics, materials, design (T207).'

Grateful acknowledgement is made to the following sources for permission to reproduce material within this course.

Figure 3 Courtesy of Black & Decker; Figure 5: © NASA; Figure 9: © Ecoscene/Ray Roberts; Figures 14 and 15: Courtesy of Glynne T. Bowsher; Figure 21: Courtesy of the The Elgar Birthplace Trust; Figure 22: Courtesy of Falcon Cycles Ltd; Figure 27: Courtesy of Carbotec UK; Figure 28: Courtesy of McLaren; Figure 38: Courtesy of Ray Matela, Open University.

Table 2: Pheasant, S. (1986) *Bodyspace: Anthropometry, ergonomics and the design of work*, Taylor and Francis.

Every effort has been made to contact copyright owners. If any have been inadvertently overlooked, the publishers will be pleased to make the necessary arrangements at the first opportunity.