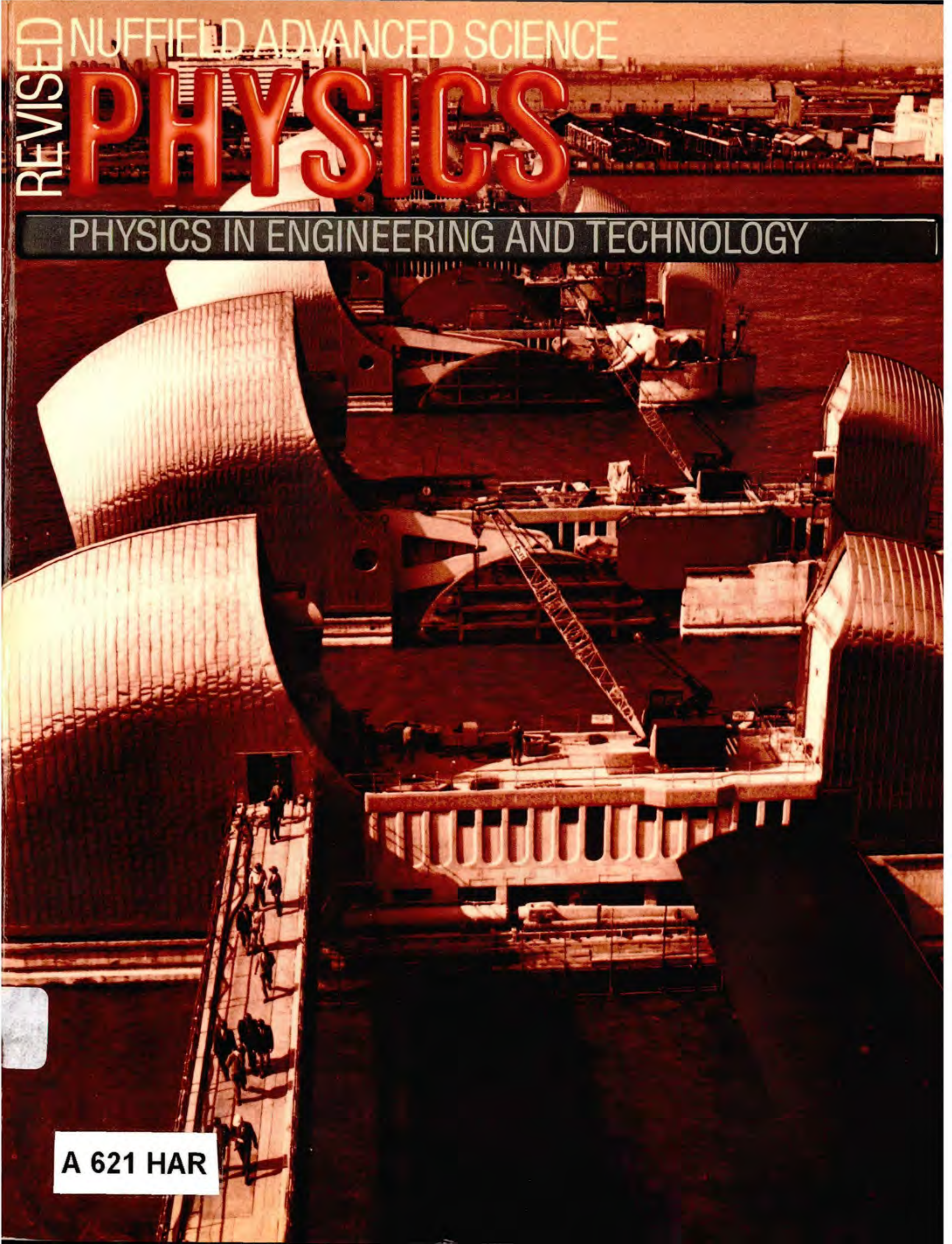
REVISED

NUFFIELD ADVANCED SCIENCE

PHYSICS

PHYSICS IN ENGINEERING AND TECHNOLOGY



C3PY31

PHYSICS
PHYSICS IN ENGINEERING AND TECHNOLOGY

REVISED NUFFIELD ADVANCED SCIENCE
Published for the Nuffield–Chelsea Curriculum Trust
by Longman Group Limited

General editor
Revised Nuffield Advanced Physics
John Harris

Consultant editor
E. J. Wenham

Editor of this book
John Harris

Contributors
Steve Banks
A. E. De Barr
David Fisher
D. E. J. Walshe
L. R. Wootton

Longman Group Limited
Longman House, Burnt Mill, Harlow, Essex CM20 2JE, England
and Associated Companies throughout the World

Design and art direction by Ivan Dodd
Illustrations by Oxford Illustrators Limited

Filmset in Times Roman and Univers
Printed in Great Britain by Scotprint Limited, Musselburgh, Scotland

ISBN 0 582 35422 6

Cover
The Thames Barrier under construction.
Costain Tarmac HBM (Holland) Joint Venture

CONTENTS

Design and construction of the Humber Bridge

DAVID FISHER
Freeman Fox & Partners, Consulting Engineers

The Humber Bridge, with a main span of 1410 metres, has the longest bridge span in the World – 112 metres more than the previous longest, Verrazano Narrows Bridge in New York. It is a suspension bridge, which is the only bridge form so far used for very long spans. A long-span bridge was necessary because the Humber estuary has a mobile bed and 1410 metres covers the likely movements of the navigation channel. Piers supporting short spans would have interfered with the natural regime of the river. The north tower is sited just above the high water line on the Hessle side and the south tower is in the estuary about 500 metres from the Barton shore line. Because of the topological and geological conditions there is a marked inequality in the lengths of the two side spans: that on the north side is 280 metres long and that on the south side 830 metres. However, the bridge is so long that the asymmetry is not readily apparent.

The bridge was built to aid industrial and social development along both banks of the estuary. Distances between the major towns in the region, for example Grimsby, Immingham, and Hull, have been cut by as much as 80 km and the bridge, together with its approach roads, forms a vital element of the integrated road system linking Humberside as a whole with the national motorway network. The structure carries dual two-lane roadways (which gives a peak capacity of about 6000 vehicles an hour or a daily capacity of about 40 000) with a combined footpath and cycle track along each side of the bridge. The underside of the road deck is a minimum 30 metres above high water to give clearance for shipping using the river. The bridge was opened by the Queen in July 1981. The whole scheme is funded by tolls and is expected to reach its traffic capacity at about the end of the century.

Design of the Humber Bridge and all the associated works were undertaken by Freeman Fox & Partners, who had been retained as consulting engineers for the project since 1928, and had proposed designs in 1930, 1935, 1955, and 1966. When the project was finally authorized by the Government in 1971, the firm began detailed design of the bridge and of the approach roads on both sides of the estuary. Under British practice the consulting engineers design the 'service' structure, that is, the completed works that will safely do the job required – in this case, carry the traffic. They also prepare the specifications and tender documents, including the drawings, that describe and define what is required; and they advise the client, in this case the Humber Bridge Board, on the appointment of contractors. The contractors have the responsibility for completing the project, with the consulting engineers supervising construction to ensure that the proposed structure is built. The consulting engineers also adjudicate on the fair price

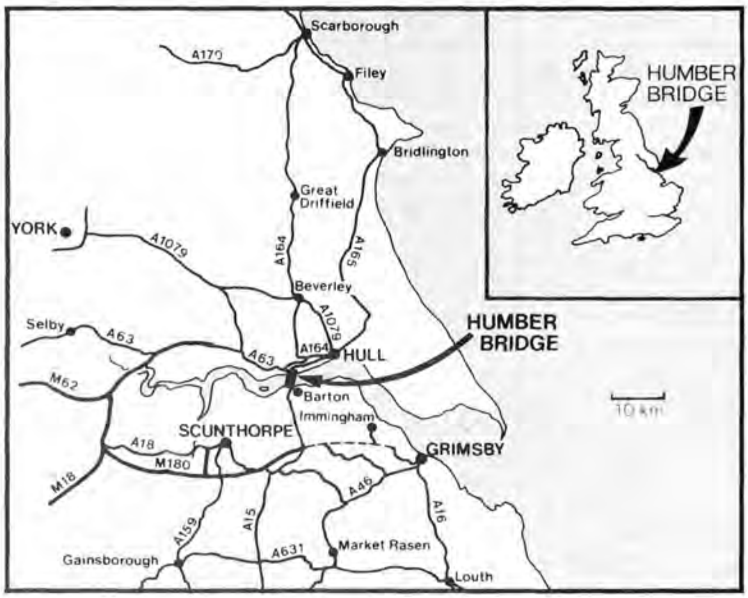


Figure 1
Location of the Humber Bridge.

to be paid to the contractors for the work done. For the Humber Bridge, the Government has loaned the Bridge Board 75 per cent of the cost, the other 25 per cent being raised in the money markets. The Board now operates the bridge and will repay the capital and interest out of toll revenues less operating costs.

The bridge has been designed to carry the highway loading intensity specified in British Standard 5400, which deals with steel, concrete, and composite bridges. The load intensity varies with the length of load considered with a minimum lane load of 8.7 $kN\ m^{-1}$; for short-loaded lengths the load intensity increases to 32.1 $kN\ m^{-1}$. In addition the bridge has been designed to carry a special vehicle with an all-up weight of 1760 kN (180 tonnes). Consideration has been given to many patterns and combinations of loading, for example, alternate spans being fully loaded with traffic on one side and empty on the other, so that the most onerous combinations of loading are catered for throughout the different elements – deck, suspenders, cables, towers, anchorages – of the bridge.

Wind loading and its effects are of major importance for a bridge of this span and the design caters for wind speeds of approximately 47 $m\ s^{-1}$ (169 $km\ h^{-1}$) on the deck structure and, for the towers, a speed that increases with height to a maximum of 66 $m\ s^{-1}$ (238 $km\ h^{-1}$). Of even greater significance than the highest wind loadings are oscillations due to dynamic instability (always remembering the case of the Tacoma Narrows Bridge, 'Galloping Gerty'). The deck girder for the Humber Bridge has a streamlined cross-section of the type previously used successfully by the designers, Freeman Fox & Partners, for the Severn Bridge (main span 988 metres, completed in 1966) and the Bosphorus Bridge (1074 metres, 1973). Wind tunnel tests were carried out at the National Maritime Institute at Teddington on models of the proposed Humber deck to establish its aerodynamic stability and to determine lift and drag coefficients (figure 3). Similarly, tests were also made on a model of a tower to determine drag coefficients for various angles of wind incidence and the possibility of tower oscillation during erection (when the tower is free standing without any restraining effect by the cables of the completed structure).

The forces involved in the structure are very large. For example, the mass of the deck structure, before the application of live load, is about 21 000 tonnes, and the maximum pull in each main cable is 1.9×10^9 N. From these and other primary loads and the geometry of the structure, stresses can be calculated for any point in the structure. However, an additional complication of any suspension bridge is that it does not respond linearly to applied loads. Furthermore, good judgment is required to assess what loads will, or will not, occur simultaneously – for example, it

Figure 2
Humber Bridge viewed from the south approach road on the Barton side.
David Lee Photography Ltd.

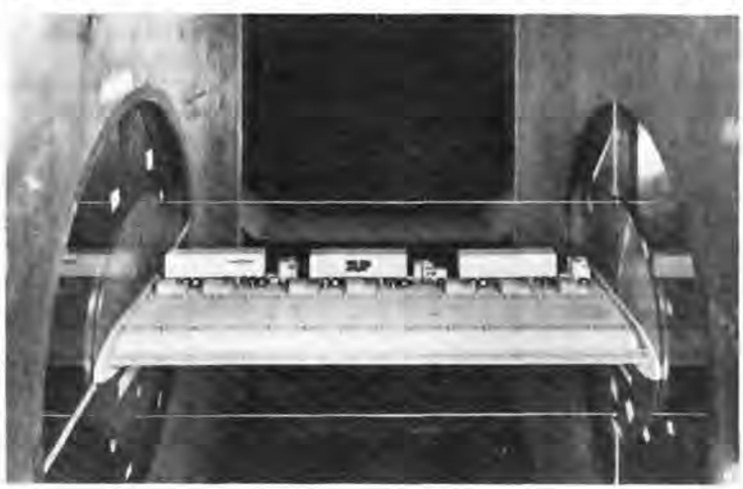


Figure 3
Model of a loaded section of the deck under test in the wind tunnel at the National Maritime Institute.
National Maritime Institute, Teddington.

is improbable that maximum traffic loads will be experienced at the same time as very high wind speeds – however the worst likely combination has to be catered for. Over-conservatism would lead to the use of excess material and an uneconomical structure. Great care is given by the designer to the refinement of structural details since it is upon these items that a trouble-free life for a bridge so much depends.

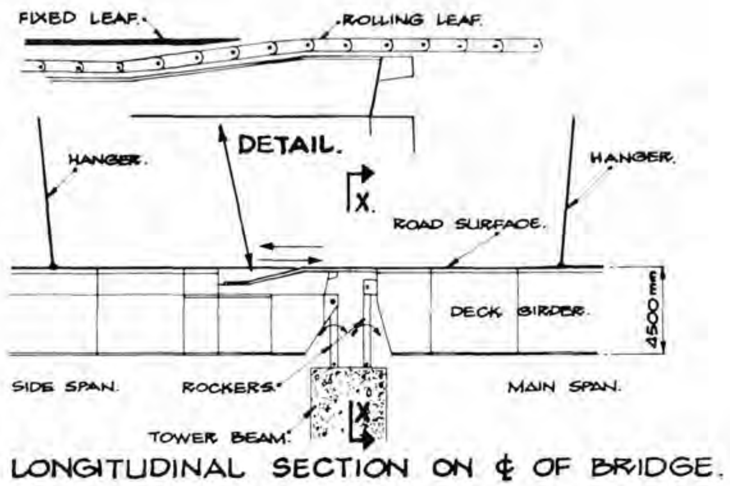
Fluctuating stresses

Bridges are now designed to have a normal life of 120 years and appropriate regard has to be paid to components subjected to fluctuating stresses so that proper allowance is made for fatigue. It is in matters such as this that sound engineering experience and appreciation of structural and materials behaviour are of great value in tempering the results of stress investigations.

Long-span bridges are very flexible structures, subject to very large movements arising both from strain due to external load and from temperature effects. To accommodate these movements between −20 °C and +60 °C the three spans at Humber are supported at each end by A-frame rockers (two-legged hinged struts) mounted on the bottom cross-beam of the towers and on the anchorage blocks. The rockers (figure 4) will permit longitudinal movement and small changes in gradient of the deck but constrain lateral movement. At the towers continuity of the road surface above the rockers is provided by a joint of the 'rolling leaf' type which works after the fashion of a roll-top desk – see figure 4. At the Barton Tower the joint can cope with a movement of 2.8 metres, caused partly by temperature change but mainly by longitudinal movement of the cables under uneven traffic loading. At the anchorages road continuity is provided by rubber joints.

Some of the other principle movements of the structure under extreme conditions of load are as follows.

1 Maximum tower top movement occurs at the Barton Tower where, under the worst conditions of highway loading, the tower top deflects by 630 mm towards the



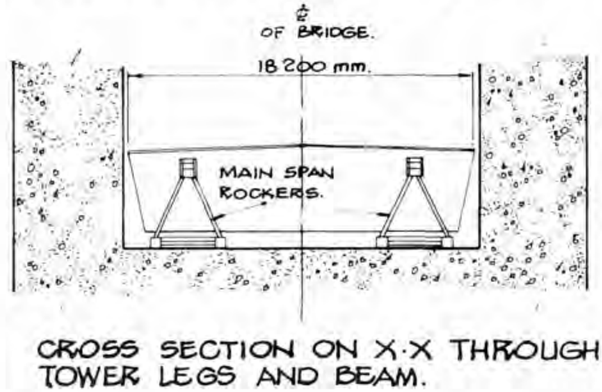
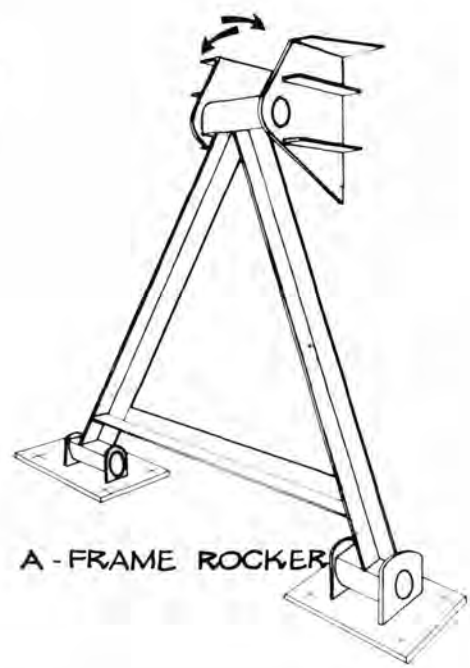


Figure 4
Diagram of the 'rockers' at the towers and detail of the expansion joint in the road surface.



main span; an increase of 20 °C would cause the tower top to deflect by a further 93 mm.

2 The maximum vertical deflection at mid-span due to the design traffic load is 3210 mm; an increase of 20 °C in the ambient temperature would increase the deflection by a further 1100 mm.

3 The maximum lateral deflection of the deck mid-span due to winds of $48\,m\,s^{-1}$ ($173\,km\,h^{-1}$) at deck level is 4510 mm; the cables would then move 4280 mm laterally.

Self-evidently, for most of the time, movements are far below these limits and, except in extreme instances, would hardly be noticeable.

Main towers

The main towers of the Humber Bridge are of reinforced concrete – a complete break from tradition. Previously, all the major suspension bridges (spans over 700 metres) had been built with steel towers. The towers, however, are essentially compression members and concrete is an inherently good material in compression. Added to this, concrete has improved immensely in reliability over the past twenty or so years, and predesign discussions with the contracting industry indicated that the recent developments in slip-forming would save construction time. On the Barton Tower the contractor (with the experience of having already built the Hessle Tower) used 12-hour shifts, seven days a week, and the two legs of the tower, 155.5 metres high, were topped out in 1750 hours ($10\frac{1}{2}$ weeks), finally at an average rise of $110\,mm\,h^{-1}$ – a notable achievement.

The towers are slightly tapered and hollow and one leg in each tower contains a service lift for maintenance purposes. The concrete is heavily reinforced. At the Forth Bridge the steel towers oscillated in the wind when free standing (before the cables and deck were in position) and on subsequent bridges with steel towers measures were taken to prevent this. At Humber, although provision was made to damp down any oscillation, no such movement was experienced. Prior to erection of the cables, the tops of the towers had to be pulled backwards by about 900 mm to ensure that they would be vertical when the deck structure was in place.

The stiffening girder at Humber is a streamlined hollow box, the upper flange of which is covered with mastic asphalt and provides the roadway surface. This shape is preferred because it greatly reduces the wind forces on the structure, at the same time enabling the required strength and stability to be achieved using less steel than with a truss girder. The box girder also bestows a number of other advantages. The smaller weight of the girder reduces the forces in (and therefore the size of) the cables and, consequently, the size of the anchorages. The towers benefit in two ways, since both the dead load and the wind load from the suspended structure are reduced. The box girder is also shallower than the corresponding truss which improves the appearance of the structure. It is also vastly easier to paint the large, flat surfaces of the boxes than the many components of a truss, with its large number of joints to clean and protect. In other words, the box dramatically reduces points where corrosion is likely to start, not least because nesting points for birds, and consequent fouling of the steelwork, are largely eliminated.

On Humber, as at Severn and Bosphorus, any incipient oscillation of the deck due to aerodynamic excitation is further checked by the 'triangulated' suspension system. Although the inclined hangers or suspenders apply some constraint on longitudinal movement of the deck, the resulting strain energy generated in each suspender by any movement is dissipated by the hysteresis characteristics of the wire ropes, wound spirally to dampen out oscillation.

Foundations and anchorages

The design and construction of the foundations of the Humber towers and of the anchorages were dictated by the geology of the site, which differs greatly between the north and south sides of the estuary. On the north side a deep bed of chalk comes to the surface and resolving foundation problems there was relatively straightforward.

On the south side, however, the chalk has been eroded by glacial action, leaving a 30-metre deep bed of boulder clay (gravel saturated alluvium) overlying a thick bed of over-consolidated Kimmeridge clay. The foundations for both the tower and anchorage had to be taken down into the Kimmeridge clay but tests had shown that the clay, when in contact with water, turned into a slurry. So it was necessary to design and construct the foundations so that the clay was exposed for only a short period, with only relatively small areas excavated at any one time.

For the foundation of the south tower the contractor first built an artificial sand island inside a steel cofferdam (a temporary wall to keep out water); severe scouring of the river bed followed and 12 000 tonnes of chalk had to be dumped in the river to provide protection. Two circular concrete caissons (open cylinders which can be sunk into the ground for bridge foundations), each 24 metres in diameter, were sunk through the sand island and into the river bed by excavating the ground below them, but the rate of sinking was much slower than expected. Further delay was experienced when the west caisson struck a pocket of underground water and the resulting inflow of water flushed away the lubricating skin of bentonite (a thixotropic clay) on the outside of the caissons, greatly increasing the friction and thus the load required to sink them. Attempts to restore the bentonite were unsuccessful and 3000 tonnes of temporary steel ingots – 'kentledge' – and 4000 tonnes of permanent concrete had to be added to each caisson before they could be sunk the required 8 metres into the Kimmeridge Clay. The bottoms of the caissons were 'plugged' with a concrete slab and the tops capped and 'bridged' to form the concrete pier that provides the base for the tower.

At each end of the bridge is an anchorage which resists the pulls of the main cables. Each anchorage comprises a massive concrete structure containing two chambers, one for each cable. Within the chamber the cable passes over a steel saddle and then divides into strands that fit over 'shoes' fixed to the back of the chamber by pre-stressed rods. The anchorage on the north bank, 65.5 metres long, 36 metres wide, and 36 metres high, is founded in the chalk 21 metres below ground level.

Figure 5
Barton anchorage nearing completion. The upper block is 65 metres long by 37 metres wide by 22 metres high. Including its 36 metres deep foundation structure. The mass of the anchorage is about 300 000 tonnes. *Handford Photography.*

The south anchorage (figure 5) consists of a main block very similar to that on the north side but built above a massive cellular box founded in the Kimmeridge Clay 35 metres below ground level. The box was built within a framework of diaphragm walls that divided the site into five longitudinal strips. These strips were excavated no more than two at a time, thereby limiting the extent of open excavation, the walls being supported by permanent concrete struts. The bottom of each trench was sealed with a concrete slab and the cellular box structure built within the protection of the diaphragm walls and backfilled with sand and water to restore ground loading. Because of the huge size of the anchorage structure itself, large amounts of concrete were used, up to 1000 cubic metres per day. To reduce the heat of hydration and lessen the risk of cracking, up to 60 per cent of the Portland cement was replaced by blast furnace slag. The anchorage, including the substructure, has a total mass approaching 300 000 tonnes and to maintain a uniform pressure on the clay and prevent any tilting of the

anchorage as the pull on the cables increased, parts of the upper block of the anchorage, including the architectural facings and the deck, were built step-by-step as the erection of the bridge proceeded.

Figure 6
Construction of a box section.
Donald Innes.

Box girder assembly and erection

The stiffening girder, or deck, at Humber is built of welded, stiffened steel plate panels that were made off site. The panels were then assembled and welded (in a disused railway yard about a mile from the bridge, figure 6) into 124 boxes, generally 18.1 metres long. These formed sections of the deck 22 metres wide and 4.5 metres deep, with 3-metre wide panels cantilevering from each side to carry the walkways. The boxes were assembled in threes, repeatedly using a completed box from the preceding batch as a template for the next two, so that they would fit closely together ready for welding when in position in the bridge. In order to position the boxes (each of about 140 tonnes) they were taken up river by pontoon and lifted into place by travelling gantries running on the main cables (figure 7).

The first boxes were erected adjacent to the two anchorages and at the middle of the main span. Erection then continued on four fronts, working towards the towers. The concentration of load at the centre of the main span caused the boxes to adopt a sagging profile that was self-correcting as erection progressed. When the boxes were correctly aligned, they were spliced together by welding. The deck was then given a mastic asphalt (a mixture of bitumen with stone chippings or sand) wearing surface.

A large number of calculations was required to analyse loads, deflections, and stresses at the different stages of erection of the bridge, when components would be lacking the support that would be forthcoming when the structure was complete. These calculations were primarily concerned with the suspended structure. For example, the most severe conditions on the tower due to wind loading occur when all the deck is suspended from the cables but not yet connected at its ends to the towers.

Prevention of corrosion is essential for all large steel structures in exposed sites. Following established practice for bridgework, and prior to assembly of the boxes, each panel was blast-cleaned and given a preliminary protective coat of primer followed by three coats of epoxy ester paint. After erection all external surfaces of the steel were given two further coats of chlorinated rubber paint. Similarly the main

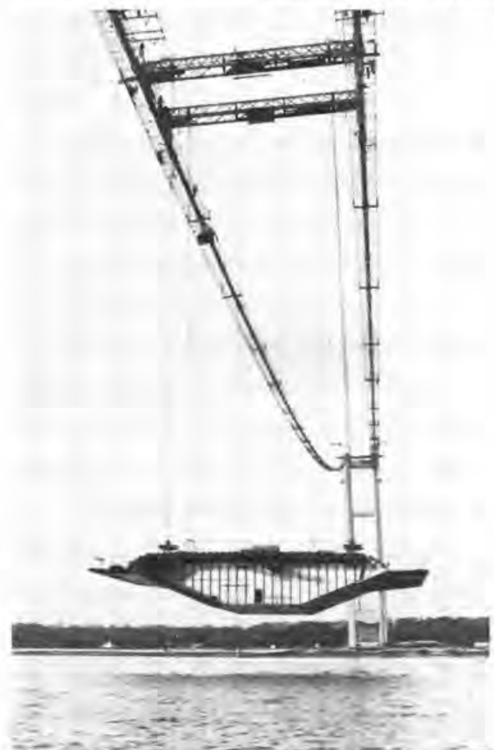


Figure 7
The first box for the stiffening girder being lifted into place. The boxes are raised by travelling gantries running on the main cables. The top flanges of the boxes form the bridge deck and, covered with mastic asphalt, provide the road surface.
Donald Innes.

cables, each of which is made up of 14 948 parallel wires of drawn steel 5 mm in diameter and heavily galvanized, have to be fully protected against corrosion. The wires were laid into place by 'aerial spinning', involving assembly of the cables four wires at a time. The cables were then compacted by a travelling ring of jacks and the cable bands, to which the suspenders are attached, put into place. With the deck in position the cables were almost fully loaded and compaction complete. The cables then received a thick coat of red lead paint and were bound with soft iron wire which formed a casing that was finally painted, along with the rest of the exposed steel.

Additional services

For efficient operation of the bridge extensive services have been installed. Their operation is directed from the central control room in the administration block where all systems are monitored and controlled. A watch is kept on traffic by closed-circuit television cameras and monitors display pictures in the control room and in the bridgemaster's office. Radio contact is maintained with patrol and breakdown recovery vehicles, and motorists' calls are taken from 30 emergency telephones sited along the bridge. The duty officer, who has an overall view along the bridge and across the toll plaza, can operate traffic signals and 'secret' signs to give warnings of speed restrictions, closed lanes, high winds, ice, and so on.

Electrical power for the bridge is obtained through two 11 kV/415 V substations at the approaches and two 3.3 kV/415 V stations in the towers. The system is also equipped with two standby generators. The main switchboard provides for a total connected load of 220 kW. In addition to lighting, communication systems, traffic signs, toll systems, and river and air navigation lights, power is supplied for the service lifts in the towers and for water pumps in the anchorages, the Barton Tower foundations, and for fire-fighting.

The toll collection system is of the 'instant audit' type, operating in lanes with the central four reversible. Each toll booth is equipped with a microprocessor which computes the toll charge from vehicle classifications entered on a keyboard, displays the charge to the vehicle driver and operator, and monitors the passage of the vehicle with the aid of detectors buried in the road surface. Recorded data are transmitted to the main computer in the central control room where information about traffic flows and cash receipts is continuously accessible on visual display screens. Toll charges can be varied by simple programming modifications and facilities are installed to introduce automatic vehicle identification and coin collection equipment if required.

From the very outset of the project the consulting engineers advised the Humber Bridge Board that a project embracing such a huge structure would have a major visual impact on the environment and, throughout the period of design and construction, the engineers sought to achieve an aesthetic grace as well as satisfying engineering and economic requirements. As well as being a fine piece of engineering, in no way does the bridge detract from the beauty of the Humber estuary.

(This article was first published in *Physics Education*, **17**, 5, Sept. 1982. It has since been revised by the author.)

Systems

A. E. DE BARR
Macclesfield

Throughout the study of physics one continually encounters examples of very different physical effects which are described and explained using essentially the same words, ideas, and mathematical expressions. Thus gravitational, electric, and magnetic phenomena are all described in terms of fields of force; optical, acoustic, and electromagnetic phenomena are all described in terms of waves. General concepts such as waves and fields of force illustrate the unity of the physical world; they are also of great help in understanding and learning about physics since knowledge acquired in the study of one branch of the subject can often be used directly to help the understanding of a different aspect of physics.

This article describes some other kinds of general concept which are useful in understanding the behaviour of physical *systems*. In the physical world nothing exists in complete isolation and everything has to be considered as part of a system – defined as 'a set of connected things or parts'. The Sun, the Earth, and other planets and their moons, for example, form the Solar System. But the term is also used in engineering to describe many other interconnected sets of parts such as hi-fi systems, measuring systems, computer systems, control systems, and so on, and it is with the behaviour of this type of technological system that this article is mainly concerned. The general concepts to be discussed here are loading, impedance, and feedback. Damping, another general concept of wide application, is discussed in the article 'Electromechanical similarities' on page 32. You will see that not only are they useful in explaining the behaviour of many different types of physical system, but that the concepts are also closely inter-related. Again, an appreciation of these relationships is often of great help when the behaviour of a new type of system has to be understood and explained.

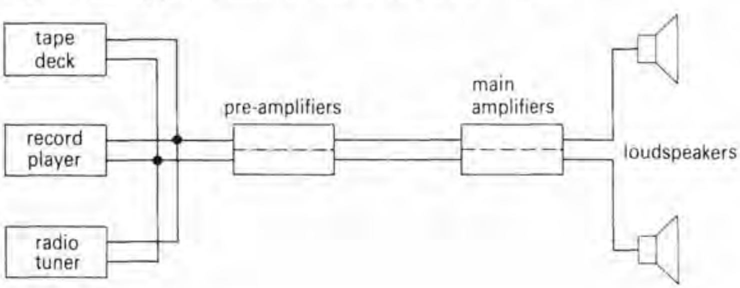


Figure 1
A typical domestic hi-fi system.

Loading

The connection to a source of energy or power of some means for transforming the energy is called *loading*; the energy or power that is transformed is the load imposed upon the source. In the case of a power station, for example, the energy transformed by working and heating in homes, factories, streets, and so on, makes up the load on the station; in a motor car parked with lights on the power consumed by the lights is the load on the car battery. In both these examples the load is usually comparable in magnitude with the capacity of the source but the concept of loading is also applicable in situations in which the load is only a small fraction of what the source could actually supply.

Suppose, for example, that we measure the p.d. across a resistor with a voltmeter; or use a dial-gauge micrometer to measure the deflection of a flexible model; or connect a loudspeaker to an amplifier; or put a thermometer in a beaker of hot water; or draw current from a battery. In each case we are loading the source of energy. Adding a load takes energy from the source and might, therefore, be expected to modify its behaviour.

Figure 2(b) shows the deflection of a flexible model – figure 2(a) – when a force is applied to it. If, as in figure 2(c), a dial-gauge micrometer is used to measure the deflection, then the force exerted on the

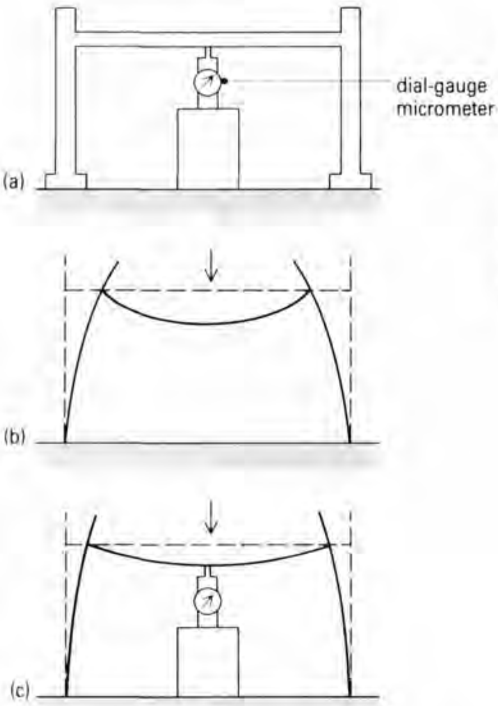


Figure 2
(a) Use of a dial-gauge micrometer to measure the deflection of a model gantry.
(b) Deflection of the model under load.
(c) A dial-gauge micrometer used to measure the deflection of the model exerts a force on the model which reduces the deflection to be measured.

model by the micrometer will tend to reduce the deflection. When the micrometer is added some of the potential energy previously stored in the model as strain energy is now stored as strain energy in the micrometer and as a result the deflection of the model is decreased; only if no force were required to operate the micrometer could a true measure be obtained of the deflection of the model caused by the applied force.

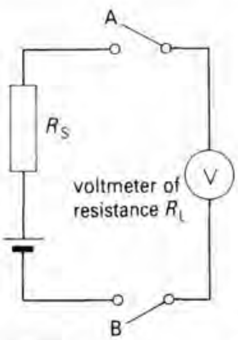


Figure 3
A voltmeter used to measure a p.d.

Measuring systems

When a thermometer is used to measure the temperature of a hot body the energy needed to heat up the thermometer to the temperature of the hot body is drawn from the body whose temperature is being measured. If this is a significant part of the energy available (that is, if the thermal capacity of the thermometer is a significant fraction of that of the body whose temperature is being measured), then the load imposed by the thermometer will reduce the temperature of the hot body. More generally, the load imposed by a measuring instrument will usually change the quantity we are trying to measure.

Thus, when a voltmeter is used to measure a p.d. the current that the voltmeter draws affects the p.d. to be measured. The power transformed in the voltmeter is the load that it imposes on the source, and as we have seen any load affects the source from which it is drawn.

In the system shown in figure 3, if V is the p.d. between A and B with no load then the p.d. across AB when the voltmeter is connected is $\frac{R_L}{R_S + R_L} \times V$. Clearly, the best estimate of V is obtained by making sure that R_L is very large compared with R_S. Under these circumstances the power required to operate the voltmeter is small compared with that available from the source but of course this power must be sufficient to operate the voltmeter. One big advantage of modern digital voltmeters is that because only a very small amount of power is needed to operate them they can have a very high resistance – perhaps $10\,\mathrm{M\Omega}$ as against $1\,\mathrm{k\Omega}$ for a moving-coil meter. They can, therefore, give valid results even when R_S is high.

It is not only measuring systems that impose loads; the connection of any system to any other imposes a load on the source of power. In some cases, for instance with measuring systems, it is desirable to minimize the load imposed upon the system to be investigated. In other cases, for example when connecting a loudspeaker to an amplifier, the objective is to maximize the amount of power that can be drawn from the source. These and other aspects of the behaviour of systems are conveniently described in terms of the input and output *impedances* of the constituent parts.

Input and output impedance

Impedance is a familiar concept in electricity and electronics but is also useful in other fields, including acoustics, mechanics and thermal transfer of energy. When referring to direct electric current, impedance is simply electrical resistance, but when alternating currents are involved – mains electricity, radio and television signals, the signal from a record player pick-up, for example – the equivalent of resistance is electrical impedance. Just as a resistor impedes the flow of direct electric current, so resistors, capacitors, and inductors present impedance to the flow of alternating current. In general, then, the impedance, Z, of an element of an electric circuit is the ratio:

$$Z = \frac{\text{potential difference}}{\text{current}}$$

The impedance of a circuit may sometimes be much greater than the sum of the electrical impedances of its component parts. The power radiated from a broadcasting aerial, for example, represents a load on the transmitter which can be described in terms of a resistance R_R, that resistance which, when multiplied by the square of the aerial current, measures the power radiated. R_R is called the radiation resistance of the aerial and appears in series with any other impedance in the circuit. For many types of aerial R_R is of the order of $80\,\Omega$, and the radiation resistance is the dominant part of the impedance of the aerial circuit.

The input and output impedances of a device are the impedances that it presents at its input and output terminals respectively. In figure 3 (page 9), for example, the input impedance of the voltmeter, R_L, is the resistance that could be measured across its terminals by applying a p.d. and measuring the current. The output impedance of the source is R_S, often referred to as its 'internal resistance', it determines how the terminal p.d. decreases as the current drawn increases.

The output impedance of the source can be determined by measuring the output p.d. with two different loads, R_{L1} and R_{L2}:

$$V_1 = \frac{VR_{L1}}{R_S + R_{L1}}$$

and

$$V_2 = \frac{VR_{L2}}{R_S + R_{L2}}$$

from which R_S can be calculated.

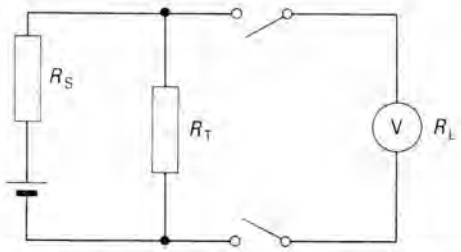


Figure 4
The p.d. measured by the voltmeter depends upon its input impedance and on the output impedance of the source.

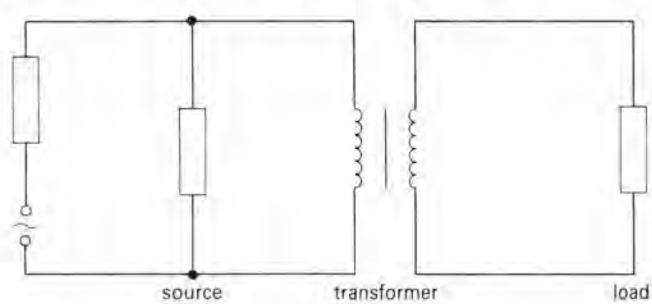
In the more general case shown in figure 4, both R_S and R_T contribute to the way in which the p.d. across the load is affected by the current drain, and it can be shown that the output impedance of the source is $R'_S = R_S R_T/(R_S + R_T)$, the value of a resistance made up of R_S and R_T in parallel. It is the resistance that would be measured across the output terminals if the battery were short circuited. The p.d. across the load is given by:

$$V \times \frac{R_L}{R'_S + R_L}$$

In general the input impedance of a measuring device, Z_{in}, should be much greater than the output impedance, Z_{out}, of the source. If this condition is not met it may be possible to improve the accuracy of measurement by inserting a 'buffer' stage between the source and the measuring device. The buffer has to have a high input impedance so as to minimize loading of the source, and a low output impedance so as not to be too heavily loaded by the measuring device.

A transformer is one form of buffer stage; a step-down transformer transforms the high-voltage, low-current output from the source into a low-voltage, high-current signal to the load (figures 5, 6). An electrometer/d.c. amplifier is another form of buffer stage; it is driven by the source and it in turn drives the load. Its input may be a

Figure 5
A transformer used to transform a high-voltage, low-current signal from a high-impedance source into a low-voltage, high-current signal to a low-impedance load.



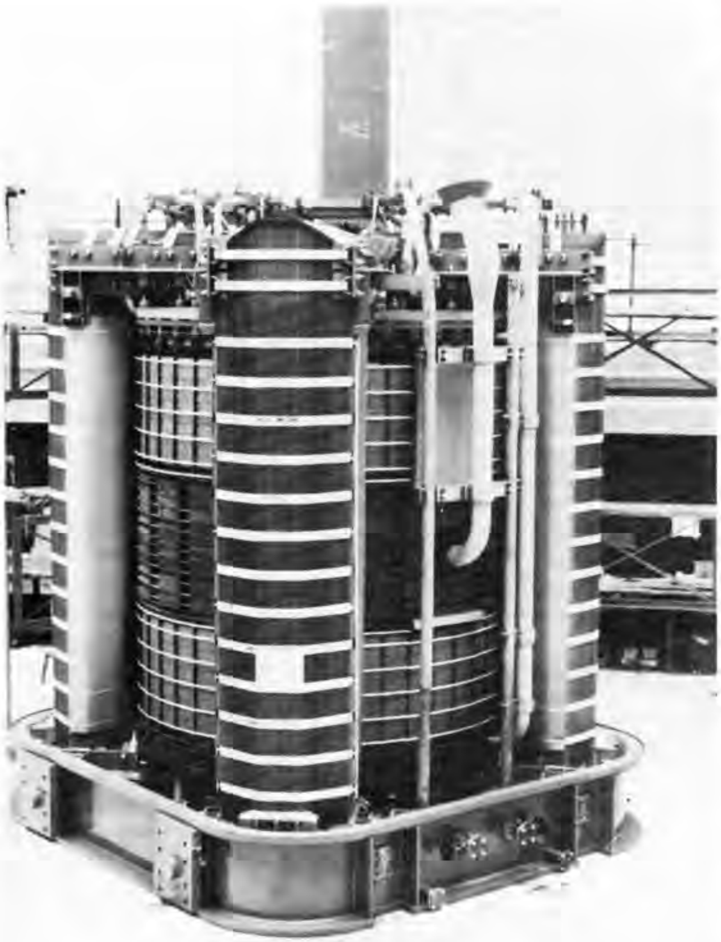


Figure 6
The photograph shows one of the twelve transformer coil and windings assemblies at Torness Power Station.
NEI Peebles Ltd., Peebles Power Transformers.

1-volt signal from a high-impedance source such as a phototube, whilst its output will be a 1-volt signal capable of driving a load of much lower impedance – a moving-coil meter, for example. A transformer, even if ideal, cannot supply to the load more power than it draws from the source but the electrometer, an active buffer, can do just this. The extra power comes, of course, from the electrometer's power supply.

An amplifier has a finite input impedance – the impedance seen looking into its input terminals – and thus imposes a non-zero load on the source used to drive it. An amplifier also has an output impedance which determines the effect on its output of a given load connected to it. In each case the effect is described by the factor $\frac{Z_{in}}{Z_{in} + Z_{out}}$, thus for the effect of the load on the signal to be minimized $Z_{in} \gg Z_{out}$. An amplifier should usually have a low output impedance in order that its performance, in this case its gain, is not significantly influenced by the nature of the load that it is to drive. An important example is an amplifier/controller driving a variable-speed electric motor, the speed of which is required to be proportional to the input signal to the amplifier. This proportionality should be maintained whatever the load on the motor, that is, whatever the amount of work that the motor has to do, and thus whatever the current that the amplifier has to supply. Such a requirement arises, for example, in paper-making machines, steel rolling mills, and electric trains. With a low output impedance the p.d. at the output terminals of the controller (which determines the speed at which the motor runs) is largely independent of the current drawn by the motor (which depends upon the work being done by the motor). The relation between output and input voltages is almost independent of output current. (In practice, negative feedback – see page 15 – is often used to assure the constancy of this relationship by effectively reducing the output impedance of the amplifier.)

Just as a resistor impedes the flow of direct electric current and a capacitor or inductor presents an impedance to the flow of alternating current, a pipe presents an impedance to the flow of water and an insulating layer presents an impedance to the thermal flow of energy – see the section on flow in the article 'Electromechanical similarities', page 32. In general impedance can be thought of as resistance to movement or flow and defined as the ratio:

$$\text{Impedance} = \frac{\text{force variable}}{\text{flow variable}}$$

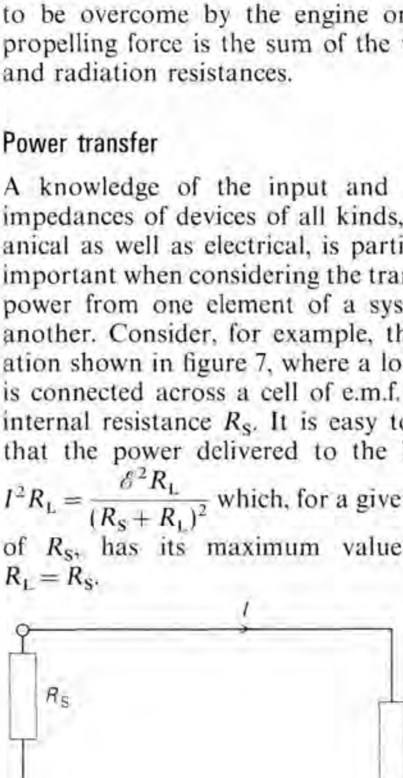
Mechanical systems have input and output impedances and, when they are joined together, behave like the electrical systems described above. Just as it is difficult to transmit a large p.d. to a load of low impedance, so it is difficult to transmit a large force to a soft cushion.

Radiation resistance can also contribute to the impedance of a mechanical system. When a ship travels through water, the viscous resistance of the water has to be overcome by the propelling force. In addition, energy is radiated as wave energy. As in the case of the broadcasting aerial this load can be represented by a radiation resistance which, multiplied by the square of the velocity – the equivalent here of electric current – gives the energy radiated as waves. The total resistive force to be overcome by the engine or other propelling force is the sum of the viscous and radiation resistances.

Power transfer

A knowledge of the input and output impedances of devices of all kinds, mechanical as well as electrical, is particularly important when considering the transfer of power from one element of a system to another. Consider, for example, the situation shown in figure 7, where a load, R_L, is connected across a cell of e.m.f. $\mathscr{E}$ and internal resistance R_S. It is easy to show that the power delivered to the load is $I^2R_L = \frac{\mathscr{E}^2 R_L}{(R_S + R_L)^2}$ which, for a given value of R_S, has its maximum value when $R_L = R_S$.

Figure 7
Power transfer from source to load.



More generally, a source of output impedance Z_{out} transfers maximum power to a load R when $R = Z_{out}$. More generally still, a device of any kind – mechanical or electrical – of output impedance Z_{out} transfers maximum power to a device of input impedance Z_{in} when $Z_{in} = Z_{out}$.

It should perhaps be noted that, when this condition is satisfied, the power transferred to the load increases as Z_{out} decreases. When the impedances Z_{out} and Z_{in} are resistive, the power dissipated in the source is equal to that supplied to the load, so that for every unit of useful energy delivered an equal amount of energy is wasted internally. Thus, even when the output is maximized, the efficiency is only 50 per cent.

The input impedance of the load presented to a generator or other source of power is often adjusted to be equal to the impedance of the generator, perhaps by using a transformer whose input and output impedances are equal to the impedances of the generator and load respectively. In this way the two parts of the system are *matched* for maximum power transfer; they might also be matched for maximum voltage transfer as described earlier, or sometimes deliberately mismatched to minimize power transfer. (In a power station the [resistive] load is matched by the [reactive] impedance of the generator so that the power dissipated in the generator is only a small fraction of that supplied to the load.)

Matching

Matching is an important consideration whenever two or more items have to be joined together – and that means in almost every physical system. In modern life few pieces of equipment are used alone; most are used as part of a larger system and matching is an important aspect of systems engineering. The output from a tape deck or radio tuner has to be matched to the amplifier which it is to drive, just as the amplifier has to be matched to the loudspeakers, or vice versa. (Similar considerations apply between the stages in amplifiers and other electronic equipment.)

Transformers of various kinds are often used to match the different parts of a system. You will be familiar with electrical transformers such as those used to transform a high-voltage, low-current signal from a high-impedance source into a low-voltage, high-current signal for a low-impedance load, or vice versa. Ideally the ratio of input and output voltages is equal to the ratio of the numbers of primary and secondary turns:

$$\frac{V_{in}}{V_{out}} = \frac{V_p}{V_s} = \frac{N_p}{N_s} = N$$

Primary and secondary currents are also related:

$$\frac{I_p}{I_s} = \frac{N_s}{N_p} = \frac{1}{N}$$

A secondary circuit load of resistance R is effectively transformed to a load resistance N^2R in the primary circuit:

$$\frac{V_p}{I_p} = N^2\frac{V_s}{I_s} = N^2R$$

For example, if an a.c. power supply has an output impedance of $100\,\Omega$ the correct matching load is $100\,\Omega$. At a p.d. of 100 V this load would take 1 A and dissipate 100 W – figure 8(a). Suppose now that a 10:1 step-down transformer is connected to the power supply. If this too is to accept 1 A at 100 V the secondary will have to be delivering 10 A at 10 V, and for this a load of $1\,\Omega$ needs to be connected across the secondary winding; this $1\,\Omega$ load will dissipate 100 W as before – figure 8(b). Thus the combination of power supply and transformer is behaving as a source of output impedance $1\,\Omega$. We say that the transformer has matched the $100\,\Omega$ output impedance of the supply to the $1\,\Omega$ load and that in this arrangement the transformer is presenting an output impedance of $1\,\Omega$.

There are mechanical, acoustical, and optical devices which perform functions akin to those of an electrical transformer. Some examples are:

i The gearbox of a car used to transform the high-speed, low-torque output of the engine to a lower-speed, higher-torque drive to the wheels. The ratio of the transformer is the gear ratio – the ratio of input and output speeds.

Figure 9
Express Newspapers P.L.C.

ii The gearbox and propeller of a ship used to transform the high-speed, low-torque output of the engine to a low-speed, high-torque thrust on the water.

iii The horn of a musical instrument used to transform a small-area source into a larger-area source which is better matched to the open air into which the sound is to be radiated. (See figure 9.)

Other examples of transformers are the sounding boards of pianos, the bodies of violins, and the non-reflecting coatings on the lenses of cameras and binoculars. The first two of these enable sound energy to transfer more easily from the source to the air whilst the third allows more light to transfer from the source to the film or the eye by providing a match between the optical impedances (refractive indices) of air and glass.

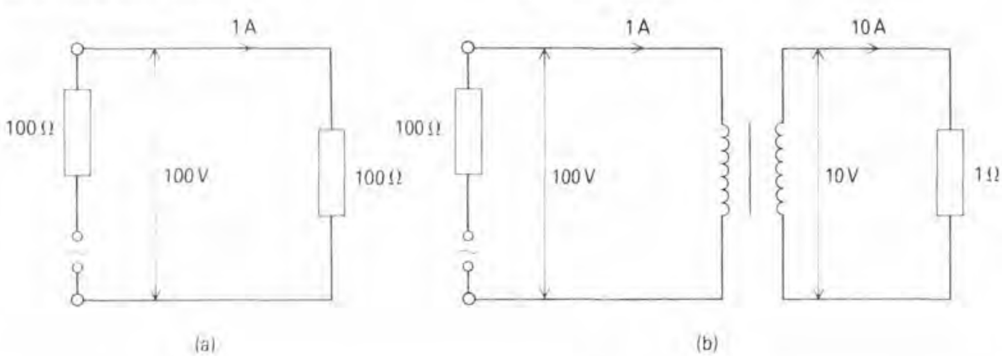


Figure 8
A source of output impedance $100\,\Omega$ supplying 100 W to matched loads:
(a) of $100\,\Omega$;
(b) of $1\,\Omega$.

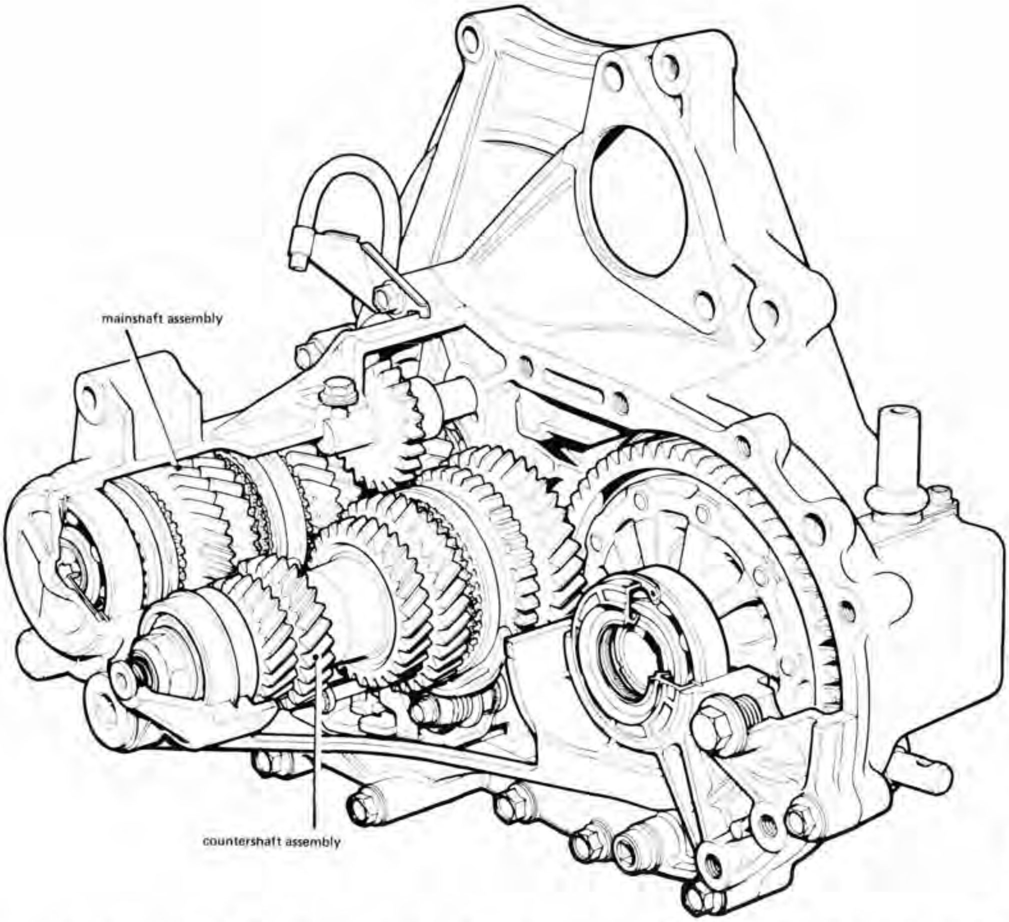


Figure 10
Sectioned view of a manual gearbox.
Austin Rover Group Ltd.

If the surface of a lens is coated with a film of material of different refractive index and of thickness $\lambda/4$, where λ is the wavelength of the light in the material, then it is easily shown that the waves reflected back into air from the inner and outer surfaces of this film are out of phase and superpose destructively, thus reducing the amount of light reflected. Further analysis shows that the intensity of the reflected wave is zero if the refractive index of the material of the film is $\sqrt{n_1 n_2}$ where n_1 and n_2 are the refractive indices of air and of the material of the lens respectively. It is not quite so easy to show that the light which is not reflected is now transmitted, thus increasing the total amount of light transmitted, but this is the effect produced by the film. The $\lambda/4$ length of material of suitable refractive index is, in effect, a transformer matching the optical impedance of the glass lens to that of the air (see figure 11). Quarter-wave transformers which act in a similar way are used in sonar systems and in radar and microwave telecommunication systems.

In practice, aerials and transmitters are often connected by a transmission line, usually a coaxial cable similar to that used to connect a television aerial to the set. The transmitter 'sees' the impedance of the transmission line which, in turn, 'sees' the radiation resistance of the aerial and, for maximum power transfer, both junctions have to be properly matched. The impedance of a coaxial transmission line is given by $Z = \sqrt{\dfrac{L}{C}}$, where L and C are the inductance and capacitance of unit length of the line. (For the coaxial cable usually used in television and hi-fi systems $Z \approx 80\,\Omega$.)

Other examples of matching include

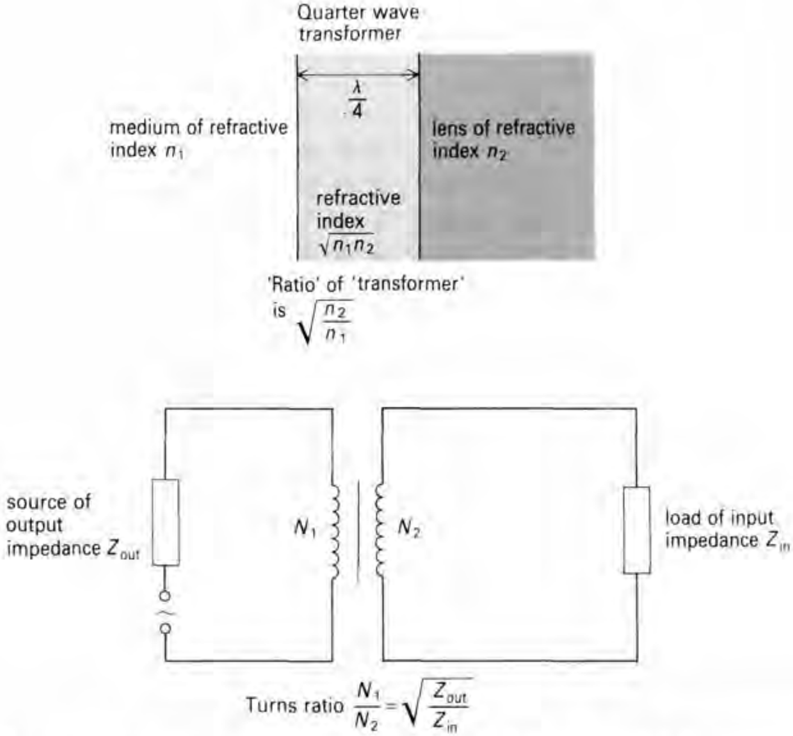


Figure 11
The anti-reflection coatings on the lenses of cameras and binoculars are, in effect, transformers matching the optical impedances of air and glass in order to facilitate the transfer of light energy from one medium to another.

the use of flippers in swimming to enable effective use to be made of the force that can be exerted by the muscles of the swimmer, and in the design of vibratory feeders for the automatic feeding of components to assembly machines or assembly stations in a factory. In the case of mechanical systems the equivalent of resistance or impedance is $\frac{\text{force}}{\text{deflection}}$, or stiffness. For maximum power transfer to the vibrating table the stiffness of the vibrator has to be matched to that of the table itself.

Sound energy is transmitted most efficiently from one medium to another when the media have approximately the same impedance. An example of this is in the use of ultrasonics for non-destructive testing of metals. The ultrasonic transducer, a piezoelectric crystal, is usually 'coupled' to the metal object to be tested via a layer of glycerine, a material whose acoustic impedance is intermediate between that of the transducer and that of the solid to be tested. The glycerine replaces air which would present a very low impedance and thus greatly reduce the acoustic energy transmitted into the solid.

On the other hand, in some situations deliberate mis-matching is often practised so as to reduce power transfer between different parts of a system. In the interest of sound insulation, for example, media of very different acoustic impedances are often introduced into the path of the sound so as to reduce the sound energy transmitted. The acoustic impedance of a dense partition is very much greater than that of air so that very little of the sound energy incident in air on such a partition is transmitted, most of it being reflected back. The effect, which is analogous to the reflection of light at the interface between glass and air, can be enhanced by using multi-layer partitions with layers of different acoustic impedance.

The effect of negative feedback on output impedance

We have seen that the effect of a low output impedance is to ensure that the output is largely independent of the nature of the load; if, in figure 3 (page 9), R_S is small then $\frac{R_L}{R_S + R_L}$ is largely independent

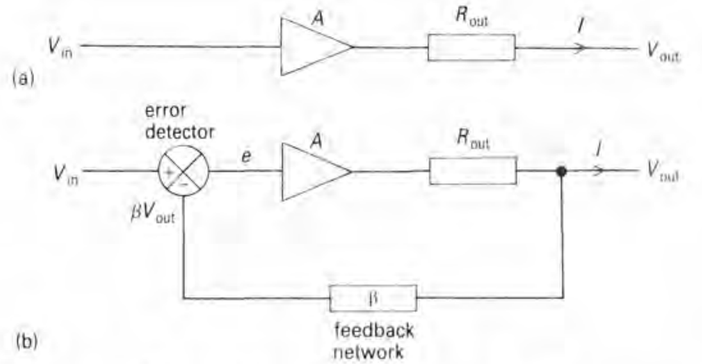


Figure 12
(a) Amplifier without feedback.
(b) Amplifier with voltage feedback.

of the value of R_L and the output voltage follows closely any variation in the e.m.f. of the source.

The effect of negative (voltage) feedback in an amplifier is to ensure that the output signal follows closely any variation in the input signal, or, in other words, to reduce the output impedance of the amplifier. With reference to figure 12(a), without feedback the output voltage:

$$V_{out} = AV_{in} - IR_{out} \qquad [1]$$

where A is the gain of the amplifier. The output impedance of the amplifier, Z_{out}, equals R_{out}. With negative voltage feedback – figure 12(b):

$$e = V_{in} - \beta V_{out}$$

and

$$V_{out} = Ae - IR_{out}$$

where β is the fraction of the output voltage that is fed back, as negative feedback, to the input. Thus:

$$V_{out} = \frac{AV_{in}}{1+\beta A} - \frac{IR_{out}}{1+\beta A} \qquad [2]$$

Comparing equations [1] and [2] it can be seen that with negative feedback the gain and output impedance of the amplifier have both been reduced by the factor $1+\beta A$, which can be large if A is large.

Without negative feedback the gain of an amplifier can, and usually will, vary with the frequency of the input signal or with its amplitude. Both effects would be undesirable in a hi-fi system: some frequencies would be reproduced louder than others, or soft sounds might be amplified more than loud sounds. By using negative feedback (and ensuring a low output impedance) a flat frequency response can be obtained and all sounds, soft and loud, amplified by the same amount.

The main effects of negative feedback are to reduce the gain and to increase the range of frequencies over which the gain of an amplifier remains constant – see figure 13. At very low and very high frequencies, however, capacitances and inductances associated with the wiring of the circuit may cause phase changes in the feedback loop which tend to make the feedback positive instead of negative; if too much feedback is used, therefore, the effect may be to cause the gain to increase at low and high frequencies. (With integrated circuits or other direct-coupled amplifiers the open-loop gain does not fall off as the frequency decreases but tends to remain constant down to zero frequency.)

Negative feedback is used in control systems to constrain the output signal to follow closely any variation in the input signal, for example, to keep the temperature of a room equal to the value set at the thermostat; or to make the speed of a servo-motor proportional to the input signal to the driving amplifier. The word servo comes from the Latin *servus*, a slave, and is used to describe a motor or system which is meant to follow exactly the instructions given by a low-power input signal. In effect the negative feedback reduces the output impedance of the system thus making it less sensitive to the nature of the load on the system.

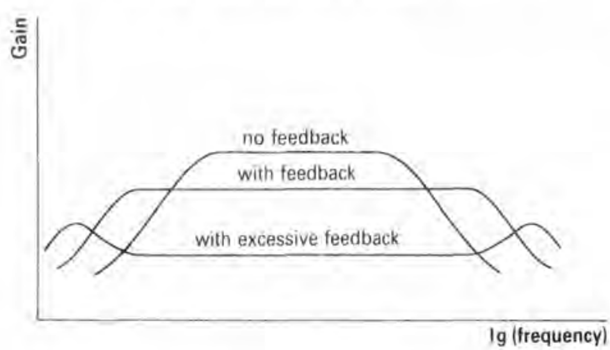


Figure 13
The effects of negative feedback on frequency response

Summary

The operation of most physical systems, including all the technological systems mentioned in this article, and many others, is characterized by the transfer of energy from one part of the system to another. In a camera light energy is transferred from the object to the photographic film; in a broadcasting system sound energy is converted into electrical energy which, after amplification, is converted into electromagnetic radiation; in a car the energy released when fuel burns in the engine is transferred, via the gearbox, to the vehicle and its occupants. The criteria for their performance vary from system to system: in some instances, as in the case of a generator supplying power to the National Grid, it is maximum power transfer; whilst in the case of a voltmeter measuring a p.d. the criterion is maximum signal (in this case voltage) transfer. In all cases, however, the performance of the system is determined by the impedances of the elements concerned. In the general case of transfer of energy through an element of a system it is its input and output impedances that are involved. Although you will be most familiar with impedance in electrical terms the concept is applicable to other types of system also, including acoustic, mechanical, and optical, and the same general relations apply to all types of system.

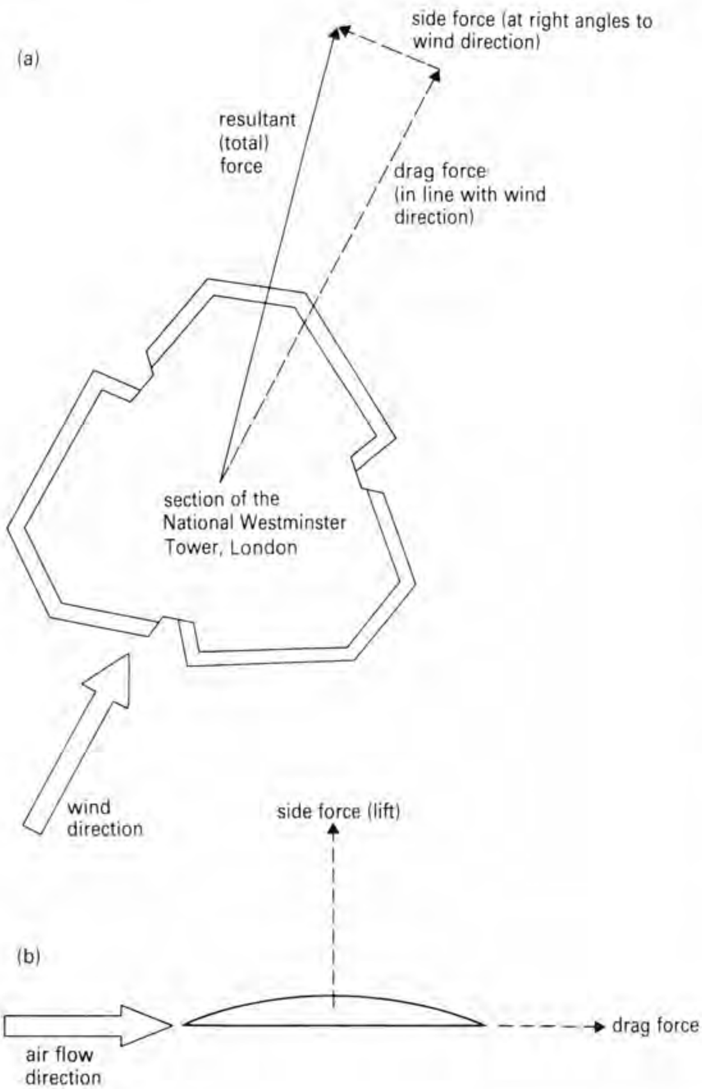
Where possible the elements of systems are designed to have the appropriate input and output impedances but, when this is not possible, various means can be adopted to modify their impedances. The most familiar of these are transformers and negative feedback. Again most familiar in the electrical context, these methods are also used in optical and mechanical systems.

General concepts like impedance are useful in thinking about new types of system as they enable ideas already learned to be applied to unfamiliar problems; they also illustrate the essential unity of the physical world.

Buildings, bridges, and wind

L. R. WOOTTON
Atkins Research and Development
D. E. J. WALSHE
British Maritime Technology, Feltham

Figure 1
(a) Side force created by wind on a non-symmetrical structure.
(b) Drag and side force on section of an aircraft wing.



Forces caused by wind

The purpose of civil engineering is to produce safe, useful, elegant structures which can be built and maintained economically. The effect of wind is a nuisance because it produces forces that must be resisted by the structure, and may cause unwanted movement which the designer must guard against. This article is about some of the problems and how they are overcome, generally without the people who use the final building, bridge, or tower even noticing.

When wind blows on any structure it produces a pressure difference between the windward and leeward sides. This pressure difference results in a force, called the drag force, which acts in the direction of the wind. If the structure is not symmetrical or if the wind blows at an angle to a side of a symmetrical structure, then it will also experience a force in the cross-wind direction, referred to as the side force – figure 1(a). For some shapes the side force is negligible, but for others it can be very large, and indeed an aeroplane wing is particularly designed so that there is a large side force (or 'lift' as it is called in this case) which may be over ten times the magnitude of the drag force – figure 1(b). In the case of civil engineering, we are primarily concerned with drag forces because they are unavoidable and all buildings must be designed to withstand them. Only in certain cases, such as grandstand roofs or specially shaped buildings, are the side forces important.

There are two features about wind loads that are obvious from our own experiences while walking on windy days. The first is that the higher the wind speed the greater the force. In fact, the drag and side forces on structures increase in proportion to the square of the wind speed. The second feature is that the natural wind is not smooth but, particularly in high winds, blustery: by this we mean that it varies continuously in speed and direction, to a certain extent in a random way. We can see this, for example, in the way the leaves and branches of trees move around in the wind, or by watching smoke from a chimney on a windy day.

Study of the wind around a body is made more interesting and understandable if we can make the flow visible by smoke, or by replacing the air flow with water and injecting coloured dye streamers into the flow. The flow patterns in air and water can be similar in form even though the fluids have quite different properties. It is often more convenient to use water to give a clear picture of what is happening (figure 2).

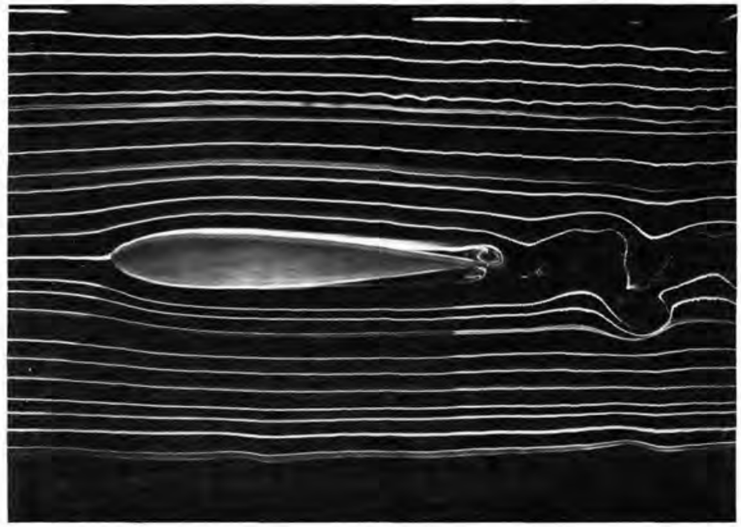


Figure 2(a)
Flow over an aerofoil.
Dr P. Bradshaw, Department of Aeronautics, Imperial College, University of London.

If the body is smooth, like a section of a wing, an aeroplane, a fast sports car, or the sail of a yacht, then wind flow passes smoothly over it and the drag is small. This is called streamlining. However, the design of buildings cannot be determined purely by aerodynamic considerations and, in any case, in the examples of the aeroplane, car,

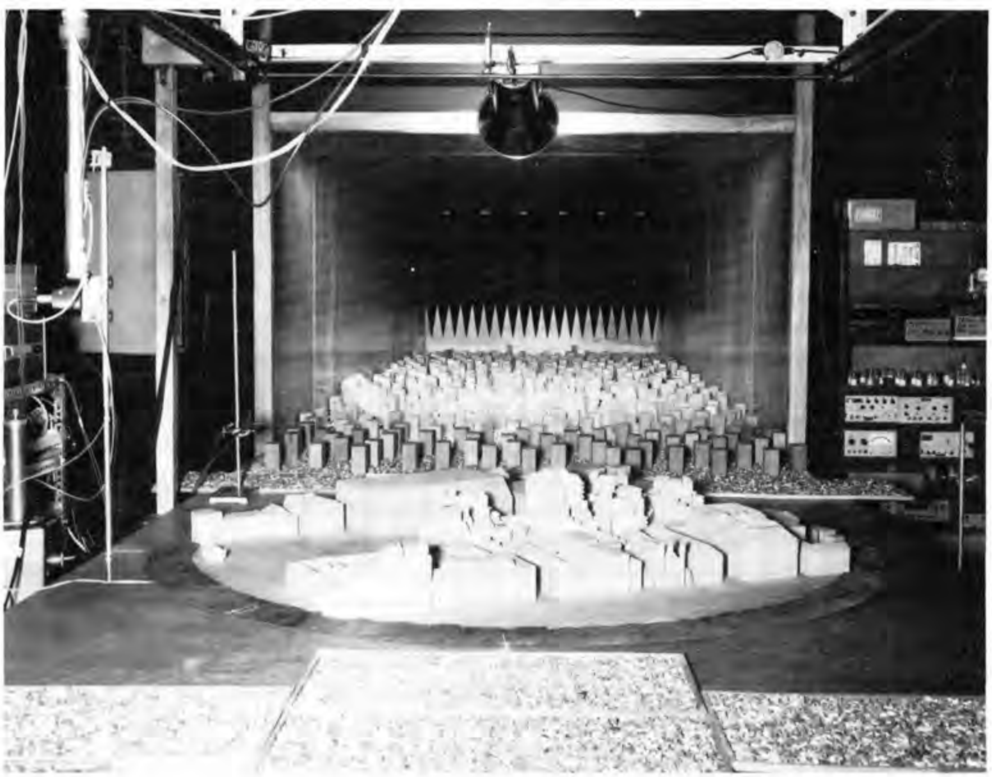


Figure 2(b)
Wind tunnel showing urban terrain simulation (blocks and spire fence). The model under test is placed on a rotating turntable.
Building Research Establishment photograph: Crown copyright.

and yacht the body moves through the fluid with a determined direction. A building, on the other hand, must face up to wind from all directions and we cannot design it to be universally streamlined. A square, round, or typical structural shape is called a *bluff* body because unlike the streamlined body the flow downwind is much disturbed. This disturbed flow is called a *wake*. A characteristic of the wake is that it produces low pressures on the leeward facing parts of the body and these integrate to give a high drag force (figure 3). In general terms, anything that can be done to reduce the width of the wake also reduces the drag force. Thus, in addition to depending on wind speed, as we have noted before, the drag force is affected by the shape and size of the body. In practice, the outside shape and size of a building are set by the specification for its use and by the architectural requirements. Unlike the design of an aeroplane, where the designer considers the flow around the external shape and its forces to be of great importance, the civil engineer seldom wishes to alter a building's shape to satisfy wind criteria alone. Instead, the internal structure of the building must be made strong enough to withstand the wind forces.

The wind usually produces forces that fluctuate in an irregular way about a mean (figure 4). So far, we have only been interested in the mean (or time-average) forces on a structure. However, the fluctuating component of the force may be important and can be irregular, periodic, or a combination of the two. This represents a particularly interesting study.

Vortex shedding

The clearest example of this is the shedding of vortices from a circular cylinder or any bluff-shaped body. This flow pattern occurs even if the approaching wind flow is smooth (figure 5). As the wind flows past the block, the flow on one side starts to curl up round the rear corner, driven on the outside by the fast moving free stream wind and retarded by the nearly stagnant air in the wake behind the block. As time goes on so this rotating whirl of air, called a vortex, gets bigger and more powerful until, at a certain point, it breaks free from the block and moves downwind. At the same time as the breakaway another vortex starts to form on the other side and a pattern of alternate clockwise and anticlockwise vortices develops in a stream downwind of the block. Precisely the same pattern is formed by a circular-section cylinder in air or water, as can be observed, for example, on the pier of a bridge in a flowing stream.

Nowadays we can also simulate the flow patterns using a mathematical model of the flow and a computer (figure 6). These calculate the development of vortices in

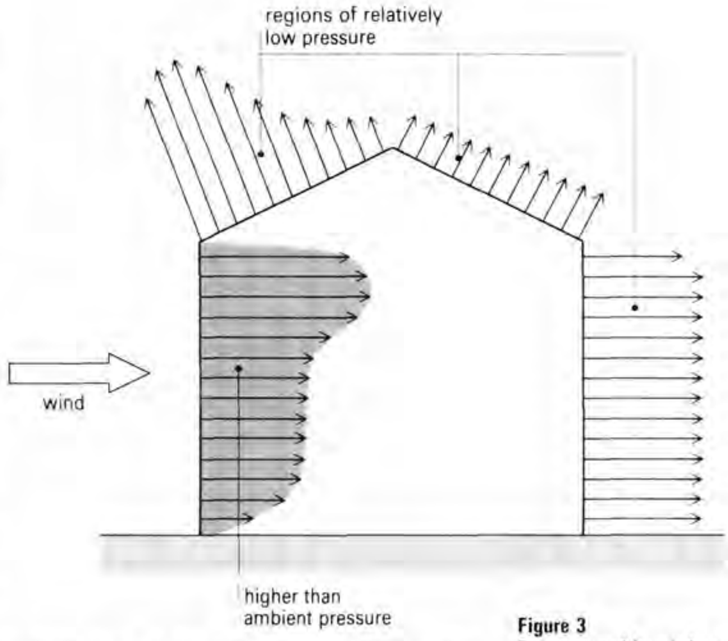


Figure 3
Forces caused by wind on the cross-section of a house.

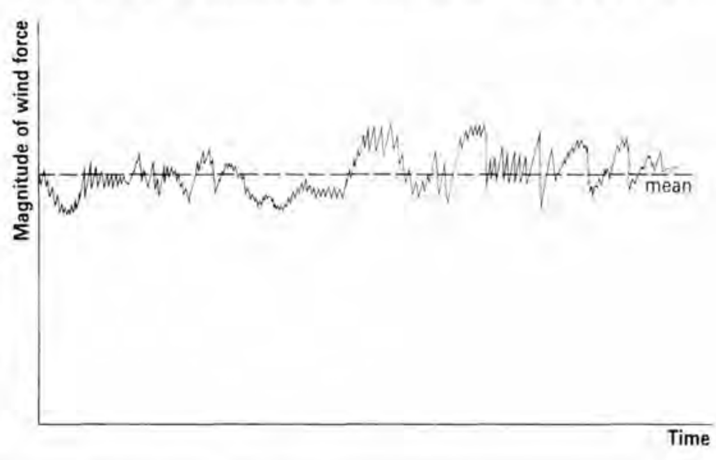


Figure 4
Typical time variation of an irregular wind force.

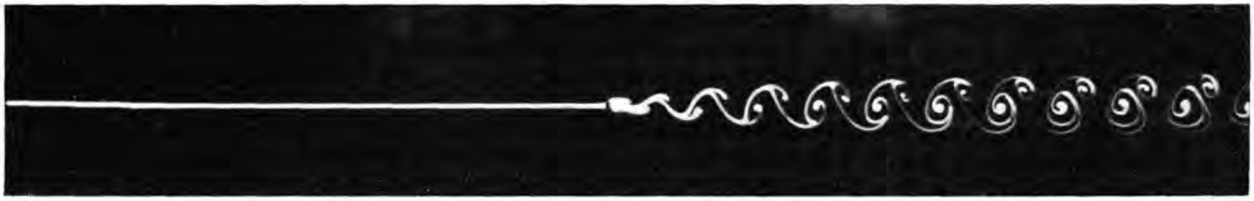
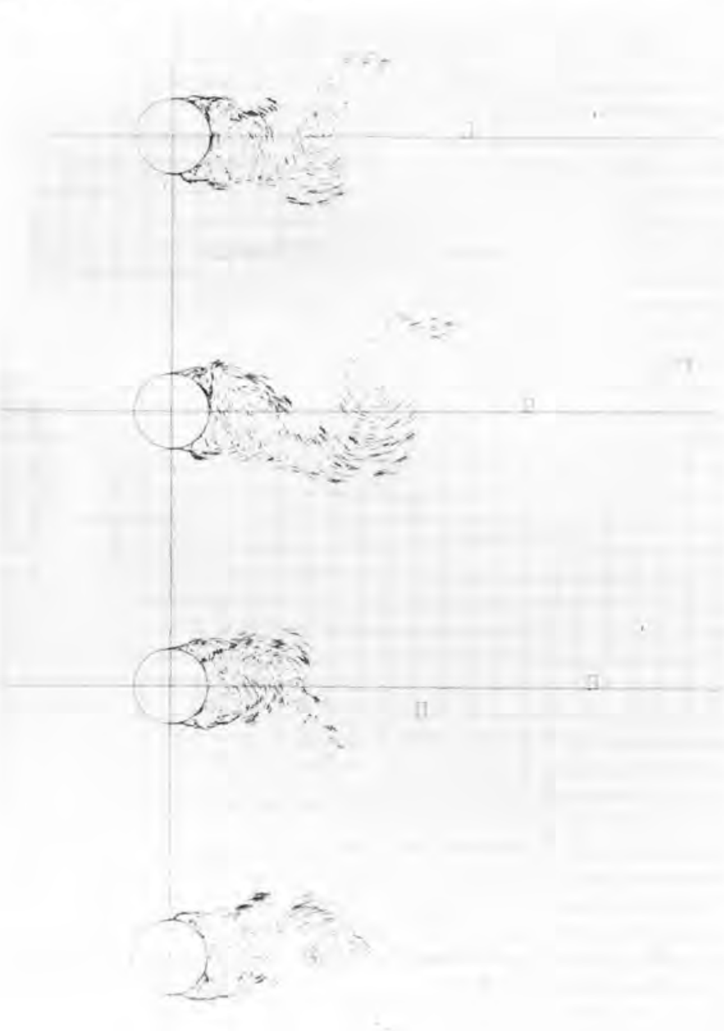


Figure 5
The formation of vortices.
British Maritime Technology Ltd.

Figure 6
Vortex shedding being simulated by a mathematical computer model.
Dr P. K. Stansby, University of Manchester.

the flow and their passage as they move downstream. This sort of calculation requires a very good understanding of the properties of flow by the mathematician and fluid dynamicist who set up the computer program. However, if well analysed, computer simulations are very useful as an addition to visual observations in real fluids.



It is interesting to note that similar patterns of vortices form in very large scale flows. Figure 7 (page 22) shows a satellite photograph of the wind flow pattern around a solitary mountain on an island where the clouds produce a visualization of the air streams.

There are two basic relationships that have been observed with respect to vortex shedding. The first is that the vortices are shed more frequently as the wind speed increases. Indeed, to a close approximation, there is a proportional relationship:

$$f \propto v$$

where f = frequency of shedding of pairs of vortices, in Hz; v = wind speed upstream of cylinder, in m s^{-1}.

The second feature is that the greater the diameter of the cylinder or the size of the block, the slower the rate of shedding. Thus, for example, the shedding of vortices from a telephone cable is very rapid; from a mountain it is very slow. Again, to a good approximation, there is a proportional relationship, but this time it is inverse, which means that:

$$f \propto \frac{1}{d}$$

where d is the diameter or size of the bluff object, in m. Combining these two relationships gives:

$$f \propto \frac{v}{d}$$

which implies that:

$$f = S\frac{v}{d}$$

where S is the Strouhal number (named after the German scientist who first studied the problem).

The interesting feature is that for any particular structure in a wind of constant

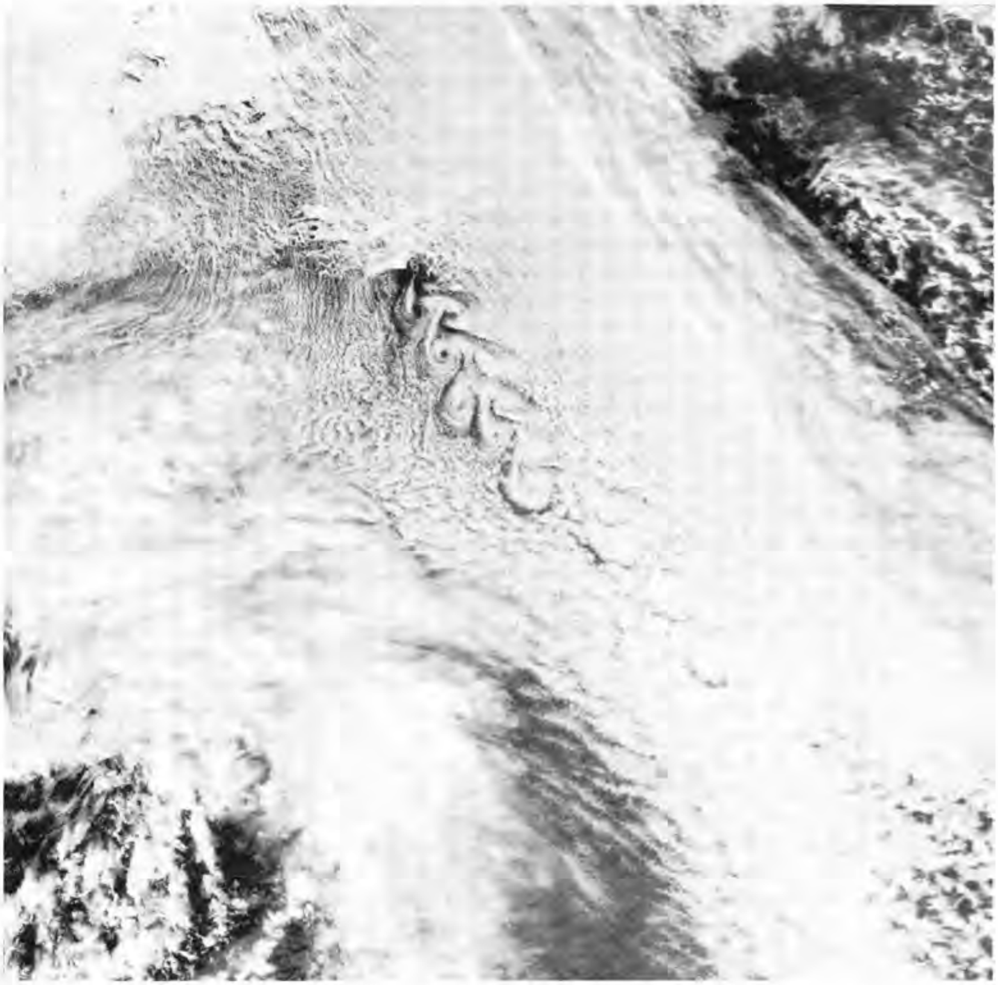


Figure 7
Wind flow over Jan Mayen Island, near Greenland.
Satellite photograph, Department of Electrical Engineering and Electronics, University of Dundee.

speed, the shedding has a very steady frequency. The value of the constant, S, has been found experimentally to be about 0.2. This would all be of purely academic interest if it were not for the fact that when each vortex is shed, it produces a sideways force on the structure. As a vortex is shed from one side of a structure it tries to suck the structure in that direction. As the vortex moves downstream with the wind, away from the structure, its effect rapidly diminishes. In time another vortex is formed on the other side and the process is repeated. The result is a force that is mainly across the direction of the flow. Since the vortices are shed from each side alternately, this force changes direction each time. Because the shedding is regular, the force is also strongly rhythmic. Now consider the example of a flexible structure such as a very tall steel chimney stack which has a clearly defined natural frequency of motion. The motion of the stack will be excited in resonance when the frequency of the vortex-shedding force coincides with the natural frequency, that is, when $f = f_{nat}$ where f_{nat} is the natural structural frequency.

Thus, the resonant motion occurs when:

$$f = 0.2\frac{v_{crit}}{d} = f_{nat}$$

or

$$v = v_{crit} = 5f_{nat}d$$

where v_{crit} is the critical wind speed for response.

This needs a little more explanation. As the wind speed increases, so the shedding frequency increases. The response of the stack is quite insignificant because the shedding frequency is below the natural frequency of the structure. At a particular wind speed, corresponding approximately to the value $5f_{nat}d$, the motion may build up because of resonance between the exciting force and the stack (figure 8). Note that it only *may* build up: there can be other reasons why a structure is prevented from large amplitude motion and these will be discussed later. But if a resonant response does occur then it is at the least a cause for concern; at worst it could cause failure or collapse. Because the cause of the motion is the sideways force of the vortices, the resulting response is across the direction of flow. Also note that, if the response does occur, the motion of the structure tends to encourage the vortex shedding at the frequency of the motion: in other words the vortices may continue to be shed at f_{nat} even though the wind speed does not remain exactly at v_{crit}.

At higher wind speeds (provided that the structure survives the resonant response!), the amplitudes of motion would decrease since the frequency of shedding would be well above the natural frequency of the structure (figure 8).

Combatting resonance in structures

However, it is obvious that many structures do withstand these forces without oscillating; why? There are several possible reasons and by understanding the reasons it becomes possible to apply remedies. One of these reasons is that the natural frequency may be so high that the resonance vibration never occurs because the critical wind speed is too high to occur in nature. The high natural frequency is a result of high stiffness relative to mass. (Recall that $f_{nat} \propto \sqrt{\frac{k}{m}}$, where k is stiffness; so for f_{nat} to be large, k must be large, particularly if m is also large.) Steel chimney stacks can have their stiffness increased by the addition of guy ropes, but a tall building must have the stiffness incorporated into the basic structure.

The second reason why motion is prevented is as a result of the ability of the structure to absorb energy. This can be appreciated through the following example. On a day when the wind is still it may be possible to force the structure into motion. This can be done by firing rockets off from one side, or by pulling it with a winch and cable and then releasing it (figure 9), or even by someone swinging on the structure so as to get it to resonate. In all these methods, energy has been given to the structure to make it oscillate. Once the energy is no longer provided, the vibratory motion dies down. The rate at which the motion decays is a measure of the structural damping: that is, a measure of the ability of the structure to absorb energy. A rapid decay indicates a high damping, and a structure that has high damping will not

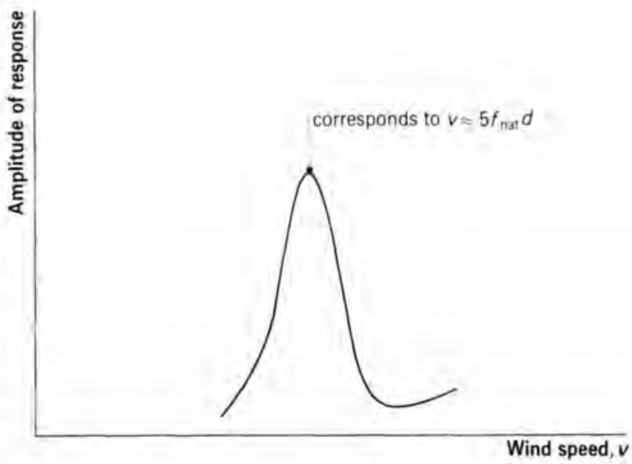


Figure 8
Amplitude of response plotted against wind speed.

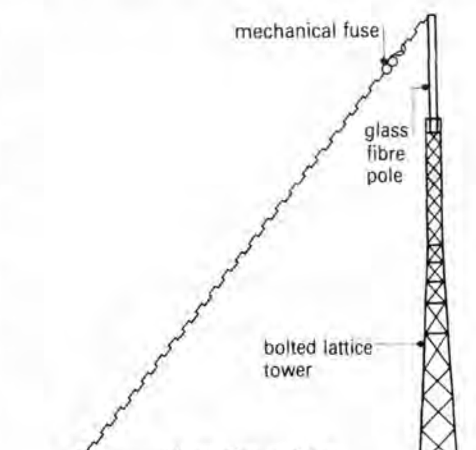


Figure 9
Method of giving an instantaneously released load to a u.h.f. television mast.

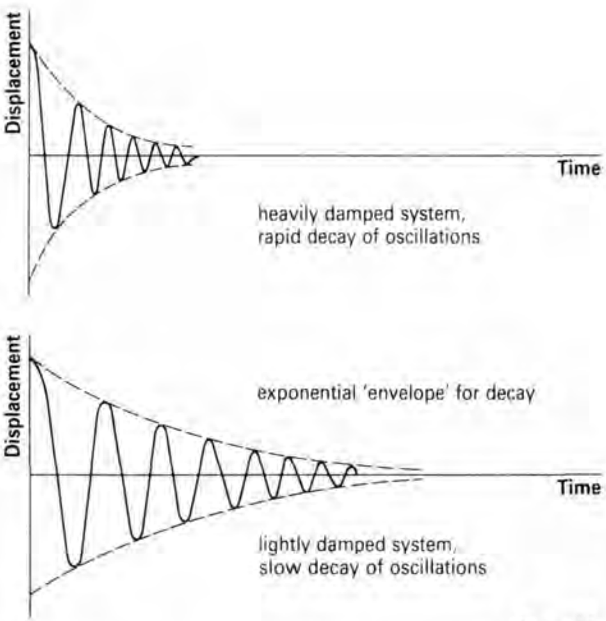


Figure 10
Decay of oscillations.

respond as much as a lightly damped structure to vortex forces (figure 10).

When a structure is vibrating continuously in the wind there must be a balance between the energy being fed into the structure by the air (there is nowhere else for the energy to come from) and the energy dissipated by the structure. This dissipation is related to the damping and the amplitude of the motion. If the structure has a high damping then it can dissipate a lot of energy even if the amplitude of the vibration is very small. If the damping is low then a given rate of energy input from the wind will produce a greater amplitude. In both cases the energy balance is maintained.

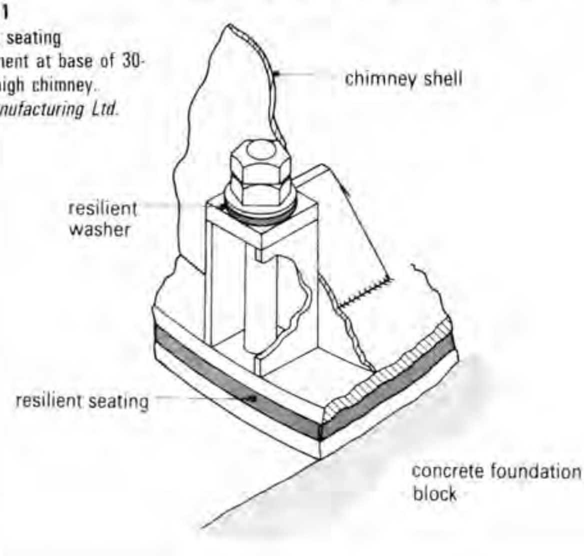
All structures have natural damping in their materials with additional damping caused by minute cracks in the concrete, the brickwork, or the welds of the steel, or by the movement of internal walls and floors in the case of buildings. Usually this is sufficient to prevent motion, but if it is not, then it may be practical to fit additional damping and this can be done in a number of ways. For chimneys and certain other structures, resilient pads of a rubber-like material can be inserted just above the foundations and these absorb considerable energy as soon as the structure starts to sway (figure 11).

A remedy used to prevent vortex-induced motion of rockets when standing vertical on the launch pad is to fit chains hanging in vertical square box-section tubes (figure 12). The chains are designed to swing if any movement starts in the wind and dissipate energy by impacting on the sides of the box.

Several other forms of damper have been used, but by far and away the most common is the Stockbridge dumb-bell damper used to eliminate vortex-induced motion of electrical overhead lines, suspension bridge cables, and so on. These dampers consist of two steel masses connected to the central fixing on the line by a stiff stranded cable. Careful choice of the mass and stiffness of the system ensures that energy is absorbed by out-of-phase movement of the masses with the electrical line (or cable) and dissipated by fretting of the strands of the cable of the dampers. These dumb-bell dampers can be seen on most high-voltage power lines in the UK, and also on the Humber Bridge (figure 13).

The third reason for the lack of response of a structure to wind is because its mass is too great. An example here is that of the stone Monument in the City of London. It is about 62 metres high but will not move in the wind, whereas a lightweight steel chimney or radio transmission mast of similar external shape might. This illustrates the main reason why wind-induced motion has been of increasing

Figure 11
Resilient seating arrangement at base of 30-metres high chimney.
Tico Manufacturing Ltd.



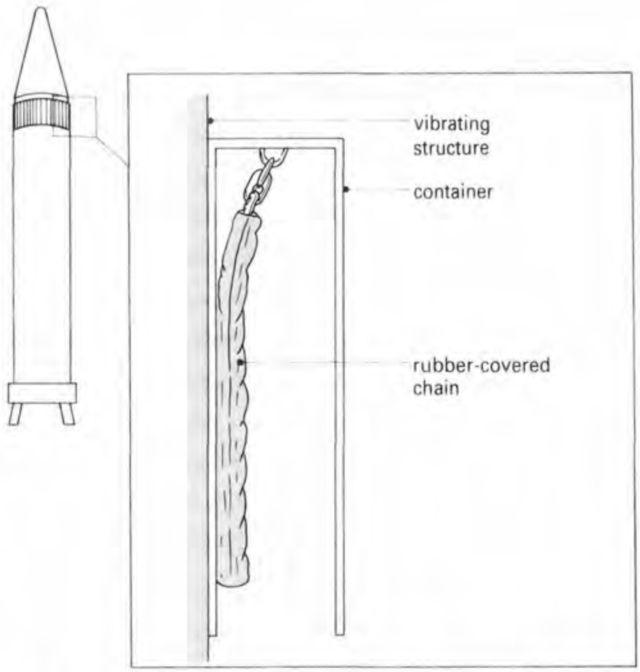


Figure 12
Chain damper installation on Jupiter launch vehicle.

interest. Over the last few years the trend to greater economy has resulted in much lighter structures. As a consequence, there is greater potential for a troublesome response.

An obvious way of eliminating vortex-induced motion in wind is to change the shape of the structure. In the case of circular steel chimney stacks, the most usual remedy is to fit helical 'strakes' which break up the strong regularity of the vortex shedding and so prevent the harmful resonant response (figure 14). An important aspect of strakes is that they act equally well in winds from any direction.

Strakes were developed as a result of tests on a model in a wind tunnel some years ago. Wind tunnels are the most important test facility for this sort of work, since although taking measurements on real buildings and structures is very valuable, it is also very expensive, and can be frustrating if the wind does not blow when you want it to. The characteristics of the natural wind can be represented in a wind tunnel at a small scale, such as 1:50 or 1:200. The model airflow can be made to have the three distinct characteristics of the real wind. The first is that the speed can be made to vary from calm to the equivalent of a storm force wind. The second

Figure 13 (left)
Stockbridge dampers on the Humber Bridge.
Colin Price.

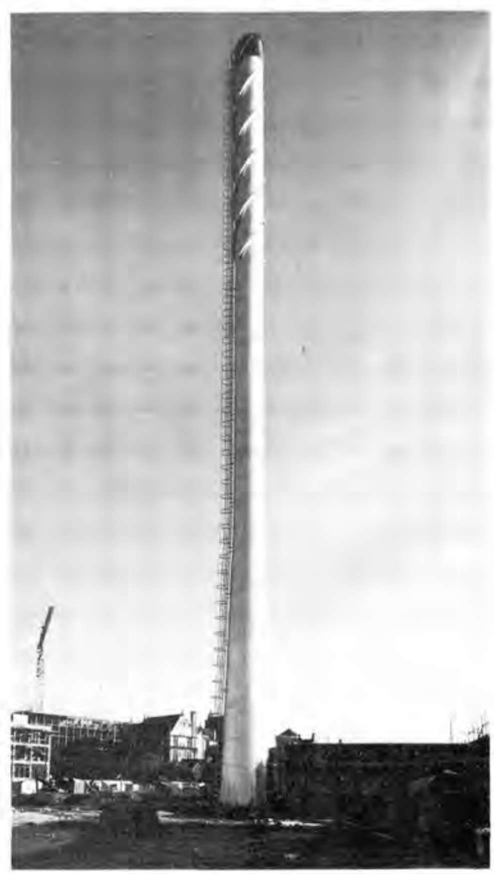


Figure 14 (right)
Strakes fitted to the top of a chimney stack.
National Research Development Corporation.

reproducible characteristic is the gradual increase of wind speed with height above the ground that occurs naturally. This rate of increase depends on whether the wind has, for example, blown over a large distance of flat, open countryside (in which case the flow is fast quite close to the ground), or whether it has been blowing over a large city (in which case the flow speed is much lower close to the ground than higher up) (figure 15). The final characteristic is that it is turbulent; that is to say, the wind speed varies continuously about a mean in a fairly random manner. The level of turbulence would be a lot lower for wind flowing over flat countryside than over urban areas.

The turbulence in the air can result in direct forcing of structures irrespective of any motion that may or may not occur due to vortex shedding. The characteristic of turbulence is that it is not strongly rhythmic so the response is not of a resonant form but, rather, builds up gradually with increasing wind speed. A very important distinction between turbulent forcing of a structure and vortex-induced motion is that the latter produces motion across the direction of the wind whilst the former forces the structure backwards and forwards in line with the wind.

The response due to direct forcing by turbulence in the approaching wind can be differentiated from responses arising from the flow around the body itself. Movement induced by flow around the body is referred to as self-excitation. Such motion can be controlled to negligible amplitudes by a certain level of structural damping; but once this is reduced, the amplitude may build up very rapidly. These forces can easily lead to instability in a structure.

Figure 16
Wind-excited asymmetric torsional oscillations of the first Tacoma Narrows Bridge.
Bulletin 116, University of Washington Engineering Experiment Station.

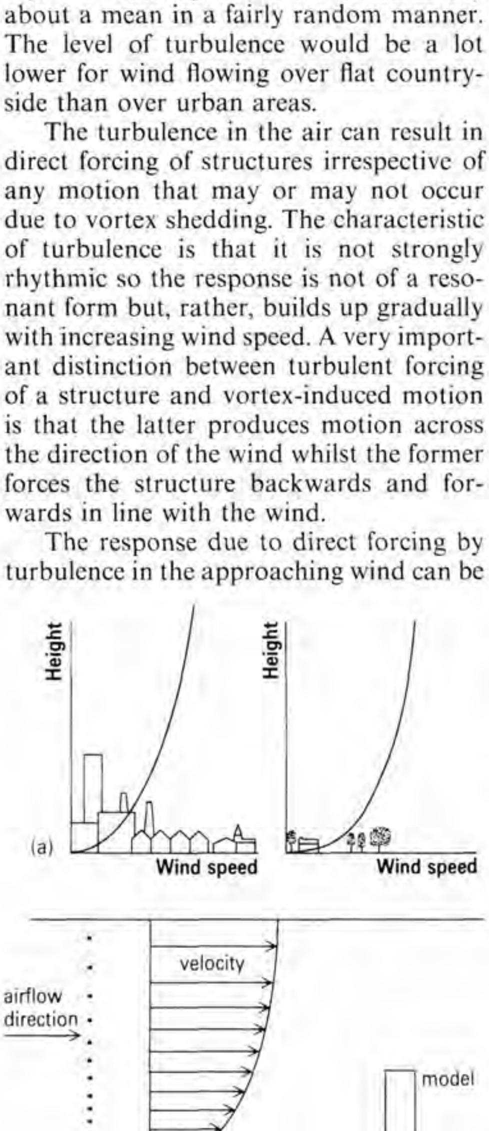


Figure 15
(a) Typical wind speed profiles over different terrains.
(b) How a velocity profile can be simulated in a wind tunnel by placing unevenly spaced bars across the working section.

There have been many instances of self-excited oscillations, but perhaps the most famous are the oscillations of the first Tacoma Narrows Bridge. This bridge was built in 1939 near Seattle in the USA. It was then, with a main span of 853 metres, one of the largest suspension bridges in the World, and was stiffened by two deep solid girders. For many months after its completion the bridge had oscillated at low wind speeds, with its road deck moving vertically up and down. This motion was not destructive and, in fact, was a source of amusement to motorists, who christened the bridge 'Galloping Gerty'. In 1940 it oscillated to destruction in a wind of about $19\,m\,s^{-1}$. The destructive oscillation induced by the wind travelling at $19\,m\,s^{-1}$ was caused not by a vertical displacement, but by a twisting motion, as shown in figure 16, which overstressed and eventually fractured the structural members and led to the complete collapse of the bridge.

It is interesting to note that this sort of thing had happened before. In 1836, the Chain Pier at Brighton partially collapsed

Figure 17
(a) The Chain Pier, Brighton, by W. Westall ARA.
(b) Sea view of the Chain Pier, Brighton, at the instant the platform separated on the memorable 29 November 1836, by W. Gleadah.

in wind after exhibiting a twisting motion identical to that seen at Tacoma. (See figure 17.) At the time the instability was not understood, and no lessons were learned which could be applied to future suspension bridges.

Tacoma, however, marked a turning point in the history of civil engineering. This time, aeronautical engineers recognized the problem as being similar to the 'flutter' oscillations of aircraft parts (see below), and they knew how to tackle it. Above all, they recognized that as knowledge of the behaviour of materials increased and as design and construction techniques improved, bridges would have less mass and lower damping than previously. They would, therefore, be more susceptible to wind-excited oscillations than before. Thus the phenomenon had to be understood and designed against before any more long-span bridges could be built. A thorough investigation into the mechanism of wind-induced oscillations began in the USA immediately after the Tacoma Narrows Bridge disaster; and in 1946, the National Physical Laboratory at Teddington in England carried out an investigation into the aerodynamic design of bridges. The primary object of the work at the National Physical Laboratory was to assist the consulting engineers responsible for the designs of the suspension bridges to span the Firth of Forth and the River Severn. The designers of these bridges were Messrs Mott, Hay, and Anderson, and Freeman Fox & Partners.

Problems of bridge design

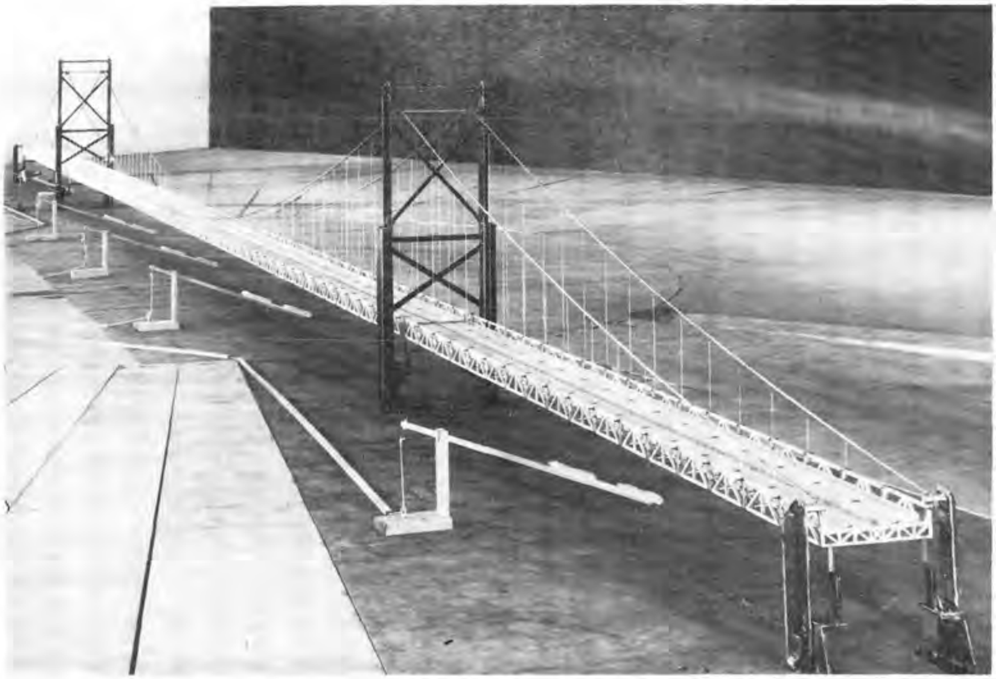
The research work started on a 'full' aeroelastic model. This is a model of the entire structure, with its geometry, mass distribution, and elastic properties scaled so that it reproduces the modes of oscillation and the dynamic behaviour in wind of the full-scale bridge. This particular model (figure 18) was scaled 1:100. It was 15.6 metres long and was tested in a specially constructed wind tunnel.

These tests, and experience elsewhere, showed that bridges could oscillate in any of three types of motion. Figures 19 to 21 show these types of oscillation in diagrams of the bridge. Vertical motion is shown in figure 19. Alternatively, a torsional (twisting) oscillation can take place (figure 20); and thirdly, the torsional and vertical oscillations may couple to give the type shown in figure 21.

These coupled vertical and torsional oscillations can occur in a number of modes of oscillation of the bridge as a whole. Figure 21 shows the fundamental mode; another possible mode is shown in figure 22.

These coupled motion oscillations are a particularly severe form of instability known as 'classical flutter'. Probably everyone reading this article has caused a flutter instability by blowing over a stiff blade of grass held between the thumbs. Blowing on the blade causes it to vibrate and this produces a piercing noise. The

Figure 18
The full aeroelastic model used in initial investigations for the Severn and Forth Road Bridges.
National Physical Laboratory, Teddington; Crown copyright.



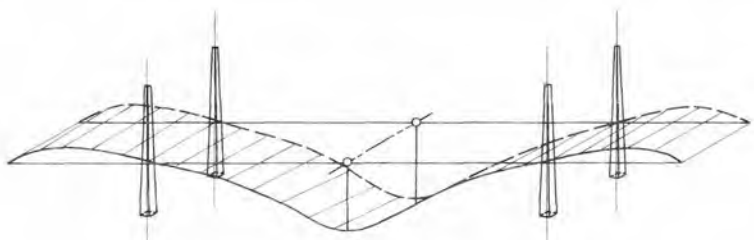


Figure 19
Vertical oscillation.

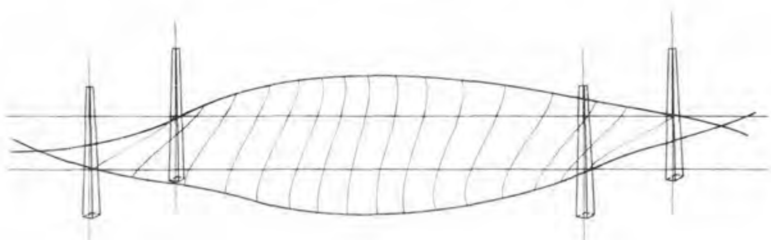


Figure 20
Torsional oscillation.

flutter type of instability has been extensively studied in the aeronautical field because of its highly dangerous nature, and every aeroplane design is thoroughly tested so that flutter is avoided. (During the First World War, classical flutter was a major cause of aircraft losses.) No attempt will be made here to give a physical picture of the cause, but for flutter to occur, the structure must be able to vibrate both vertically and in torsion. To avoid flutter on suspension bridges it was found experimentally and theoretically that the natural frequency of the fundamental torsional mode must be substantially higher than the fundamental frequency of the vertical mode. This makes the occurrence of the coupled motion much less likely.

As a result of the initial research, it was found that the oscillations of a complete

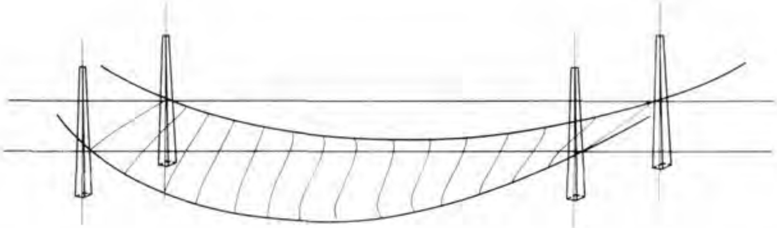


Figure 21
Coupled vertical and torsional oscillation: fundamental mode.

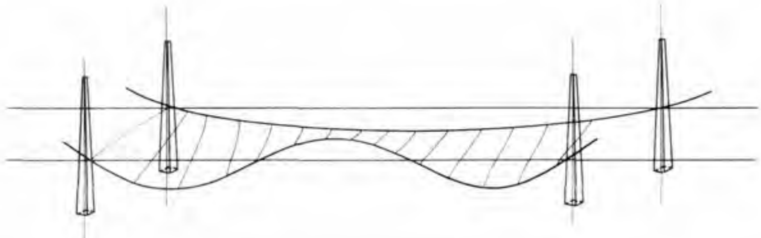


Figure 22
Coupled vertical and torsional oscillation: higher mode.

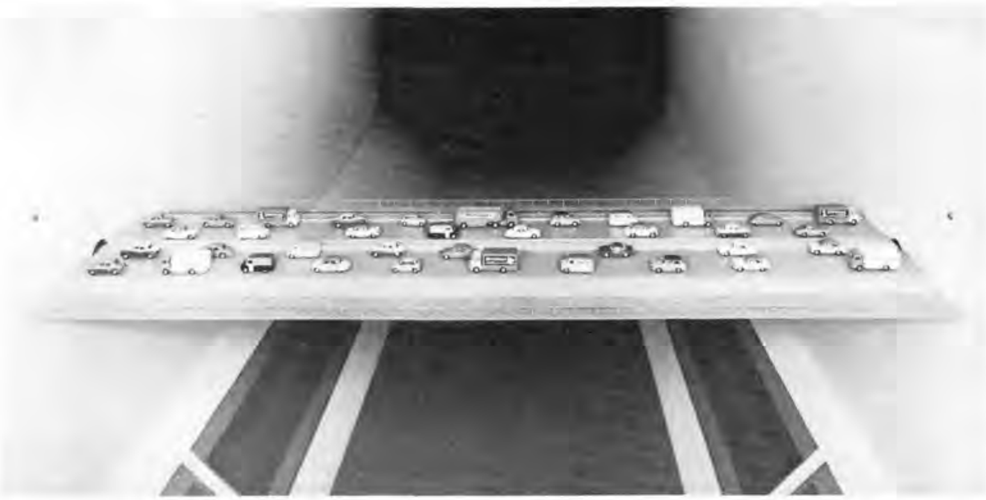


Figure 23
The sectional model used in wind tunnel tests on the Severn Bridge.
National Physical Laboratory, Teddington; Crown copyright

bridge could be reproduced to a satisfactory approximation by using a model of a typical section of the deck. A diagram of a sectional model mount is shown in figure 23. The model itself is rigid but is mounted on springs which reproduce the required elastic characteristics so that it can oscillate in wind with vertical motion to simulate the vertical motion of the full-scale bridge and with a pitching motion to simulate the torsional motion of the bridge. This particular method is now usually used in place of the far more expensive full model method of test.

From such test methods, the design for the Forth Road Bridge finally evolved and a picture of the bridge is shown in figure 24. You will note that it is stiffened in torsion by an open-lattice box. This raised the fundamental torsional frequency well above that of the fundamental vertical mode and avoided the classical flutter instability.

A far cheaper way of increasing the torsional frequency is to use a solid box, but a solid box is notoriously unstable in wind. This research work was carried on for the Severn Bridge to find a shape which would be free from aerodynamic instabilities. The final shape chosen is shown in figure 25. This design has several major advantages. The drag forces are low, vortex excitation is minimal, and flutter speeds are very high. The lessons learned during the research have now been applied successfully to several more bridges:

Figure 24
Forth Road Bridge.
W. Ralston Ltd for Mott, Hay & Anderson Holdings Ltd.

Figure 25
A section of the Severn Bridge being raised into position. Note the streamlined box shape.
William Tribe.

among them is the Humber Bridge, which is the subject of the first article in this book.

A comparison between the Forth, Severn, and Humber Bridges, shown in table 1, emphasizes in terms of mass per unit length the saving effected by using an enclosed box section, despite, in the case of the Humber Bridge, its considerably longer length. This saving in mass also means that towers, tower foundations, and suspension cables can all be lighter.

We have so far been emphasizing the need for a high-stiffness torsion box to fight the flutter. Unfortunately, the box required gets deeper as spans lengthen. The implication is that as spans lengthen, the cables have more and more weight per metre to carry, which leads to a rapid escalation in cost with length until it

Figure 26
A model of the solid twin deck suspension bridge design proposed for the Messina Straits.
British Maritime Technology Ltd.

becomes prohibitive. There is a proposal to span the Messina Straits between Italy and Sicily, and the bridge must have a main span of 3300 metres. The usual type of suspension bridge would be out of the question, but Dr N. C. Brown of Freeman Fox & Partners suggested that it would be possible to build a bridge of the required span if the bridge consisted of a deck with judiciously placed perforations, but no torsion box. The purpose of the perforations would be to decrease the flutter forces, thereby eliminating the need for a torsion box. Subsequent tests showed that his suggestions were valid. Further experimental and theoretical work by J. R. Richardson at the National Maritime Institute (now called British Maritime Technology) showed that the same effect could be achieved by using solid twin decks separated by a large gap but joined every so often by cross girders, so that the two decks worked as a whole; a model of the concept is shown in figure 26. This could well be the very long span bridge of the future.

Table 1

Bridge†	Length of main span/m	Length of side spans/m	Depth of stiffening truss/m	Depth of box/m	Mass per unit length/10^3 kg m^{-1}
Forth	1006	408	8.38		20.9
Severn	988	305		3.05	11.7
Humber	1410	280/530		4.50	13.7

Bridge†	Fundamental symmetric mode‡		Fundamental antisymmetric mode		Height of towers/m
	N_θ/Hz	N_z/Hz	N_θ/Hz	N_z/Hz	
Forth	0.26	0.15	0.35	0.12	156
Severn	0.37	0.14	0.51	0.13	122
Humber	0.30	0.11	0.52	0.10	155.5

†The Forth Bridge has an open-lattice box structure; the Severn and Humber Bridges have enclosed box sections.
‡N_θ refers to torsional oscillations; N_z refers to vertical oscillations.

Electromechanical similarities

A. E. DE BARR
Macclesfield

The behaviour of mechanical systems – which in this context covers systems made up of solids, liquids, and/or gases – is often found to be similar (analogous) to that of electrical circuits. An appreciation of these similarities, or *analogues*, can be helpful in understanding the physics of both electrical and mechanical systems. These analogues also form the bases for the design and use of analogue computers, computers used to study the behaviour of mechanical systems by observation of the behaviour of electrical analogues of the systems to be studied (figure 1). (Now that digital computers are cheap and powerful they are often used in preference to analogue computers. But, when large, complex systems are involved, a digital computer may take a long time to process all the data and produce a result, whereas an analogue computer can often operate in real time and simulate the behaviour of the system as it actually happens. Thus for studying the behaviour of a large system – a chemical plant, perhaps – analogue computers are still useful.)

Figure 1
An analogue computer. The HP 1000 Model 5 Graphics Microsystem performs many 'Automation Workstation' tasks in manufacturing. The low-cost microsystem can monitor and display real-time process or production status information in clear and understandable graphic form.
Hewlett-Packard Ltd.

In this article some of the analogies between electrical and mechanical systems will be discussed with particular reference to
1 flow,
2 vibrating systems, and
3 the use of analogue computers.

Flow

We are all familiar with the sight of water flowing in a river or from a tap, and with the idea of rate of flow as the quantity of water – or other fluid – passing a given point in unit time. The gravitational force on the water causes rivers to flow downhill; the water flows from a place of high gravitational potential to one of lower potential. In general, the greater the gravitational potential difference, that is, the difference in height between the ends of a pipe or two points along the course of a river, the faster the water will flow between them. But the flow rate also depends upon the resistance to flow offered by the pipe or channel; the smaller the resistance the greater the rate of flow. Flow in a closed circuit, such as a circular channel, the arteries and veins in our bodies, or the hydraulic system of a robot, is also inversely proportional to the resistance of the circuit and proportional to the pressure difference across the pump which is supplying the energy necessary to maintain the flow.

The concept of flow with which we are familiar from our experience of the behaviour of fluids is also useful in describing and explaining other physical phenomena, particularly electrical, thermal, and magnetic phenomena. In each case we think of a rate of flow, or a flux, dependent upon the ratio of a driving force to a resistance to flow. In the case of electric current, for example, we have $I = V/R$, and relationships of the form

$$\text{flow} = \frac{\text{potential difference}}{\text{resistance}}$$

can be applied to a variety of systems (table 1).

For historical and other reasons the language describing flow, resistance, and so on is most familiar in the context of electrical circuits and it is often convenient to discuss hydraulic, pneumatic, thermal, and magnetic phenomena in terms of their electrical analogues. It must always be remembered, however, that the analogy between the behaviour of fluids in pipes and that of current in electrical circuits is not exact. And this caution is even more necessary in using the concept of flow to describe magnetic flux or thermal transfer of energy.

The analogy provides a convenient way of describing the behaviour in both qualitative and quantitative terms; it does not necessarily mean that the physical processes involved are identical. In the case of electricity, for example, current is

Table 1
Flow in a cylinder of radius r and length l

Electric current	$I = \dfrac{\text{potential difference}}{\text{resistance}} = \dfrac{V}{l/\sigma A}$
Fluid flow	$Q = \dfrac{\text{pressure difference}}{\text{resistance}} = \dfrac{\Delta p}{8\eta l/r^2 A}$
Heat flow	$\phi = \dfrac{\text{temperature difference}}{\text{thermal resistance}} = \dfrac{\Delta T}{l/kA}$
Magnetic flux	$\phi = \dfrac{\text{current turns}}{\text{reluctance}} = \dfrac{NI}{l/\mu A}$

$A = \pi r^2$ = cross-sectional area
$\sigma = \dfrac{1}{\rho}$ = electrical conductivity
η = viscosity
k = thermal conductivity
$\mu = \mu_r \mu_0$ = magnetic permeability

the rate of flow of electric charge. But, in contrast to the way in which fluid flows, the particles which carry charge past one particular point in a circuit are not necessarily of the same kind as those that carry charge at the same rate past another point in the circuit – see figure 2. In the case of thermal flow no flow of physical particles is involved at all; what is said to flow is energy. (And in the case of magnetic flux no flow is involved but the quantity of magnetic flux does depend on a driving force and a resistance in the same way that the rate of flow of a liquid or electric charge depends upon these factors.)

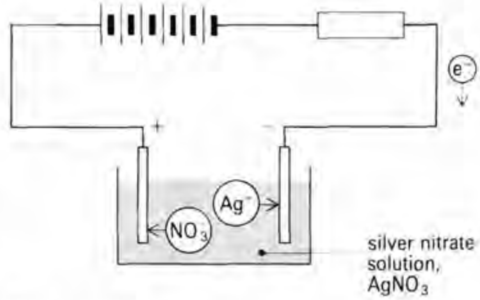


Figure 2
In this electrical circuit the size of the current is the same at all points; in the connecting wires it is carried by electrons but in the electrolytic cell it is carried by ions.

Electric current

The electric current I in a conductor of cross-sectional area A, length l, and electrical resistivity ρ, with a potential difference V between its ends is

$$I = \frac{V}{\rho l/A} \qquad [1]$$

$\rho l/A$ is the resistance of the conductor.

Hydraulic systems

The rate of flow Q of fluid in a cylindrical pipe of radius r and length l is given by

$$Q = \frac{p_1 - p_2}{8\eta l/\pi r^4} = \frac{p_1 - p_2}{8\eta l/r^2 A} \qquad [2]$$

where p_1 and p_2 are the pressures at the ends of the pipe and η is the viscosity of the fluid. (Viscosity is the property of a fluid that makes it hard to pour; treacle has a high viscosity whilst that of water is much lower.) The quantity $8\eta l/r^2 A$ is the resistance offered by the pipe to the flow of fluid. (This equation holds whilst the flow in the pipe is smooth and streamlined; at very high rates of flow the flow becomes turbulent and the 'resistance' of the pipe is no longer constant but varies with the rate of flow. It behaves like a non-linear resistor, one whose resistance depends upon the current in it; a thermistor, for example.)

As might be expected from the similarity of equations [1] and [2] the flow of fluid in a system of pipes in series and parallel behaves like the flow of current in a similar set of resistors. In a badly designed domestic plumbing system, for example, flushing a W.C. on the ground floor may well reduce considerably the flow of water into a basin alongside it – see figure 4(a); or even stop altogether the flow into a basin on an upper floor – see figure 4(b).

Flow in water mains and sewage systems depends upon potential differences and the resistances of pipes and channels; the potential differences may be gravitational or pressure differences created by pumps. The operation of other types of hydraulic system, such as those in industrial robots and earth-moving vehicles, also depends upon pressure differences and resistances and such systems can be designed and explained in terms of their electrical analogues.

Figure 3
Pipelines used to transport fluid at ICI's nylon plant, Wilton.
Imperial Chemical Industries P.L.C., Dyestuffs Division

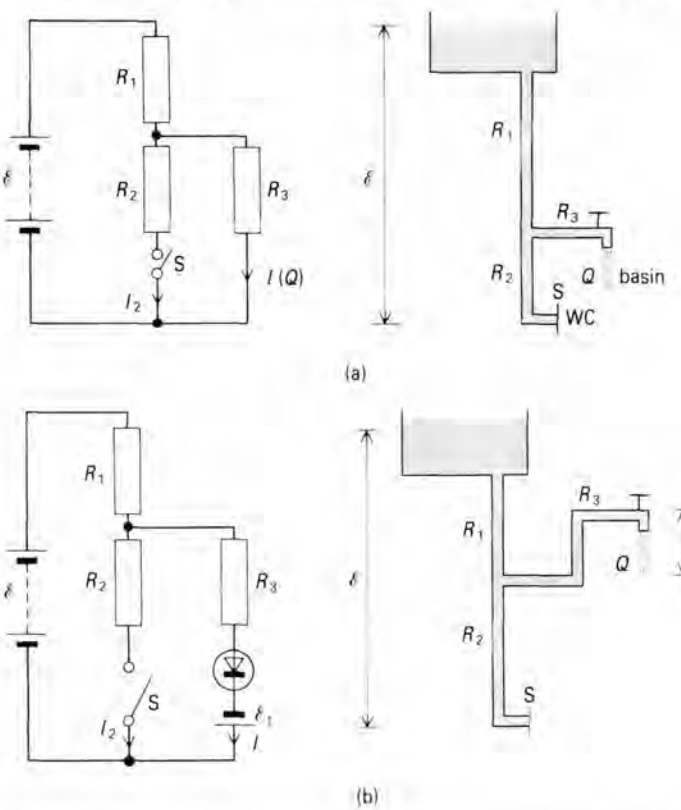


Figure 4
The electrical equivalents of domestic plumbing circuits.
(a) In the electrical circuit, if $R_1 > R_3 > R_2$ then closing switch S greatly reduces the current I. In the water circuit, flushing the W.C. greatly reduces the flow, Q, into the basin.
(b) If, with S closed, $R_2 I_2 < \mathscr{E}_1$ then $I = 0$. In the equivalent plumbing circuit, flushing the W.C. reduces Q to zero.

Pneumatic systems

The flow of gases in pipes or channels obeys the same relationship as that of liquids and there are some interesting applications of the relation between pressure, flow, and resistance.

The coefficient of friction between a packing case and the floor of a warehouse is probably about 0.2; the frictional force that has to be overcome to slide a 1-tonne packing case along the floor is thus about 2000 N. With a cushion of high-pressure air between the packing case and the floor, however, the frictional force is reduced to zero and the only force required to keep the case moving is that required to overcome the viscous resistance of the film of air, probably about 5 N. The force provided by the air under pressure supports the weight of the packing case. In practice the packing case is supported on an air bearing, a skate or pad which is supplied with air under pressure.

The advantages of air bearings are obvious, and are perfectly highlighted by an example that everyone will recognize but possibly not appreciate as being dependent on air bearings. The hovercraft's inflated cushion allows it to skim across the top of the water, thus greatly decreasing the frictional resistance to the forward movement (figure 7). Air bearings are also used to make it easier to move oil storage tanks over rough ground, and in many other applications (figure 8).

Figure 5 (above left)
An industrial robot is used in the production of engine support brackets.
Cincinnati Milacron Limited.

Figure 6 (above right)
A 39-tonne Caterpillar is used in the preparations for the installation of a pipeline.
H. Leverton and Co. Ltd.

Figure 7
A diesel-powered hovercraft.
British Hovercraft Corportion.

Figure 8
A $13\frac{1}{2}$-tonne road paver on air-bearing pads.
Airfloat Systems Ltd.

Figure 9
The Talyrond 200 precision roundness-measuring machine with the Talydata 2000 microcomputer.
Rank Taylor Hobson Ltd.

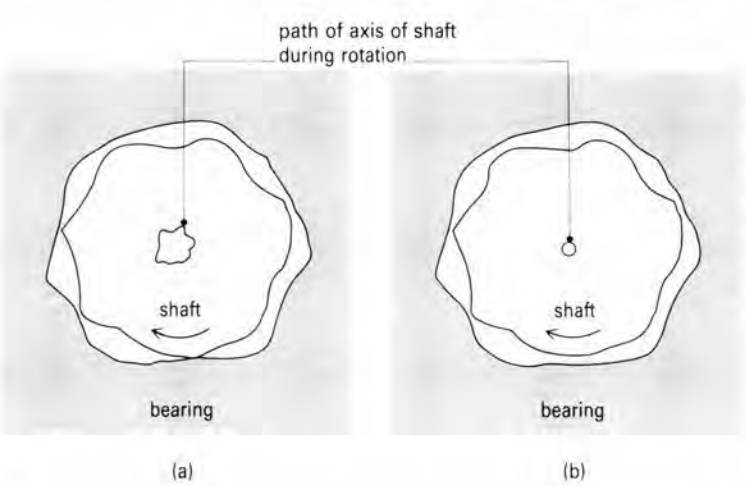
Figure 10
An air bearing helps to ensure accurate rotation of a shaft. (a) When the shaft is supported directly by the bearing the irregularities on the surfaces (shown here greatly exaggerated) cause the axis of the shaft to move about as it rotates. (b) With a film of air under pressure between the shaft and the bearing the movement of the axis of the shaft is greatly reduced; the action of the air bearing tends to maintain the average thickness of the air film constant.

Air bearings can be used to support rotating shafts in precision machines such as grinding machines and roundness-measuring machines (figure 9). As there is no physical contact between the shaft and the bearing pad the accuracy of rotation of the shaft can be much greater than when other types of bearing are used. The air cushion effectively smooths out any irregularities in the form of the bearing surface or shaft and greatly increases the accuracy of rotation – see figure 10.

An air bearing is designed so that the force provided by the air under pressure always balances exactly the weight or other force to be supported. There are two main ways in which this equilibrium can be ensured and both are easily explained in terms of flow and resistance.

i If, in figure 11(a), air is supplied to the bearing pad at a constant rate then the pad will rise until the upward force on the pad just balances the load to be supported; the pressure at the edge of the pad is atmospheric whilst that at the centre increases until the above condition is just satisfied*. The instantaneous effect of a sudden increase in load will be to reduce the gap under the pad and increase its resistance to flow, thus increasing the average pressure under the pad. (The flow is constant as is the pressure at the edge of the pad; an increased resistance to flow must, therefore, increase the pressure at other points under the pad.) Thus the gap will increase again until equilibrium is once more attained with a gap only very slightly less than the original value.

ii In another type of air bearing – see figure 11(b) – air at constant pressure is



(a) (b)

*As the pressure under the pad varies from the centre to the edge the upward force has to be found by integrating $p\,dA$ over the area under the pad.

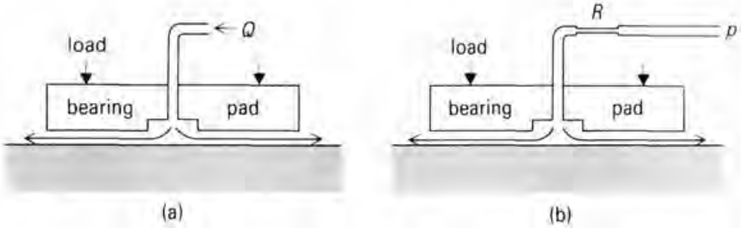


Figure 11
Air bearing pads.
(a) With constant flowrate Q.
(b) Supplied at constant pressure p.

applied to the pad through a fixed resistance R (usually a capillary tube or a small hole in a plate). In equilibrium the upward force on the pad is again just equal to the load. A sudden increase in load will tend to reduce the gap and therefore to reduce the rate of flow of air. The pressure drop across R will decrease so that the pressure at the centre of the pad will increase; equilibrium will be attained when the force again balances the load.

Draw and explain the electrical analogues of both types of air bearing.

Thermal transfer of energy

The rate of thermal flow of energy along a bar ('heat flow'), ignoring losses from the sides of the bar, is given by

$$\phi = \frac{T_1 - T_2}{l/kA} \qquad [3]$$

where T_1 and T_2 are the temperatures at the ends of the bar;
k is the thermal conductivity of the material of the bar;
l is the length of the bar;
A is the cross-sectional area of the bar.

In terms of our analogy, $T_1 - T_2$ represents the potential difference between the ends of the bar, and l/kA represents its resistance to the thermal flow of energy. (Or compare with the electrical resistance of a bar $= \rho l/A = l/\sigma A$ where ρ is the electrical resistivity of the material of the bar and $\sigma = 1/\rho$ is its electrical conductivity.)

It is not easy to imagine or depict realistic circuits representing the thermal flow of energy since, in practice, the elements concerned are rarely of simple geometric shape. Furthermore, unless they are all perfectly insulated at their sides, flow along the circuit elements is accompanied by flow out of the circuit by convection and/or radiation from the sides of the conductors. In the case of flow of fluids this flow out of the circuit would be referred to as leakage, a term that is also used for electrical current and magnetic flux. Nevertheless the concepts of flow and resistance are often useful when considering ways either to increase or to decrease the energy transfer from a hot source.

In a machine tool, for example a lathe, some of the energy dissipated in the driving motor causes it to become hot. Because of the temperature difference created, energy flows from the motor to the structure of the machine tool and consequently some parts of it become hotter than others. The resulting differential expansion can distort the structure and cause loss of accuracy in machining. One way of overcoming this problem is to try to reduce the rate at which energy is transferred from the motor to the machine by interposing a high thermal resistance (for example a layer of material of low thermal conductivity) between them. An alternative approach is to reduce the thermal resistance of heat paths from the motor to the machine so as to reduce the temperature difference associated with a given rate of energy transfer.

Magnetic flux

The magnetic flux ϕ in a toroid of cross-sectional area A and length l is given by

$$\phi = \frac{NI}{l/\mu A} \qquad [4]$$

where μ is the magnetic permeability of the material of the toroid. Electric current in a wire wrapped around a toroid gives rise to the magnetic flux ϕ (figure 12). For a given length of magnetic circuit the more turns of wire N, and the larger the current I, the greater the flux in the toroid. The product of N and I ('current turns') is sometimes called the magnetomotive force; in equation [4] it corresponds to the p.d. or temperature or pressure difference in other flow equations. The denominator in equation [4], which corresponds to the resistance term, is called the 'reluctance' of the magnetic circuit.

The quantity μ is the magnetic permeability of the material of the toroid, it is large

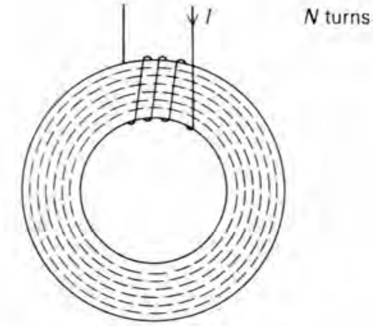


Figure 12
Magnetic flux in a toroid.

for iron and other ferromagnetic materials, and approximately unity for most other materials. (The permeability μ can also be written as $\mu_r\mu_0$, where μ_r is the relative permeability of the material and μ_0 is a constant, the permeability of free space.) Large electromagnets are often used for moving large quantities of iron and steel (figure 13).

Figure 13
A mechanical mobile crane handles scrap with a magnet.
Jones Cranes Ltd.

Series and parallel circuits

Equations [1]–[4] apply to a single homogeneous circuit element. More generally, however, circuits are composed of a number of different elements and, in the case of elements in series:

$$\text{flow} = \frac{\text{total driving force}}{\text{resistance}_1 + \text{resistance}_2 + \text{resistance}_3 + \ldots}$$

With electric current, for example – see figure 14(a):

$$I = \frac{\mathscr{E}}{l_1/\sigma_1 A_1 + l_2/\sigma_2 A_2 + l_3/\sigma_3 A_3 + \ldots}$$

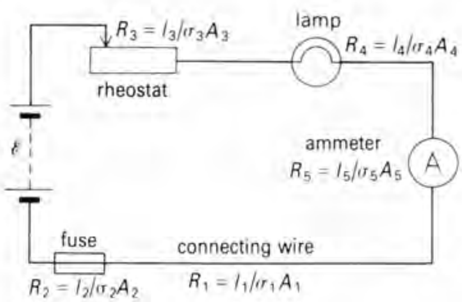
With magnetic flux, and assuming no leakage of flux from the magnetic circuit:

$$\phi = \frac{\text{magnetomotive force}}{l_1/\mu_1 A_1 + l_2/\mu_2 A_2 + \ldots}$$

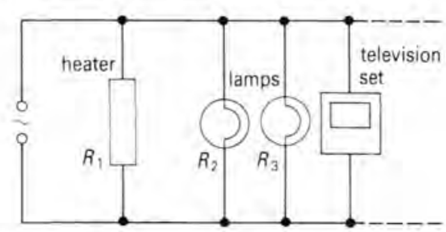
Just as a small air gap – l_1 small and σ_1 very very small – in an electrical circuit reduces the current virtually to zero, so a small air gap in a magnetic circuit (such as the air gap in the magnetic circuit of the moving coil meter in figure 15) greatly reduces the flux in the circuit. And just as a short section of high resistance in an electrical circuit (for example, a 'dry' joint) greatly reduces the current flowing, so a small restriction in a pipe greatly reduces the rate of fluid flow produced by a given pressure difference. Similar relationships apply to the flow of fluid and energy and, in general:

$$\text{flow} = \frac{\text{total driving force}}{\text{total resistance}} = \frac{\text{total driving force}}{R_1 + R_2 + R_3 + \ldots}$$

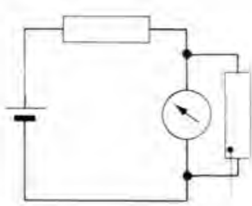
Figure 14
(a) Electrical circuit with elements in series.
(b) Electrical circuit with elements in parallel.
(c) The low resistance shunt greatly reduces the current through the meter.



(a)



(b)



(c) low-resistance shunt for meter

Similarly, with elements in parallel the total resistance, R, is given by

$$\frac{1}{R}=\frac{1}{R_1}+\frac{1}{R_2}+\frac{1}{R_2}+\dots$$

Thus

$$\text{flow}=\frac{\text{total driving force}}{\text{total resistance}}$$

$$=\text{total driving force}\times\left(\frac{1}{R_1}+\frac{1}{R_2}+\frac{1}{R_3}+\dots\right)$$

In the case of electric current – see figure 14(b):

$$I=V/R$$

$$=V\times\left(\frac{\sigma_1 A_1}{l_1}+\frac{\sigma_2 A_2}{l_2}+\frac{\sigma_3 A_3}{l_3}+\dots\right)$$

An element of high conductivity in a parallel circuit greatly increases the current drawn from the source or, if the total current is fixed, the addition of the high-conductivity element greatly reduces the current through the other element, which is in parallel with it – figure 14(c).

In electrical terms these relationships and deductions may seem obvious; they are, however, equally applicable (and probably more useful) in thinking about magnetic flux, the flow of fluids, or the thermal transfer of energy. In figure 16(a), for example, the flow of water is shared between the pipe and the radiator according to the inverse ratio of their hydraulic resistances. In figure 16(b) some of the magnetic flux passes through the air gap in the recording head whilst the rest goes through the magnetic coating on the recording tape. Similarly, a joint between two surfaces may present a high resistance to the thermal flow of energy and so create a temperature difference across the joint which may be undesirable. This can be reduced by placing a small piece of material of high thermal conductivity across (in parallel with) the joint.

Vibrating systems

We have seen that different kinds of flow behave in similar ways and can be described by similar relationships. This is true of vibrating systems also, and in this section we shall first look at some of the similarities between different kinds of vibrating system and then go on to discuss the practical importance of one particular feature of such systems.

A system – a simple pendulum, say, or a mass/spring combination – can be excited into vibration by giving it a small amount of energy, for instance by disturbing it from its rest position. There is then a

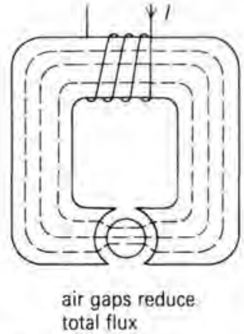


Figure 15
Magnetic flux in an open magnetic circuit.

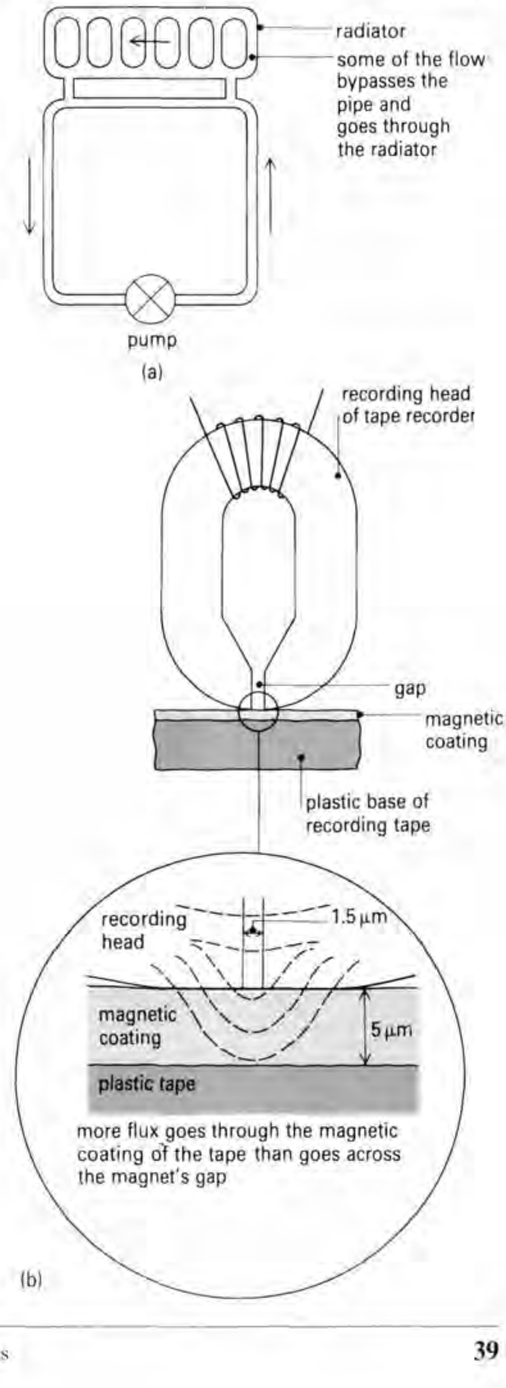


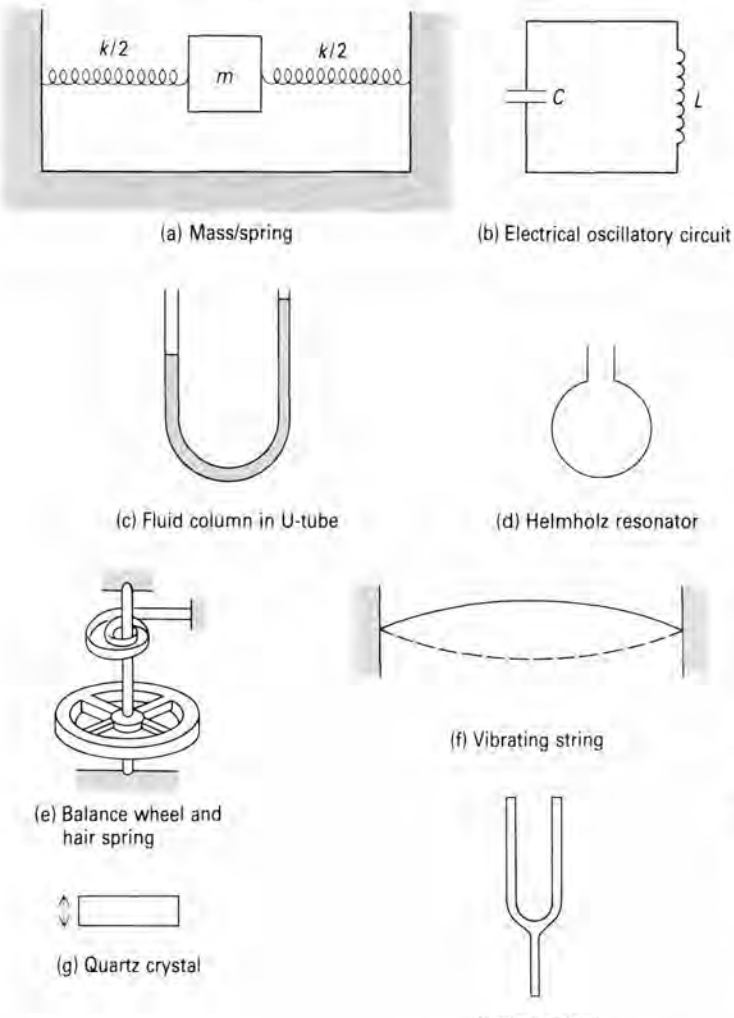
Figure 16
Parallel circuits.
(a) Hydraulic.
(b) Magnetic.

continual interchange of this excess energy between two different forms, leading to the oscillations with which we are familiar. In the case of the mass/spring system the energy takes the forms of kinetic energy of the mass and strain energy in the spring and is concentrated first in one component of the system and then in the other, changing continuously from one to the other as the mass vibrates on the spring. In the simple pendulum the exchange is between the kinetic energy and potential energy of the pendulum bob. If there were no dissipation of energy, once a system was disturbed it would vibrate indefinitely at the same amplitude. But in practice energy is dissipated continuously and, unless maintained in vibration by fresh injections of energy, the system will eventually dissipate all its excess energy and come to rest (see 'Damping', below).

This continual interchange of energy characterizes all vibrating systems. Although almost any system can be excited into vibration by giving it energy in excess of that required for equilibrium, in practice the vibrating systems which we are most aware of are those in which the additional energy is dissipated only slowly; systems which oscillate with only a slowly decreasing amplitude.

The simple pendulum and the mass/spring system are two familiar examples of mechanical vibrating systems; others include tuning forks, bells, violin strings, and the balance wheels of watches. A few other examples of vibrating systems, indicating the forms between which the energy alternates, are given below – also see figure 17.

Figure 17
Examples of vibrating system.



(a) Mass/spring

(b) Electrical oscillatory circuit

(c) Fluid column in U-tube

(d) Helmholz resonator

(e) Balance wheel and hair spring

(f) Vibrating string

(g) Quartz crystal

(h) Tuning fork

Quartz crystal	Strain energy in crystal Kinetic energy of vibration of crystal
Fluid in a U-tube	Potential energy of fluid Kinetic energy of fluid
Helmholtz resonator	Potential energy of compression and rarefaction of air Kinetic energy of air (mainly in the neck)
Tuned (*LC*) circuit	Energy stored in the *E*-field produced by charge on the capacitor Energy of the *B*-field produced by current in the inductor

Period of oscillation

In the case of the mass/spring system the period of oscillation, T, is given by $T = 2\pi\sqrt{m/k}$, where m is the mass and k is the spring constant. A similar relationship applies to all other vibrating systems and, in general,

$$T \propto \sqrt{\frac{\text{inertia factor}}{\text{force factor}}}$$

Thus the period of oscillation of some of the other systems mentioned on page 40 will depend upon:

Table 2

System	Factor determining period T
Mass/spring	$\sqrt{\dfrac{\text{inertia}}{\text{spring force}}} = \sqrt{\dfrac{\text{mass}}{\text{spring constant}}} = \sqrt{\dfrac{m}{k}}$
Balance wheel/hair spring	$\sqrt{\dfrac{\text{inertia}}{\text{spring force}}} = \sqrt{\dfrac{\text{inertia}}{\text{spring constant}}} = \sqrt{\dfrac{I}{k}}$
Quartz crystal	$\sqrt{\dfrac{\text{inertia}}{\text{elastic force}}} = \sqrt{\dfrac{\text{mass}}{\text{elastic constant}}} = \sqrt{\dfrac{\text{thickness}}{\text{elastic modulus}}}$
Fluid in a U-tube	$\sqrt{\dfrac{\text{inertia}}{\text{gravitational force}}} = \sqrt{\dfrac{\text{length of column}}{g}} = \sqrt{\dfrac{l}{g}}$
Tuned circuit	$\sqrt{\dfrac{\text{inertia}}{\text{potential difference}}} = \sqrt{\dfrac{\text{inductance}}{1/\text{capacitance}}} = \sqrt{LC}$

Damping

The dissipation of energy in a vibrating system is referred to as damping and has three main effects.

1 It causes the amplitude of the free vibrations of the system to decrease with time giving rise to damped vibrations, see figure 18.

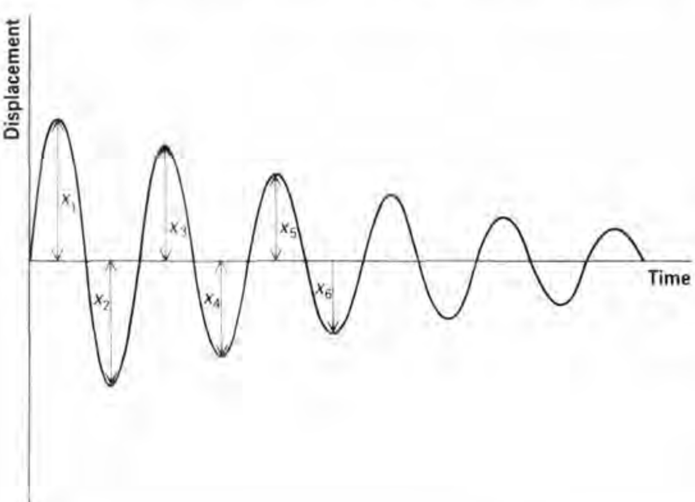
2 The greater the energy dissipation, the greater the energy that has to be supplied to maintain the system in oscillation at a given amplitude, that is, to maintain forced vibrations. Here are some examples of systems deliberately maintained in oscillation:

Clocks and watches	pendulum or balance wheel with energy supplied from weights or a spring via an escapement
	quartz crystal with energy supplied electrically via the piezoelectric effect
Violin string	energy supplied via the bow
Organ pipe	energy supplied by blowing
Tuned circuit	energy supplied electrically

3 At a constant amplitude of vibration, the rate at which energy is dissipated in damping is exactly equal to that at which energy is supplied to the system.

In many mechanical vibrating systems the main source of damping is friction, for example at the pivot of a clock pendulum. The similarity between different types of vibrating system extends also to damping and it is friction – or the equivalent resistance to movement or flow – that is responsible for the damping in all systems.

Figure 18
Damped vibrations.



Friction at pivots and joints

Friction at pivots is easily understood but it may not always be appreciated that, when structures containing bolted or sliding joints vibrate, flexing of the joints produces relative motion between the two surfaces in contact and the frictional resistance to this motion dissipates energy.

Internal friction

Energy is dissipated inside solid materials when they vibrate, the mechanism being referred to as internal friction; it is associated with the movement of dislocations in the crystal structure. In a perfectly elastic material there is no internal friction but real materials are not perfect and their vibration always involves some dissipation of energy which heats up the material.

Internal friction in solids varies greatly from one material to another. The metal from which bells are made, for example, has very little internal friction and the vibration of the bell is only lightly damped. Lead, on the other hand, has so much internal friction that it is almost impossible to excite vibrations in it.

Viscous resistance of gases and fluids

In fluids the equivalent of friction is viscous resistance to flow. The movement of a pendulum in air or of a column of fluid in a U-tube is opposed by viscous resistance and the energy dissipated heats the fluid and its surroundings. The compressions and rarefactions that take place when a column of air vibrates also involve heating of the air.

Electrical resistance

In electrical circuits it is the resistance of the conductors that opposes movement – flow of current – and dissipates energy. Real inductors and capacitors always have some electrical resistance so that resistive losses always cause some damping of the oscillations in *LC* circuits.

Radiation resistance

The energy radiated as sound by a tuning fork comes from the excess energy supplied when the fork is struck; it causes damping of the vibration and thus decay of the amplitude (and sound intensity) with time. A vibrating tuning fork held gently by its base in air is very lightly damped and continues to vibrate for some time after being struck, but radiates sound only at a low intensity. If, however, after being struck the base is placed upon a table or other sounding board much more sound power is radiated and the vibrations of the fork are rapidly damped out. Putting the tuning fork on a sounding board improves the match (see the article on 'Systems', page 12) between the output impedance of the fork and the impedance of the air into which it is radiating sound, and so makes the transfer of energy from the fork to the air easier. Clearly, very little energy would be transferred via the base of the stiff tuning fork to a soft cushion, say, but a hard table top provides a better match and sound energy is radiated more rapidly. The intensity of the sound radiated is increased but the greater rate of loss of energy causes the vibrations of the fork to decay more rapidly.

Vibrations of the air in organ pipes and other wind instruments are also damped by the radiation of sound, and energy has to be supplied (by blowing) to maintain the sound.

Electrical tuned circuits radiate energy as electromagnetic radiation and this damps the oscillations.

Measures of damping

One measure of the damping in a system is given by the ratio of the energy dissipated per cycle to the excess energy in the system. This quantity is often expressed in inverse form as

$$2\pi \times \frac{\text{energy stored}}{\text{energy dissipated per cycle}}$$

and is denoted by the symbol Q, large values of Q being associated with small damping and sharp resonance.

Another measure of damping is given by the rate at which the oscillations decay. The damping in many vibrating systems is of the viscous type, in which the damping force is proportional to the velocity, and in these circumstances the ratio of the amplitudes of successive excursions is constant. The usual measure of the rate of decay is the natural logarithm of the ratio of the amplitudes of successive excursions on the same side of the mean position, and is called the logarithmic decrement, δ. With reference to figure 18 (page 41):

$$\delta = \ln\left(\frac{x_n}{x_{n+2}}\right) = 2\ln\left(\frac{x_n}{x_{n+1}}\right)$$

The logarithmic decrement is related to the quality factor Q by the relationship $\delta = \pi/Q$.

The internal friction of solid materials is often described in terms of their 'specific damping capacity' – the proportion of the energy of vibration which is dissipated each cycle. This quantity is usually designated by the symbol S, and is related to other measures of damping – logarithmic decrement δ and quality factor Q – by the relationships:

$$S = 2\delta$$

$$S = \frac{2\pi}{Q}$$

Specific damping capacity is usually expressed as the percentage $100S$. The specific damping capacities and Q factors of some solid materials are indicated in table 3. (The values given are only approximate as damping varies widely from sample to sample.)

Table 3

	Specific damping capacity, $100S$ (%)	Q
Quartz crystal	0.0006	10^6
Bell metal	0.3	2×10^3
Violin string	0.6	10^3
Steel	2	300
Cast iron	10	60
Mn–Cu alloy	30	20
Rubber	60	5–10

The practical importance of damping

Two main types of situation can be distinguished – those in which damping is kept as small as possible and those in which high damping is desirable.

1 For many purposes a high value of Q is desirable so as to maximize the duration of transient vibrations and/or minimize the energy required to maintain forced vibrations. In these circumstances steps are taken to reduce to a minimum all sources of damping by minimizing external friction and choosing materials with low internal friction; bells, musical instruments, tuned circuits, clocks, and watches are typical examples.

A quartz crystal is, in effect, a mass/spring system in which the only source of damping is internal friction in the quartz, and in well-grown crystals this can be very small (see table 3), so that very little power is required to keep the crystal in oscillation. Because of the piezoelectric properties of quartz this power can be supplied electrically, and so a quartz crystal can be used as a stable, constant-frequency source of electrical signals in clocks and watches.

Stringed musical instruments such as violins are designed to radiate (as sound) most of the energy supplied by the bow to make the strings vibrate. It is desirable that other sources of damping such as friction and internal friction are as small as possible and the materials and methods of construction are carefully chosen to this end. Some bronzes – alloys of copper and tin – have very low internal friction and are used for bells; very little of the energy of vibration induced by the clapper is lost in heating the material of the bell and most is radiated as sound energy.

2 In other circumstances, however, high damping – large energy dissipation – is desirable and again two types of situation can be distinguished: those in which the objective is to maximize energy dissipation and those in which the objective is to minimize the amplitude of vibration.

a Some vibrating systems, for example broadcasting aerials, in which electrical oscillations are maintained by the transmitter, are designed to radiate as much electromagnetic energy as possible and the radiation resistance of the aerial is matched to the output impedance of the transmitter so as to maximize the output of energy. Musical instruments such as bells, violins, and horns are also designed in this way.

b Although resistance is often incorporated into the electrical circuits of, for example, amplifiers in order to dissipate energy and prevent unwanted oscillations, many of the situations in which the energy dissipation in a system is deliberately increased relate to mechanical systems.

There are three main ways in which the damping in a structure or other mechanical system can be increased:

i By using materials of high internal friction; cast iron and concrete have higher internal friction than steel and some alloys such as manganese bronze have very high internal friction. (Lead has very high internal friction indeed but is too soft to be of use for structures.)

ii By incorporating joints; when, for example, a bolted joint vibrates there is usually some relative sliding motion between the surfaces in contact and energy is dissipated by the resultant frictional forces.

iii By the addition of auxiliary vibrating systems designed to dissipate energy by internal friction or friction. Devices such as these, called dampers, are often used to restrain the vibrations of rotating machinery such as diesel engines, but they are also used on machine tools and other machinery.

Perhaps the best-known examples of dampers are the viscous dampers or shock absorbers fitted to cars and other vehicles to damp vibration of their suspension systems. In other situations much smaller amplitudes of vibration are involved and dampers may consist of an auxiliary mass connected elastically to the structure in such a way that the auxiliary system resonates at about the same frequency as the structure itself – figure 19. The auxiliary system is also provided with some means for dissipating energy, often by using rubber of high internal friction for the elastic suspension, but frictional dampers and viscous dampers can also be used.

Machine tools such as lathes, drilling machines, and milling machines are often excited into vibration by some suddenly applied force such as that which might result from a small, hard inclusion in the material being machined and, as a result of interaction between the machining process and the vibration of the structure of the machine tool, there is a tendency for such vibrations to increase in amplitude. Vibration causes a poor surface finish on the workpiece and is thus undesirable. One way of preventing the development of such continued vibration is to increase the damping in the structure. Most of the damping in machine tools comes from friction in joints and the viscous resistance of oil films, but useful increased damping can often be obtained by making parts of the structure in concrete instead of iron or steel. Auxiliary dampers are also used.

There are, however, other situations in which steps are taken to make Q as small as possible. For example, propellers for submarines are sometimes made from manganese bronze alloys (figure 20); the high internal friction of these alloys inhibits vibration of the propeller blades and reduces the amount of noise that they radiate, thus helping the submarine to escape detection by underwater listening devices. Cast iron or concrete would not have this effect. A manganese bronze alloy propeller is shown below.

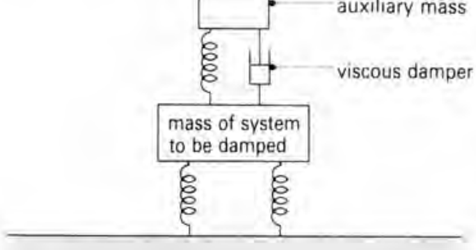


Figure 19 (left)
Auxiliary damper for increasing the damping in a vibrating system.

Figure 20 (right)
The high internal friction of a manganese–bronze alloy propeller reduces the amount of noise radiated.
Stone Manganese Marine Ltd.

Tall structures such as chimneys, cooling towers, power cables, and suspension bridges can be excited into vibration by sudden gusts of wind or other variation in wind forces – see 'Buildings, bridges, and wind', page 18. Interaction between the vibrating structure and the wind forces may cause the structure to be maintained in vibration at its natural frequency; the amplitude of vibration will increase until the energy dissipated in damping is just equal to that supplied by the wind forces. Clearly, if large-amplitude vibrations are to be avoided, high damping is necessary. Buildings of brick or stone contain many joints and the minute movements across these joints when the structure vibrates generate frictional forces which dissipate energy. As a result buildings of this type are much less prone to wind-excited vibrations of large and damaging amplitude than are metal or steel-framed concrete structures.

Auxiliary dampers are sometimes used to damp out unwanted oscillations in power cables.

Electromechanical analogues

The electromechanical similarities introduced earlier can be extended and generalized as shown in table 4.

Table 4
Electromechanical analogues

Electrical	Mechanical
Potential difference	Force (torque)
Current	Velocity
Charge	Displacement
Inductance	Mass (inertia)
Resistance	Damping coefficient
Capacitance	Compliance $\left(=\dfrac{1}{\text{stiffness}}\right)$

Reference to the electrical analogue is not only useful in understanding how a mechanical system will behave (and vice versa) but electrical analogues are used to study, quantitatively as well as qualitatively, the behaviour of systems such as buildings, chemical plants, and aircraft. Consider, for example, the mass–spring–damper system shown in figure 21(a).

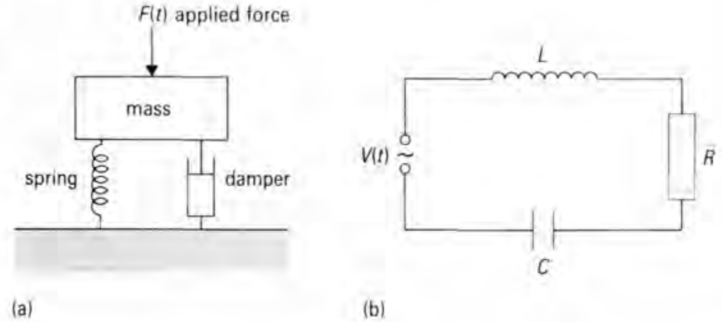


Figure 21
(a) Mass–spring system with viscous damper (damping proportional to velocity).
(b) Electrical analogue of part (a).

The equation of motion of the mass when subjected to a force $F(t)$ is

$$F(t) = ma + Dv + kx \qquad [5]$$

where a is the acceleration of the mass; v is its velocity; x is its displacement from its equilibrium position; D is the damping coefficient; k is the spring constant.

The terms on the righthand side of equation [5] are, respectively, the force accelerating the mass, the force overcoming viscous resistance to motion, and the force stretching the spring.

The electrical analogue of the system of figure 21(a) is shown in figure 21(b). The equation connecting current and voltage is:

$$V(t) = L\frac{dI}{dt} + RI + \frac{Q}{C} \qquad [6]$$

where Q is the charge on the capacitor. The terms on the righthand side of the equation [6] are the p.d.s across the inductor, the resistor, and the capacitor respectively.

Equations [5] and [6] can be written

$$F(t) = m\frac{d^2x}{dt^2} + D\frac{dx}{dt} + kx \qquad [7]$$

$$V(t) = L\frac{d^2Q}{dt^2} + R\frac{dQ}{dt} + \frac{Q}{C} \qquad [8]$$

With suitably chosen values for L, R, C, and V the current in the circuit shown in figure 21(b) represents exactly the velocity of the mass in figure 21(a). Real mechanical systems, such as buildings, power plants, and bridges, are, of course, much more complex than the simple system of figure 21(a); but they are all made up of elements with mass, stiffness, and damping and their behaviour can be represented by a number of interdependent equations

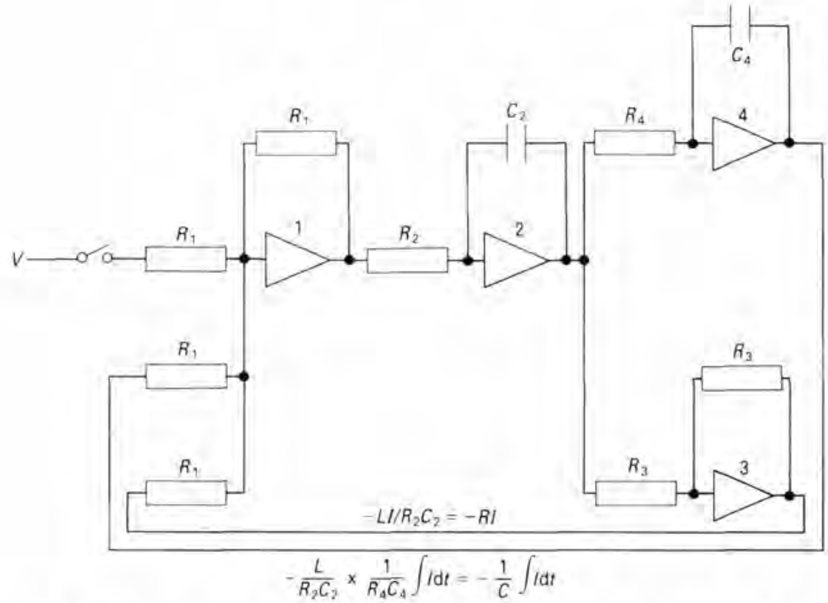


Figure 22
Operational amplifiers making up an analogue computer representing the mass–spring–damper system in figure 21(a).

such as equation [8]. Their behaviour can, therefore, be simulated by a number of interconnected electrical circuits such as that shown in figure 21(b).

The use of electrical circuits to study and calculate the behaviour of mechanical systems is known as analogue computing; an analogue computer consists of a large number of electrical elements – usually operational amplifiers – which can be interconnected via capacitors and resistors in any desired fashion. Such circuits can be used to add, to integrate, and to perform other mathematical operations on electrical signals. In an analogue computer there is some means for applying voltages varying with time – to represent the applied force, for example; there is also means for studying the resultant voltages at different places in the system which represent displacement, velocity, and so on.

With a simple system such as that shown in figure 21(a) it would be possible to study its behaviour by observing the currents and voltages in the electrical analogue, figure 21(b). When more complex systems are involved, with many interconnections, operational amplifier networks provide a more convenient alternative. Figure 22 shows an operational amplifier circuit which is the analogue of the system of figure 21(a).

The first operational amplifier adds the three signals V, $-RI$, and $-\frac{1}{C}\int I\mathrm{d}t$. Since $\int I\mathrm{d}t = Q$, and bearing in mind the change of sign introduced by the operational amplifier, the output of amplifier 1 is $RI + \frac{Q}{C} - V = -L\frac{\mathrm{d}I}{\mathrm{d}t}$ (see equation [6]). The second amplifier integrates this signal and its output is $\frac{LI}{R_2C_2}$. The third amplifier is simply an inverter and its output is $-\frac{LI}{R_2C_2}$. A second integrator (amplifier 4) produces the signal $-\frac{1}{R_4C_4} \times \frac{L}{R_2C_2}\int I\mathrm{d}t$. The values of R_2, C_2, R_4, and C_4 in the analogue computer circuit are chosen so that $R_2C_2 = L/R$ and $R_4C_4 = RC$ where L, R, and C are the values of the components in figure 21(b). In turn, L/R represents m/D, and RC represents D/k in the mechanical system being modelled.

A large analogue computer may contain hundreds of operational amplifier units, each of which can be used in one of several different modes. Nowadays the main advantage of analogue computers over digital computers is that they can more readily be set up to operate in real time representing the behaviour of the system as it actually happens.

The scope and relevance of control engineering

STEVE BANKS
Department of Control Engineering, University of Sheffield

1 Introduction: what is control engineering?

Control engineering is concerned with the alteration of the behaviour of a physical system in some way by an appropriate choice of the inputs to the system. By a 'physical system' one usually means some process which can be modelled mathematically and which can be modified by means of various inputs. In addition, it is essential that information may be obtained in the form of system outputs; that is, measurements may be made on the system. A simple example of a system is a tank of molten glass where the temperature of the glass is to be controlled. In this case the input to the system is the energy transferred thermally, Q, and the output is the (measured) temperature, T_m; the system may be thought of as a 'black box' as in figure 1.

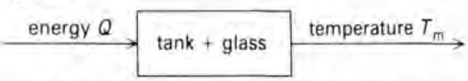


Figure 1

If the temperature is measured by a thermocouple then the actual system output will be a voltage, V_m, dependent on T_m. Such an element, which changes a signal from one form to another, is called a *transducer*. Similarly, the input to the tank may be produced by an electrical heater which is activated by positive voltage. The heating element which provides this input is therefore called the *actuator*. The 'system' may then be regarded as the tank plus the actuator and transducer, and then figure 1 is replaced by the system S in figure 2.

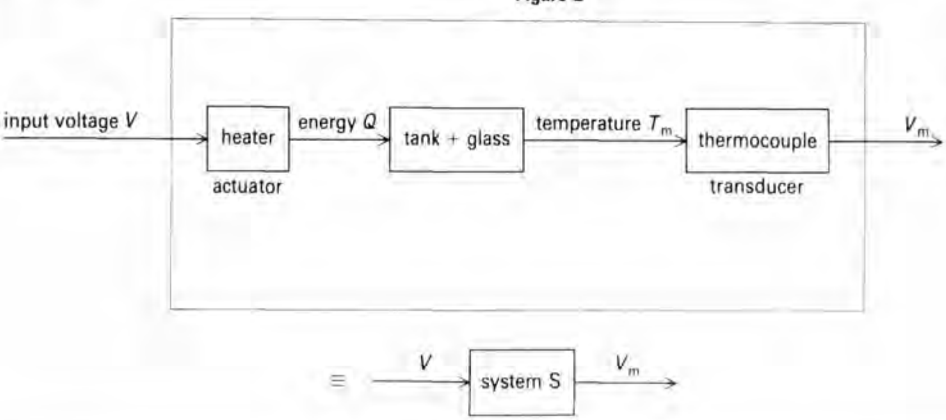
Suppose now that you want to control the temperature of the glass to some fixed value T_d (which is equivalent to a thermocouple output voltage of V_d). Then the error signal $e = V_d - V_m$ can be fed into the heater and, under suitable conditions, will provide the desired temperature regulation. The overall controlled system will then appear as in figure 3.

Note that the error voltage $V_d - V_m$ may be very small and may not be sufficient to drive the actuator (heater). In this case a power amplifier is added to multiply the error voltage by the amplification factor k shown in figure 3. (k is often variable to allow for drifting of the system parameters.)

The feedback system designed above for the glass tank is very crude and is only likely to be effective close to the thermocouple.

So far no mention has been made about the dynamics of the physical processes

Figure 2



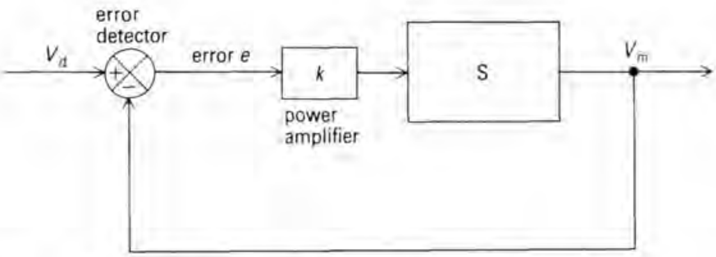


Figure 3

involved in the molten glass. These involve thermal diffusion, fluid dynamics of the glass, characteristics of the tank shape, and actuator and transducer positions. In order to achieve better temperature regulation it is necessary to *model* the system S mathematically. This may be very complex and requires a knowledge of all the physical processes involved in the system. However, supposing that it is possible to obtain a mathematical model, S_m, of S, then one can naturally ask 'how good a representation of S is S_m?'

One of the main tasks of the control engineer is to evaluate the validity of the model S_m under various input conditions. This is usually achieved by analogue or digital *simulation* of S_m and comparison of the computed outputs with the actual recorded outputs from S. If the errors are small we may have reasonable faith in our model S_m.

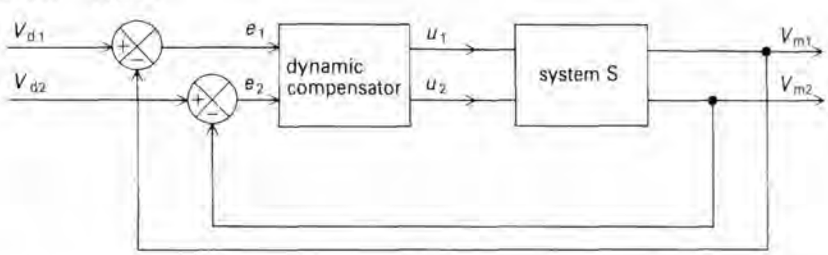
In a more sophisticated control system the glass tank may have several actuators and measuring devices so that the system is multi-input, multi-output. For example, we may also have a variable speed stirrer to mix the glass during the heating process and thermocouple sensors may be placed at several different points in the tank. In this situation it is most unlikely that a simple power amplifier with variable gain, k, as in figure 3, will be sufficient to provide the proper regulation, and it will generally be necessary to design a *dynamic compensator*. The overall system is now shown in figure 4, where the inputs and outputs of the system are defined as follows:

u_1 = voltage supply to stirrer
u_2 = voltage supply to heater
V_{m1}, V_{m2} = measured voltages from thermocouples

The dynamic compensator in most traditional control systems consists of a collection of operational amplifiers and passive electronic network elements (resistors, capacitors, and so on) and must be designed by control engineers for each system that they meet. Mathematically, (u_1, u_2) and (e_1, e_2) are related by a differential equation (which may be quite complex). The advent of very cheap and reliable microprocessors has meant that the dynamic compensator has been replaced by a microcomputer system which processes the inputs e_1 and e_2 to produce the appropriate controls u_1, u_2. Since the microcomputer operates on digital (as opposed to continuous) signals the errors (e_1, e_2) must first be *sampled*: that is, values of e_1 and e_2 are obtained at certain discrete times (usually equally spaced) by an *analogue-to-digital (A-to-D) converter*. Similarly, since the real system, S, operates only on continuous signals, the outputs of the microprocessor must be passed through a *digital-to-analogue (D-to-A) converter*, which essentially smooths out the discrete pulses. The control 'law' relating (e_1, e_2) to (u_1, u_2) is now just a computer algorithm, which again the control engineer must design for a specific system. The problem of connecting a microprocessor to a real physical system is called *interfacing*. The feedback system of figure 4 now takes the final form in figure 5.

The main objectives of control engineering have now been introduced. To

Figure 4



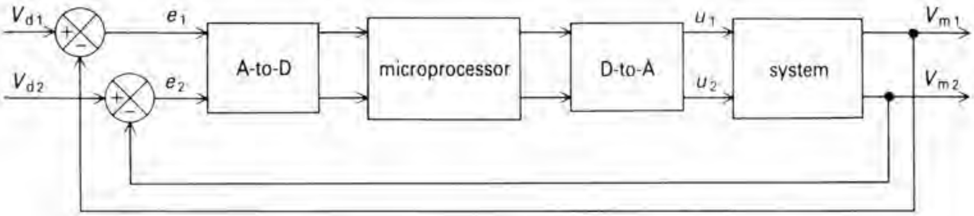


Figure 5

summarize, they are as follows (for a given physical system, S):

1 Determine a mathematical model for S, including the chosen actuators and measuring devices.
2 Simulate the model under various input conditions and compare the results with the real system.
3 Design an appropriate control algorithm and write the computer software to implement the algorithm.
4 Interface the system with a microcomputer.
5 Check the responses of the closed-loop system and, if the results are not as required, improve the system model or the control algorithm.

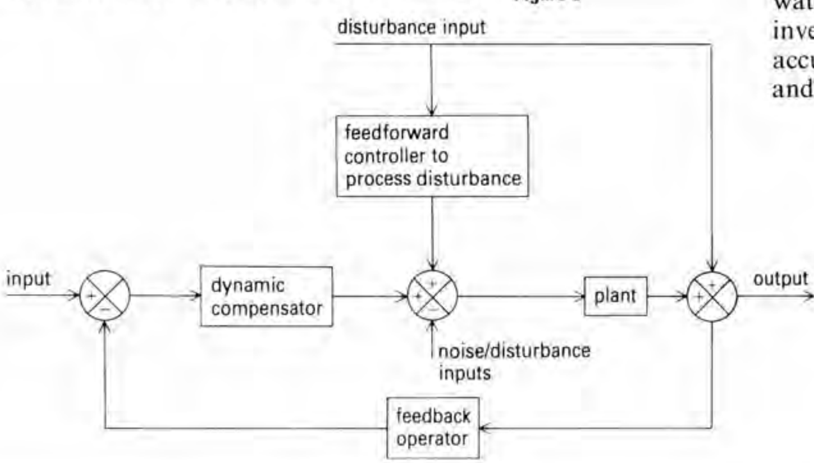
The feedback structure in figure 5 is typical of many applications and a generalized version is shown in figure 6, where the physical system to be controlled is now (as is usually the case) called the *plant*. Note that, in addition to the dynamic compensator introduced above, more general control systems include a feedback operator (often specified by the characteristics of the measuring devices) and may also be subject to general noise and disturbance effects. Feedforward elements are included for various reasons, one of which is to reduce the effects of unknown disturbances.

In the following discussion, after a brief history of automatic control, we shall show how a number of examples of technological systems can be represented in the form of figure 6. Although the examples will be physically quite different, the structure of the overall feedback systems will be seen to be very similar. The plant is usually assumed to be *linear*; that is, a sinusoidal input produces a sinusoidal output with altered magnitude and phase. If the output from such an input is not sinusoidal the system is *nonlinear*. In this case it is usually possible to remove the nonlinear effects by considering small deviations from some *equilibrium* or operating condition. Such a system is said to be *linearized*.

2 A short history of control

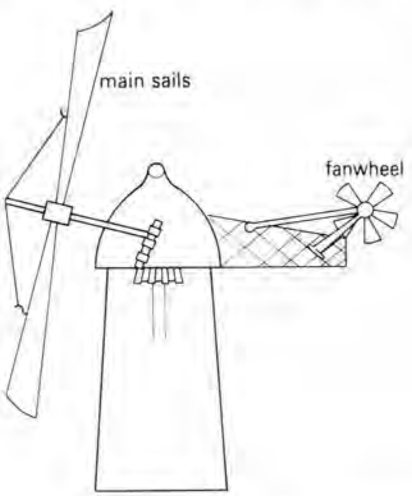
It is generally assumed that automatic control is a 'new' subject involved with high technology. In fact, some of the basic ideas of automatic control can be traced back to approximately 250 BC in Alexandria where the water clock of Ktesibios used a float regulator to control the level of water in a reservoir. Although the clock invented by Ktesibios was not particularly accurate, the basic ideas remained in use and were refined over many centuries.

Figure 6



The next major development in automatic control was the temperature regulator of Drebbel (1572–1633) in Holland, which he applied to an incubator and a furnace, with a great deal of success. With the increasing use of windmills in the eighteenth century it became clear that some form of speed regulator was necessary, and the windmill sails were pivoted so that the area of the sails facing the wind changed automatically with increasing wind speed. In fact, as early as 1745, Lee had patented the fan-tail device, which is a small fan wheel set at right angles to the main sails and which will automatically point the sails into the wind (figure 7).

Figure 7

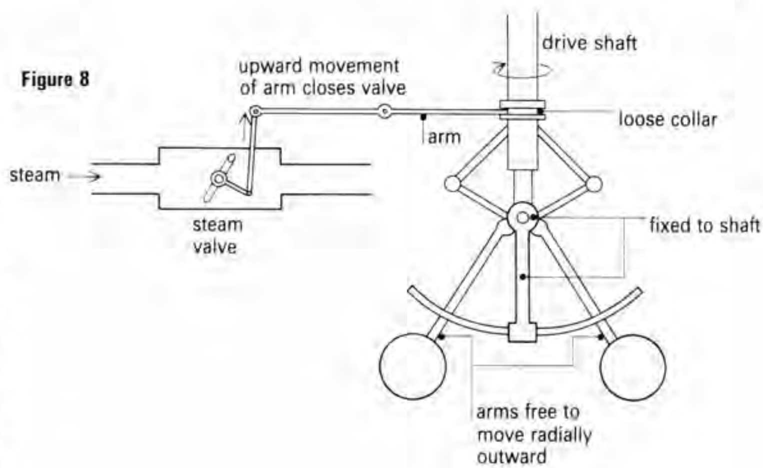


Moreover, a 'centrifugal' speed governor was developed for the main mill drive shaft which predates the famous flyball governor of Watt. The flyball governor consists of a pair of masses attached to the drive shaft of a steam engine as shown in figure 8. As the speed of the shaft increases, the masses spin further from the shaft, and a linkage closes a steam valve, reducing the steam flow to the cylinder. The shaft therefore slows down and the masses drop back, thereby producing the desired regulation.

The Industrial Revolution of the nineteenth century was the turning point for control theory. Although many of the intuitive ideas of control had been known for some time, it was the use of the flyball governor in steam engines which led to a theoretical study of control systems. A high feedback gain was known to be desirable for good regulation, but it became apparent that large feedback gains led to instability. In 1868 J. C. Maxwell gave the first complete theoretical discussion of the stability of Watt's governor. It quickly became obvious that the control of sophisticated systems could no longer rely entirely on intuition and since Maxwell's seminal paper there has been a rapid advance in the theoretical development of control, each new phase being stimulated by some particular application. For example, the desire for heavier-than-air flying machines led directly to the development of much of the classical control theory and later the possibility of space flight led to the foundations of the theory of optimal control (see below).

This development is, of course, still going on, and today the study of systems theory represents one of the most formidable aspects of our desire to understand and control our environment. Indeed, the ideas of control have now penetrated into almost every field of human endeavour – biological and economic systems, chemical plants, aeronautics, and even sociology, where changes in the structure of society and government policy constantly feedback to produce changes in individual behaviour, which in turn alter the society.

Figure 8



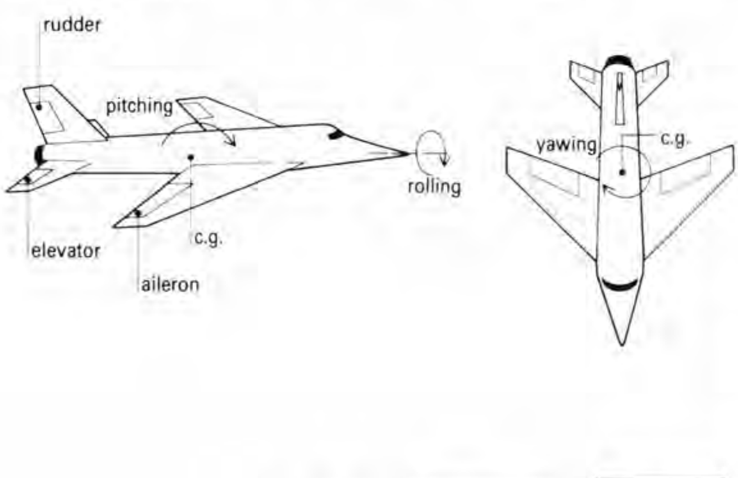
3 Examples of control systems

3.1 Aircraft flight control

The early pioneers of manned flight found (to their cost) that flying machines are inherently unstable. Even so, many designers hoped to achieve 'inherent' stability (that is, without feedback) by setting the angle of the tailplane in accordance with the relative wind and by a correct design of the wing cross-section to provide lateral stability. Although significant progress was made in this direction, systems with a large amount of inherent stability in the presence of steady wind conditions were very prone to being upset by gusts.

It was realized that some form of feedback control was necessary and one of the first reliable systems was introduced by Sperry in 1910. This system, in principle, forms the basic stabilizing controller of most modern aircraft. The elevator, rudder, and ailerons of an aircraft are variable 'flaps' which may be altered by the control system. These are called the aircraft *control surfaces* (figure 9).

Figure 9
Pitch, yaw, and roll of an aircraft (c.g. = centre of gravity).



Suppose that the pilot wants to fly a straight, level course. Since the air through which the plane is flying is not perfectly uniform and the aircraft is not stable, small changes in atmospheric conditions, if not counteracted by the control, will produce disastrous results in the flightpath. However, small changes in the attitude angle of the aircraft from the horizontal can be detected by a gyroscope, the axes of which, when pivoted on frictionless bearings, will tend to remain at a fixed orientation in space. This error can be fed back to hydraulically operated actuators on the control surfaces (such as rudder, elevator, and ailerons) and with the correct feedback compensation will stabilize the aircraft. Note that although the aircraft and its associated aerodynamics are nonlinear, since we are considering small changes it suffices to consider a linearized model. The resulting feedback system then appears as in figure 10.

A similar control system may be designed for the roll and yaw motions, which are usually taken to be independent, although interactions are sometimes considered. In modern aircraft, under more widely varying flight conditions, the compensator will require not only angular displacement measurements but also measurements of angular velocities and accelerations and so it is necessary to include rate gyros and accelerometers in the measuring devices.

The control system described above is suitable for the stabilizing of the aircraft for small deviations about the desired attitude. However, modern flight systems not only require such 'local' stabilizing control but are also subjected to overall guidance strategies. That is, not only must we stabilize the aircraft but we must control it to fly

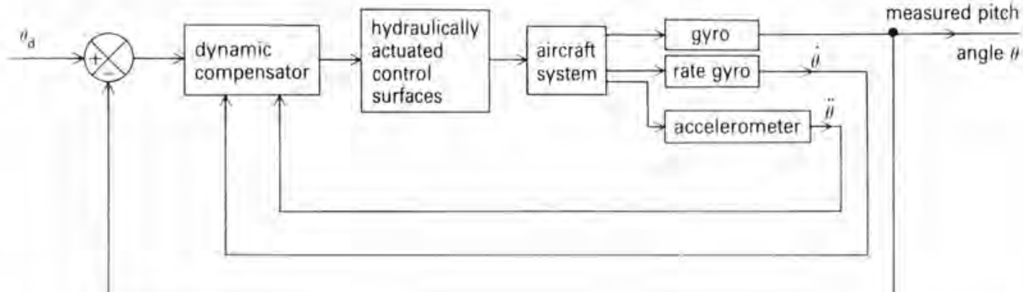


Figure 10
$\dot{\theta}$ is the rate of change of θ;
$\ddot{\theta}$ is the rate of change of $\dot{\theta}$.

along a pre-arranged flight path. For example, most airports transmit a radio beam which guides the aircraft during landing. This provides the onboard computer with beam error information which is, however, corrupted by noise and unwanted disturbances arising from atmospheric conditions, other aircraft, and so on. The resulting signal must be decoded and compensated by the computer. An overall longitudinal control system for vertical approach is shown in figure 11, and is typical of any modern aircraft such as Concorde (figure 12).

Finally, in unmanned flights the guidance control system must be very sophisticated and may contain equipment which senses the local geometry of the terrain and is linked to an onboard guidance computer. Since the flight control system parameters are measured relative to the airframe axes, it is necessary to convert the angular velocities relative to the vehicle axes to those relative to the

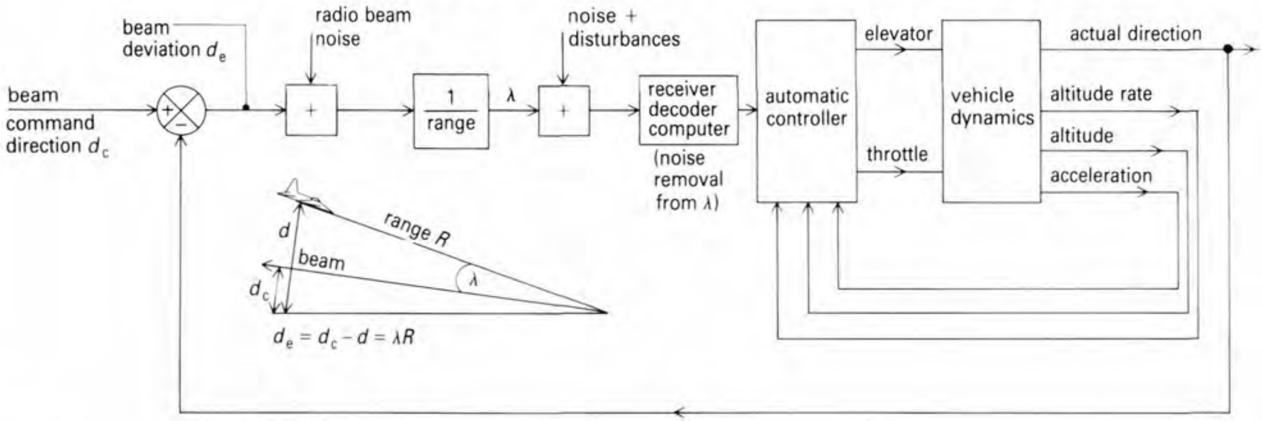


Figure 11

Figure 12
Concorde.
British Aerospace.

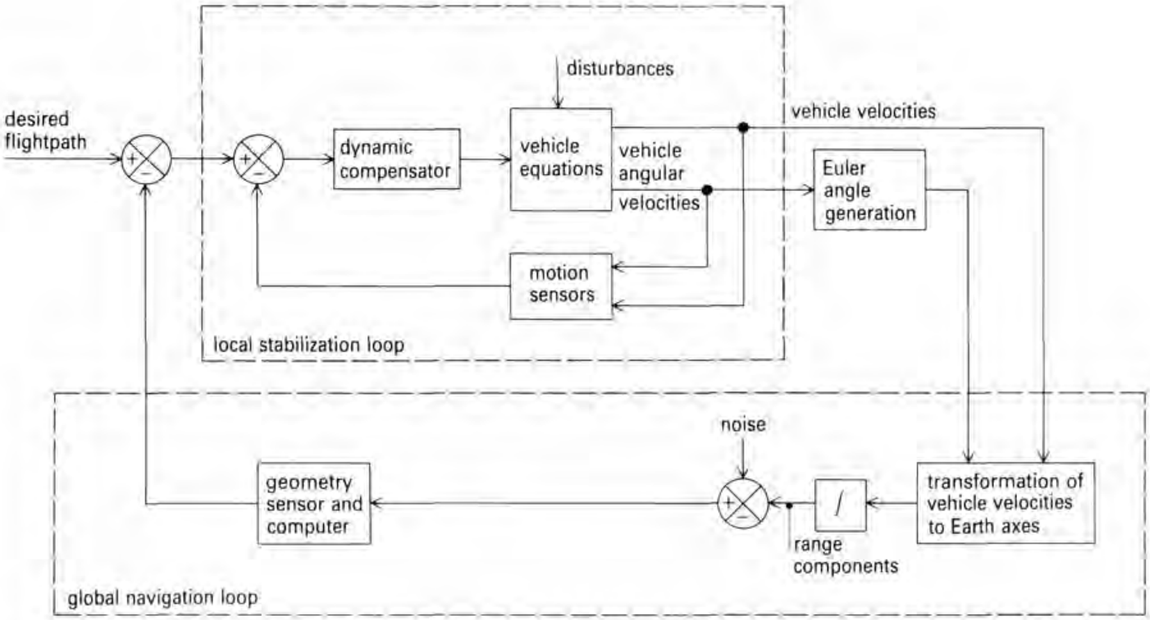


Figure 13

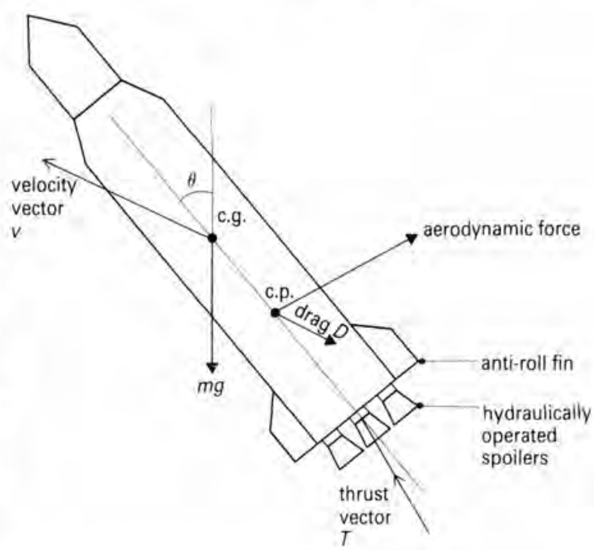
Earth axes by means of the Euler angles (which describe the attitude of the aircraft relative to the Earth). Such a system is shown in figure 13.

3.2 Rocket attitude control

The development of rockets was initiated in Germany by von Braun in the 1930s, to a large extent for military purposes. After many unsuccessful trials it was realized that the basic rocket structure is very unstable and fixed direction thrust motors would not work, no matter how precisely they were aligned. The reason for the basic instability is easy to understand by considering the forces on a modern space rocket shown in figure 14. In most dynamic systems the motion may be reduced to a combination of translations and rotations about the centre of gravity (c.g.) of the system.

However, the aerodynamic forces act through another (generally different) point called the centre of pressure (c.p.). Hence, even if the thrust can be perfectly aligned to pass through the centre of gravity, any deviation of the rocket from vertical will produce a couple around an axis on the line joining c.p. to c.g., thus tipping the rocket further over. This toppling effect is counteracted by introducing gyroscopes and hydraulic spoilers (which vary the thrust direction) on the outlets of the motors; accelerometers are again used to sense angular position, velocity, and acceleration. However, the rocket uses fuel rapidly and hence the centre of gravity changes very quickly. The onboard computer must therefore take the shifting c.g. into account in developing the appropriate input commands. The resulting control system (figure 15) is then very similar to that of an aircraft, shown in figure 10.

Figure 14



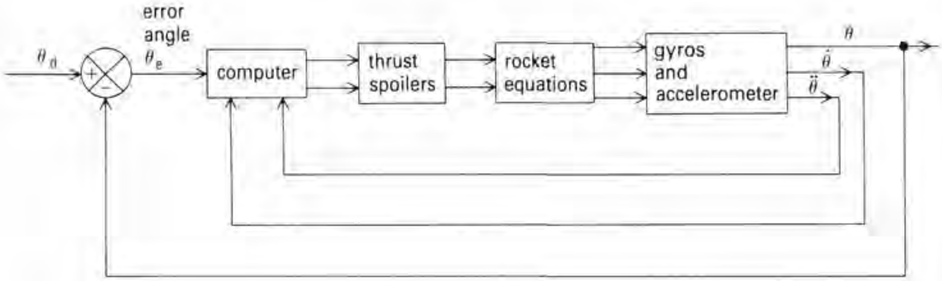


Figure 15

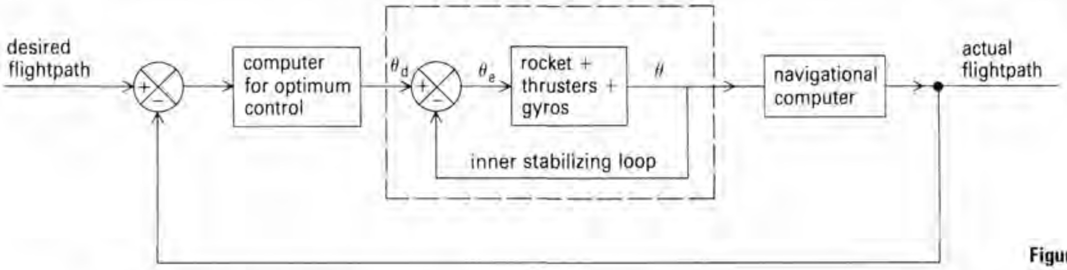


Figure 16

In just the same way as for aircraft, the overall guidance control must also be included in the system. However, since the rate of fuel consumption is so important the controller will be 'optimized' for minimum time of flight or minimum fuel usage. The actual flightpath is determined by a navigational computer with some form of navigational sensors (gyrocompass, stellar observations, and so on) and the complete system is then as shown in figure 16.

3.3 Nuclear reactor control

Another type of system where proper feedback control is vital is the nuclear reactor. The principle of the nuclear fission reactor is that if a neutron collides with an atom of ^{235}U then fission of the uranium nucleus produces, among other things, more neutrons which collide with other atoms of ^{235}U. This chain reaction is characterized by the factor:

$$k = \frac{\text{number of neutrons in one generation}}{\text{number of neutrons in a preceding generation}}$$

(k is called the reactivity. See figure 17.)

An approximate equation for the number of neutrons, y, per unit volume is:

$$\frac{dy}{dt} = \frac{k-1}{l}y \qquad [1]$$

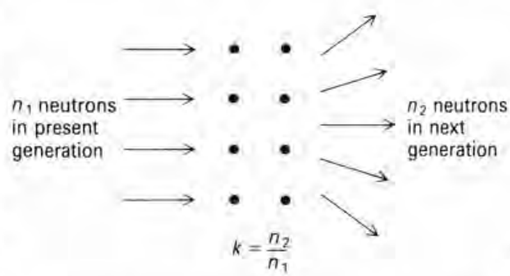


Figure 17

where l is the mean lifetime of a neutron from emission to impact. If $k < 1$ the reactor is subcritical and the number of neutrons decreases exponentially. If $k > 1$ the reactor is supercritical and leads to an exponential increase in the number of neutrons. If $k = 1$ the reactor is described as critical and corresponds to a steady state system. However, the reactor is then only critically stable, that is, any small variation in k above unity will lead to a runaway chain reaction. This simple model suggests that we must run the reactor with $k < 1$, in which case it is impossible to maintain a continuous operation. However, it is fortunately the case that a small proportion ($\approx 0{\cdot}75\,\%$) of the neutron emissions does not occur with the fission of ^{235}U but is delayed and appears a minute or so later from the fission products by radioactive decay. If a fraction, β, of the neutrons is delayed and is produced by the fission products, z, then the original equation is replaced by the equations:

$$\frac{dy}{dt} = \frac{k-1-\beta}{l} y + \alpha z \qquad [2]$$

$$\frac{dz}{dt} = \frac{\beta}{l} y - \alpha z \qquad [3]$$

where α is a decay constant. The control of the neutron level is implemented by control rods which are essentially absorbers of neutrons. These are made of boron or cadmium and may be raised or lowered in the reactor. The more the rods are raised, the fewer neutrons they will absorb, and vice versa. If thermal effects are taken into account, then a simple model is obtained by replacing k in the above equations by $k_c - \gamma T$, where k_c is the reactivity due to the control rods, γ is a constant, and T is the temperature. In addition to equations [2] and [3], we have the temperature equation:

$$\frac{dT}{dt} = ay - b(T_c - T) \qquad [4]$$

where a, b are constants and T_c is the control temperature. The three equations [2], [3], and [4] represent a simplified model of the reactor, and these equations may be linearized to obtain a suitable control model. A compensator is designed which will raise and lower the control rods under computer control and the overall feedback system is shown in figure 18.

One of the fuel rod charge/discharge machines of Dungeness A nuclear power station is shown in figure 19.

Figure 19
A charge/discharge machine at Dungeness A nuclear power station.
Central Electricity Generating Board.

3.4 Biological systems control

Most biological systems are adaptive. That is, they can change various internal parameters in response to changes in the environment. This adaptivity is due entirely to biological feedback systems. For humans, in most cases the control algorithm (that is, the compensation introduced above) is performed by the brain, either consciously or subconsciously. Consider, for example, the 'simple' act of picking up an object. The brain, via nerve impulses and muscle contraction, drives the arm towards the object and with a very sophisticated position control calculation, recognizes the shape of the object and orients the hand to the correct position over the object and then closes the fingers.

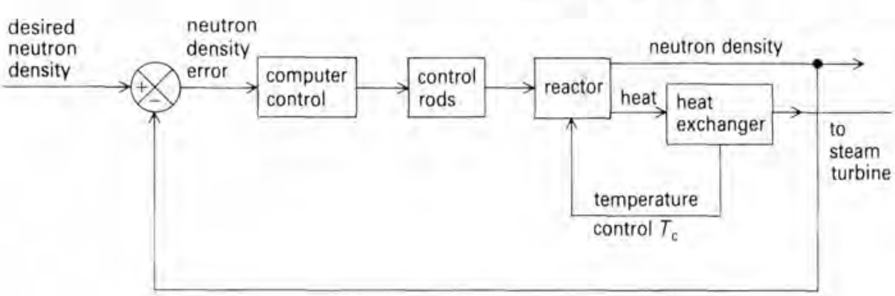


Figure 18

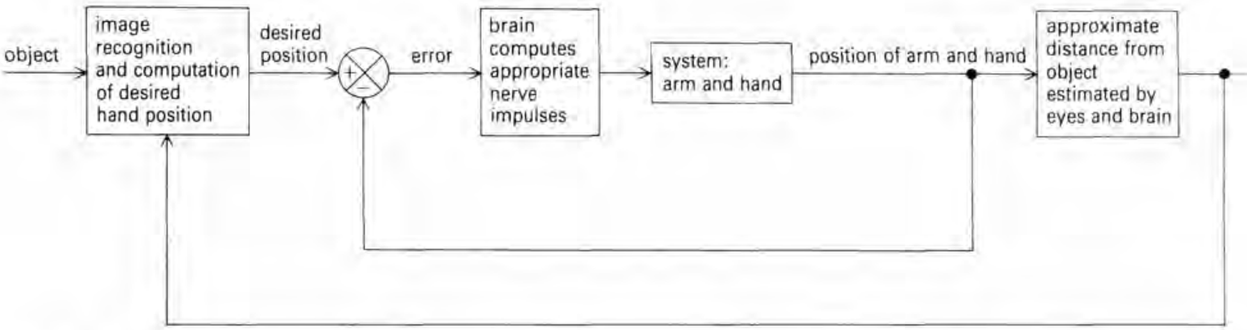


Figure 20

In figure 20 the basic control system is shown. Note that before the object can be approached, the brain must recognize the shape and estimate the weight of the object. From this, a computation of the hand and arm position which will be effective in picking up the object is made. However, this will only be approximate and may itself change adaptively as the hand approaches the object. This apparently simple operation is therefore seen to be, in fact, a very sophisticated process for the brain. It appears simple to us because most of the processing software has been developed during learning in the early years of life and is stored as subroutines in the memory. Instead of having to think about the object, we merely call up the appropriate 'programs' from our memory, and adapt them to the particular object considered.

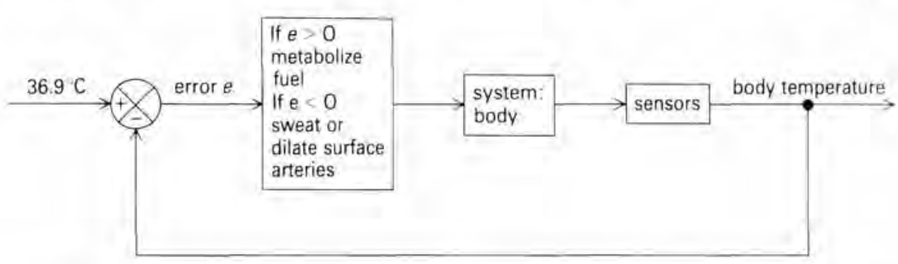
Another example of feedback control in the animal body is the precise control of temperature. The actuators are of basically two types; those which heat up the body and those which cool it down. In the former type, the metabolism of carbohydrates to sugar in the liver produces heating. The cooling devices are of various forms, but consist mainly of sweating and increasing the diameter of the blood vessels. By sweating the body loses energy in the form of the latent heat of evaporation. The body has many temperature sensors distributed throughout the skin and other organs.

Note that the system in figure 21 is not as simple as it may seem, since the control is 'distributed'. This means that one part of the body may be too hot while another is too cold and so the system has many (several million) inputs and outputs.

3.5 Chemical plant control

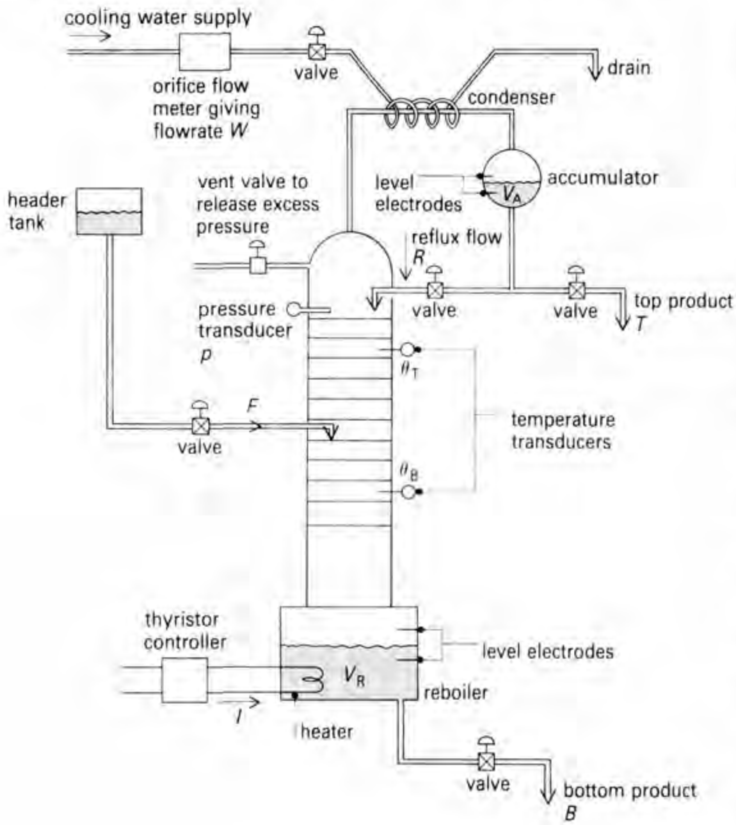
In order to demonstrate the wide applications of control, the next example will consider a typical chemical plant. The particular system which we have chosen is a distillation column, but the principles are similar for many types of chemical engineering systems. A typical distillation column is shown in figure 22, and consists of a series of trays containing small holes through which the vapour passes. A header tank feeds a mixture such as alcohol and water (we consider a binary column with just two products) to the midpoint of the column with a flow rate F. This will be assumed constant, but may itself be a control variable. The reboiler at the base of the column contains an immersion heater which is controlled by a thyristor circuit. As the heated liquid vapor-

Figure 21



izes, the vapour rises through the trays becoming richer in alcohol as it proceeds. At the top of the column the vapour (which is now mainly alcohol) is condensed by passing cooling water through the condenser and the product (called the *top product*) is stored in the accumulator. Some of the top product is removed and some is returned to the top of the column. In the steady state operation of the column, the bottom product from the reboiler should be mainly water.

Figure 22



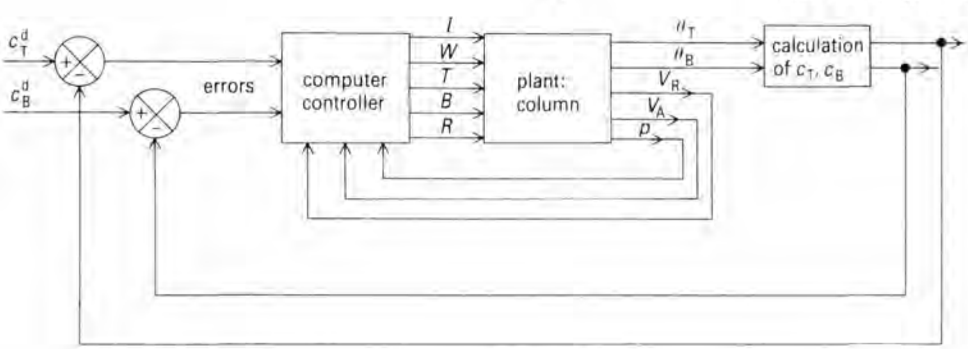


Figure 23

The control inputs to the system are the current, I, to the heater; the rate of cooling by the condenser, which is controlled by adjusting the cooling water flowrate W; the rates T, B, R of the top product removal, bottom product removal, and reflux flow respectively. The measurements are the temperatures θ_T, θ_B at the top and bottom of the column (respectively); the volumes V_R and V_A of liquid in the reboiler and accumulator; and the pressure, p, in the column.

The temperature outputs, θ_T, θ_B may be used to estimate the concentrations, c_T, c_B of the top and bottom products respectively. If the object is to control the concentrations c_T and c_B to some desired values c_T^d, c_B^d then the control system will appear as in figure 23.

3.6 Robotics

In the final example we shall consider one of the most exciting developments of control, which is still in its infancy: that is, robotics. The robots which we have all seen on television working in industry look very impressive, but most are operating in an open-loop way. In other words, they are merely repeating over and over simple commands which are stored in the computer memory. Research in robotics is now mainly into 'intelligent systems' which can adapt to different situations and which must therefore be closed-loop.

As a typical application consider designing a robot system which will perform the task of picking up an object of arbitrary shape, which we considered in terms of human behaviour earlier. If the robot is to find the object itself it must be provided with 'eyes'. These will probably be television cameras which can scan an image and produce a digital output for a computer (the 'brain' of the robot). This output will be processed by the computer to remove noise and other unwanted disturbances introduced by the camera. *Image processing* of this type is well-known from the space flights which have sent pictures of the planets to Earth. The computer brain of the robot must then use pattern recognition software to develop an internal 'model' of the object which it is to pick up. This part of the chain is still under intensive research and it is likely to be some time before a completely satisfactory solution is found. Next, the actual control software must be written; this has to contain certain subroutines which are common to the acquisition of any object (for example, to move the robot near the object and position the 'hand' and 'fingers' close to the object), and others which are adaptive. The latter must be able to adjust the controller for the particular object which is being lifted and the writing of this software is another very difficult part of the problem. Finally, the robot must have some form of ranging device, which at present is usually an ultrasonic sensor, but for much more accurate tasks the distance from the object could be found stereoscopically, for instance, by giving the robot two 'eyes' (cameras). Figure 20 may be redrawn for the robot system as in figure 24. A computer-controlled robot arm is shown in figure 25.

Figure 24

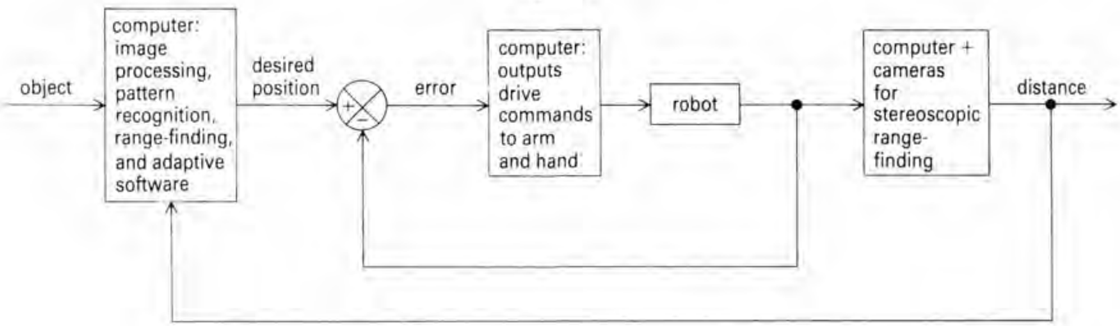


Figure 25
A computer-controlled robot arm.

4 Modern trends in control

4.1 Optimal control

Most of the control systems considered in section 3 have been successfully treated by using essentially classical techniques. These techniques are based on determining the *frequency response* of the plant, which may be defined as follows. If the plant has a single input and a single output and we input a sine wave of angular frequency ω and unit amplitude, then the plant is said to be *linear* if the output is a sine wave. However, the output sine wave will be changed in both amplitude, A, and phase, ϕ (figure 26). The values of A and ϕ will depend on ω and so are denoted as functions $A(\omega)$, $\phi(\omega)$. The complex function $A(\omega)e^{i\phi(\omega)}$ is called the *frequency response* of the system and contains enough information to design feedback controllers.

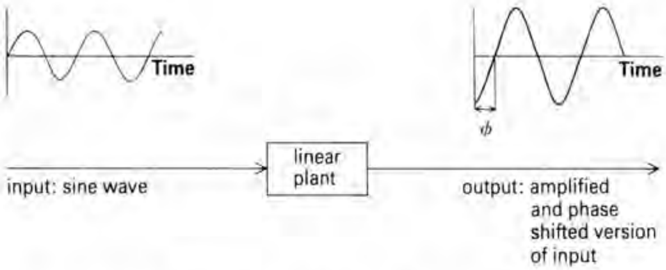


Figure 26

Although such controllers are adequate in many situations, they are usually far from being the best. In most industries in which energy and time are at a premium, interest in determining optimal controllers has been growing recently. Mention has already been made of the desire for a minimum fuel control in the rocket system. To develop this example suppose that the dynamics of the rocket can be modelled by a mathematical equation depending on a set $\{x\}$ of variables for the position, velocity, and so on of the rocket, and a set $\{u\}$ of variables describing the control inputs. If, for simplicity, the rocket burns fuel at a constant rate, then minimizing fuel is equivalent to minimizing the time of flight. Moreover, if one of the controls, u_1, is the angle θ of the spoiler then we require:

$$|\theta| = |u_1| \leqslant \theta_{\max} = \text{maximum possible spoiler deflection} \qquad [5]$$

Finally, if it is to fly on a prearranged path $x_{\mathrm{d}}(t)$, then we would wish to minimize the integral:

$$J = \int_{t_0}^{t_f} (x - x_{\mathrm{d}})^2 \, \mathrm{d}t + \int_{t_0}^{t_f} \mathrm{d}t \qquad [6]$$

where t_f is a variable final time which is determined by the theory. (The first term guarantees closeness to the desired trajectory and the second term guarantees minimum time.) The integral in equation [6] must be minimized over all choices of x which satisfy the equations of the system and the constraint of equation [5], and represents a basic type of problem in optimal control. The solution is very difficult, in general, and often requires a numerical solution on a computer.

4.2 Noise filtering

In the examples in section 3, we have also mentioned the problem of unwanted noise which may be injected into the system and which may be beyond our control. (In radar, for example, many unwanted echoes are received which 'clutter' the picture.) This noise is often modelled by introducing a term v into the equations of the system, where v is not known completely (although some of its statistical properties may be known). In fact, v is often assumed to be 'white' noise. Also, it is often the case that the 'states', x, of the system are not all measurable directly and instead of receiving x from our transducers, we actually measure a function:

$$y = g(x) + \mu \qquad [7]$$

of x, with the associated measurement noise μ. Then we wish to find an estimate, $\hat{x}$, of the state x which will minimize the 'cost function':

$$J = \text{mean}\,|y - g(\hat{x})|^2 \qquad [8]$$

Again this is a difficult problem and we shall not expand on it here.

4.3 Adaptive control

In the preceding discussion it has been assumed that one has a good model of the system dynamics. In general this will not be the case. Indeed, it may not even be desirable, since a poor quality model may suffice for control purposes. The model may be uncertain because the physics of the system is not well understood and so the structure of the equations describing the process may be unknown. Although we may be able to write down the form of the equations, certain parameters may be unknown or may vary with time. For example, the dynamic equations for an aircraft are well understood, but parameters such as the speed of sound vary with altitude, air pressure, and so on. These considerations, coupled with the fact that real systems are nonlinear and contain noise, have led in recent years to a different approach to controller design. These controllers are called *adaptive* since they can adapt to changes in the system structure or parameters. The basic idea is to introduce a model which may be nonlinear and which may or may not be very close to the real system. (Of course, the closer our model to the real system, the better.) In a *model reference adaptive system* (figure 27) the plant is controlled by a classical feedback compensator which has variable coefficients. The desired output is also fed into the system model and the error between the real system and the model is monitored and used by an adapting algorithm which changes the coefficients of the compensator. If the system is noisy and the output, y, is a function of the desired state, x, then an estimator must be included to find $\hat{x}$. Essentially, we are trying to 'force' the real system to follow the model, S_m.

Another approach is used if the uncertainty in the system is in various unknown parameters, and is called *self-tuning* control. In this case a model of the system is obtained with unknown parameters included and a controller is designed on the basis of the model. The self-tuning device essentially provides an on-line estimation of the parameters which are required by the controller. If the model is a set of equations depending on a collection, $\{\theta\}$, of unknown (or noisy) parameters, then a controller designed by the methods of sections 4.1 and 4.2 will depend not only on the system equations but also on θ. If the control is represented by $u(\theta)$ then the self-tuner provides an estimate $\hat{\theta}$ of θ and the control which is implemented is $u(\hat{\theta})$ (figure 28).

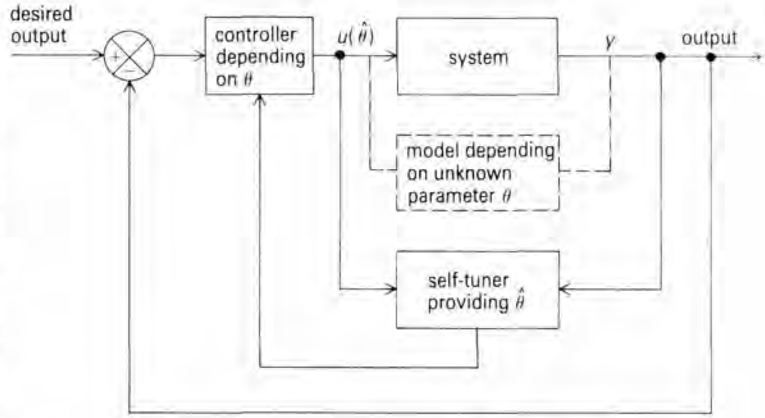


Figure 28

4.4 Further developments and research

Current research in control engineering is concerned mainly with the problems of uncertainty in the system models and also with the development of a more general and unified theory of nonlinear systems. An important technique which has been introduced in recent years is that of considering very large scale systems (for example, in power distribution or telephone networks) to be divided into a series of interconnected subsystems. The control theory of each subsystem is studied independently and the network topology then defines a 'hierarchical control structure' on the overall system.

The related fields of data processing, pattern recognition, and information technology are also rapidly expanding areas of research interest and there are likely to be many very interesting problems for the next generation of scientists and mathematicians who would like to work in this field in the coming years.

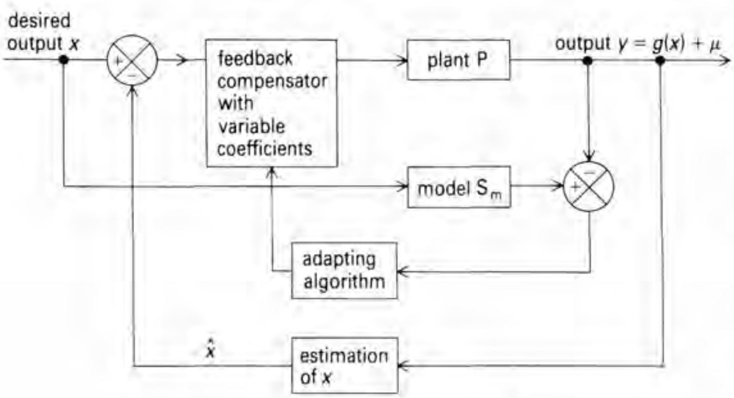


Figure 27

General editor, Revised Nuffield Advanced Physics
John Harris

Consultant editor
E. J. Wenham

Editor of this book
John Harris

Contributors
Steve Banks
A. E. De Barr
David Fisher
D. E. J. Walshe
L. R. Wootton

Physics in engineering and technology **is one of the background Readers for the Revised Nuffield Advanced Physics course. The five articles in this book demonstrate the wider applications of physics in engineering and technology. Thus the topics include a description of the design and building of the Humber Bridge, a study of systems, an article on the effects of wind on buildings and bridges, a discussion of the use of models as analogues of the behaviour in mechanical systems and electric circuits, and an introduction to the concepts of control engineering.**

ISBN 0 582 35422 6

Longman